



# Universality of scholarly impact metrics



Jasleen Kaur\*, Filippo Radicchi, Filippo Menczer

Center for Complex Networks and Systems Research, School of Informatics and Computing, Indiana University, Bloomington, USA

## ARTICLE INFO

### Article history:

Received 6 June 2013

Received in revised form 5 September 2013

Accepted 9 September 2013

### Keywords:

Impact metrics

Discipline bias

Universality

Bibliometrics

## ABSTRACT

Given the growing use of impact metrics in the evaluation of scholars, journals, academic institutions, and even countries, there is a critical need for means to compare scientific impact across disciplinary boundaries. Unfortunately, citation-based metrics are strongly biased by diverse field sizes and publication and citation practices. As a result, we have witnessed an explosion in the number of newly proposed metrics that claim to be “universal.” However, there is currently no way to objectively assess whether a normalized metric can actually compensate for disciplinary bias. We introduce a new method to assess the universality of any scholarly impact metric, and apply it to evaluate a number of established metrics. We also define a very simple new metric  $h_s$ , which proves to be universal, thus allowing to compare the impact of scholars across scientific disciplines. These results move us closer to a formal methodology in the measure of scholarly impact.

Published by Elsevier Ltd.

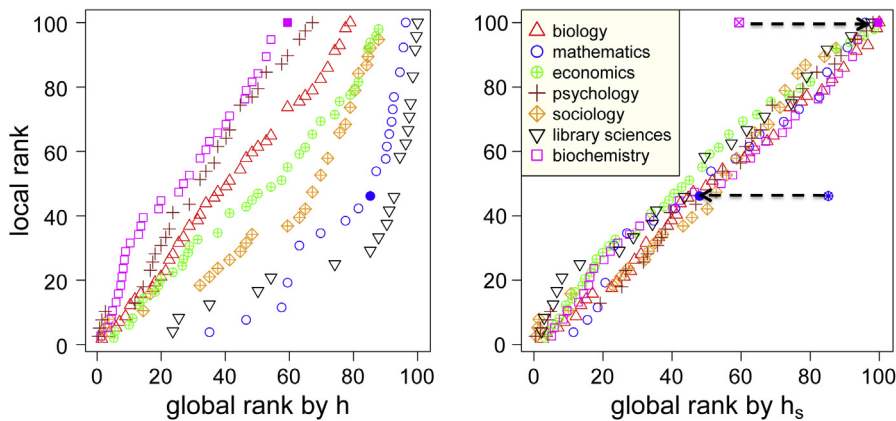
## 1. Introduction

Objective evaluation of scientific production – its quantity, quality, and impact – is quickly becoming one of the central challenges of science policy with the proliferation of academic publications and diversification of publishing outlets (Abbott et al., 2010). Many impact metrics have been and continue to be proposed (Van Noorden, 2010), most of them based on increasingly sophisticated citation analysis (Lane, 2010). These metrics have found wide applicability in the evaluation of scholars, journals, institutions, and countries (Davis & Papaneck, 1984; Garfield, 2006; King, 2004; Kinney, 2007). Unfortunately, there is very little work on quantitative assessment of the effectiveness of these metrics (Bollen, Van de Sompel, Hagberg, & Chute, 2009; Radicchi & Castellano, 2012) and the few existing efforts are proving highly controversial (Lehmann, Jackson, & Lautrup, 2006). This is alarming, given the increasingly crucial role of impact analysis in grant evaluation, hiring, and tenure decisions (Bornmann & Daniel, 2006).

Discipline bias is probably the most critical and debated issue in impact metric evaluation. Publication and citation patterns vary wildly across disciplines, due to differences in breadth and practices. These differences introduce strong biases in impact measures – a top scholar in biology has a very different publication and citation profile than one in mathematics. This has led to a recent burst of interest in *field normalization* of impact metrics, and the emergence of many “universal” metrics that claim to compensate for discipline bias (Colwell et al., 2012). Fig. 1 illustrates the idea of field normalization. If we rank scholars across all disciplines according to an unbiased (universal) metric, a scholar in the top 5% among mathematicians should be ranked the same as a scholar in the top 5% among biochemists. A biased metric on the other hand may favor some disciplines and penalize others.

\* Corresponding author. Tel.: +1 812 266 0621; fax: +1 812 855 0600.

E-mail address: [jakaur@umail.iu.edu](mailto:jakaur@umail.iu.edu) (J. Kaur).



**Fig. 1.** Effect of field normalization on ranking bias. We rank the top 5% of authors in seven JCR disciplines according to two metrics,  $h$  and  $h_s$  (see text). We compare the rank (top percentile) globally across disciplines versus locally within an author's own discipline. Due to discipline bias, biochemists are favored and mathematicians are penalized according to  $h$ , as illustrated by the two highlighted authors. The global ranking according to the normalized metric  $h_s$  is more consistent with the rankings within disciplines.

## 2. Methods

### 2.1. Data

We used the data collected by Scholarometer ([scholarometer.indiana.edu](http://scholarometer.indiana.edu)) from November 2009 to August 2012. Scholarometer is a social tool for scholarly services developed at Indiana University, with the goal of exploring the crowdsourcing approach for disciplinary annotations and cross-disciplinary impact metrics (Hoang, Kaur, & Menczer, 2010; Kaur et al., 2012). Users provide discipline annotations (tags) for queried authors, which in turn are used to compare author impact across disciplinary boundaries. The data collected by Scholarometer is available via an open API. We use this data to compute several impact metrics for authors belonging to various disciplines, and test the universality of these metrics. As of August 2012, the database had collected citation data about 38 thousand authors of 2.2 million articles in 1300 disciplines. Further statistics for authors and disciplines are available on the Scholarometer website (Kaur et al., 2012).

### 2.2. Impact metrics

The bibliometrics literature contains a plethora of scholarly impact metrics, and it is not feasible to evaluate all of them. Therefore we focus on a small set of metrics that are widely adopted and/or specifically designed to mitigate discipline bias. Our analysis of universality is performed on the following impact metrics:

- $c_{avg}$  is the average number of citations received by an author's articles.
- $h$  index** is defined as the maximum number of articles  $h$  such that each has received at least  $h$  citations (Hirsch, 2005). The  $h$  index is the most widely adopted impact metric. It summarizes the impact of a scholar's career using a single number without any threshold.
- Redner's index**  $c_{total}^{1/2}$  is defined as the square root of the total number of citations received by an author's articles (Redner, 2010).
- $h_m$  index** attempts to apportion citations fairly for papers with multiple authors (Schreiber, 2008). It counts the papers fractionally according to the number of authors. This yields an effective rank, which is utilized to define  $h_m$  as the maximum effective number of papers that have been cited  $h_m$  or more times.
- $g$  index** is the highest number  $g$  of papers that together receive  $g^2$  or more citations (Egghe, 2006). It attempts to mitigate the insensitivity of the  $h$  index to the number of citations received by highly cited papers.
- $i_{10}$  is proposed by Google and is defined as the number of articles with at least ten citations each (Google scholar citations open to all, 2011).
- $h_f$  index** was proposed as a universal variant of  $h$  (Radicchi, Fortunato, & Castellano, 2008). The number of citations  $c$  received by each paper is normalized by the average number of citations  $c_0$  for papers published in the same year and discipline. The rank of each paper  $n$  is rescaled by the average number  $n_0$  of papers per author written in the same year and discipline. The  $h_f$  index of the author is the maximum rescaled rank  $h_f$  such that each of the top  $h_f$  papers has at least  $h_f$  rescaled citations.

**Table 1**

Top JCR (ISI) disciplines. In this and the following tables, we display the average  $h$  index computed across authors in the same discipline.

	Discipline	Authors	( $h$ )
1.	Computer science, artificial intelligence	1922	15.96
2.	Biology	1147	19.66
3.	Economics	972	17.02
4.	Engineering, electrical & electronic	936	14.77
5.	Neurosciences	840	22.95
6.	Political science	794	15.81
7.	Psychology	774	21.18
8.	Biochemistry & molecular biology	766	22.37
9.	Sociology	749	16.70
10.	Mathematics	516	13.55
11.	Philosophy	501	13.63
12.	Information science & library science	480	11.15

**Batista's  $h_{i,norm}$**  involves normalizing the total number of citations in the  $h$ -core (the papers that contribute to the  $h$  index) by the total number of authors contributing to them. The resulting  $h_i$  of each author is then normalized by the average  $h_i$  of the author's discipline (Batista, Campitelli, & Kinouchi, 2006).

**New crown indicator  $(c/c_0)_{avg}$**  was proposed by Lundberg (2007) as the item oriented field-normalized citation score (FNCS) and implemented by Waltman, van Eck, van Leeuwen, Visser, and van Raan (2011). It is calculated as the average field-normalized number of citations  $c/c_0$  across an author's publications.

**$h_s$  index** is proposed here as a normalization of the  $h$  index by the average  $h$  of the authors in the same discipline. Numerical tests show that the distribution of  $h$  is not scale-free and therefore the mean is well defined. Despite its simplicity, we are not aware of this metric being previously defined in the literature. Note that within a discipline,  $h_s$  produces the same ranking as  $h$ . Therefore,  $h_s$  is very similar to the percentile score but slightly easier to compute. Percentiles have been proposed for normalization of journal impact factors (Leydesdorff & Bornmann, 2011).

### 2.3. Disciplines

To test the universality of the different impact metrics, we consider three distinct ways to define disciplines, i.e., to sample authors from multiple disciplines. When a user queries the Scholarometer system, she has to annotate the queried author with at least one discipline tag from the JCR science, social sciences, or arts & humanities indices. Additionally, the user may tag the author with any number of arbitrary (JCR or user-defined) discipline labels. Based on these annotations, we consider three disciplinary groupings of authors:

**ISI:** The 12 JCR disciplines with the most authors (see Table 1).

**User:** The top 10 user-defined disciplines (Table 2).

**Manual:** 11 manually constructed groups of related disciplines (Table 3).

In Section 4, we present results based on the ISI classification. In Section 4.1, we analyze the robustness of our results against the three disciplinary groupings of authors.

**Table 2**

Top user-defined disciplines.

	Discipline	Authors	( $h$ )
1.	Computer science	656	16.02
2.	Physics	200	18.66
3.	Computer networks	130	16.25
4.	Bioinformatics	125	16.50
5.	Engineering	115	11.46
6.	Medicine	104	23.47
7.	Chemistry	103	13.92
8.	Human computer interaction	94	17.72
9.	Computer science, security	82	19.32
10.	Image processing	80	18.39

**Table 3**  
Manually clustered disciplines.

	Manual label	Disciplines	Authors	(h)
1.	Computer science	Computer science, artificial intelligence Image processing Computer networks Computer science Computer science, theory & methods Computer science, software engineering Computer science, information systems Computer science, hardware & architecture Computer science, cybernetics	4342	15.79
2.	Biology	Plant sciences biology Zoology Plant sciences Evolutionary biology Entomology Biology Biodiversity conservation Biochemistry & molecular biology	2385	19.56
3.	Behavioral sciences	Sociology Psychology, social Psychology, applied Anthropology Psychology Behavioral sciences	1846	17.97
4.	Engineering	Engineering, mechanical Engineering, electrical & electronic Engineering, biomedical	1302	14.93
5.	Economics	Economics	972	17.02
6.	Mathematics	Statistics & probability Mathematics, applied Mathematics	860	15.53
7.	Political science	Public administration Political science	812	15.74
8.	Physics	Physics, applied Physics, multidisciplinary Physics, condensed matter Physics	675	19.63
9.	Business	Business, marketing Management Business, finance Business	665	15.59
10.	Education & educational research	Education technology Education & educational research	305	12.18
11.	Humanities, multidisciplinary	Humanities, multidisciplinary Humanities	122	9.00

### 3. Theory

An objective, quantitative assessment of metric universality is missing to date. To fill this void, we introduce a *universality index* to evaluate and compare the bias of different metrics. Our index allows for the first time to gauge a metric's capability to compare the impact of scholars across disciplinary boundaries, creating an opportunity for, say, mathematicians and biologists to be evaluated consistently.

The proposed universality index looks at how top authors according to a particular metric are allocated across different disciplines, and compares this distribution with one obtained from a random sampling process. This approach is inspired by a method for comparing expected and observed proportions of top cited papers to evaluate normalized citation counts (Radicchi et al., 2008). The idea is that each discipline should be equally represented in a sample of top authors. For example, if we rank scholars across all disciplines according to an unbiased (universal) metric, the top 5% of scholars should include the top 5% of mathematicians, the top 5% of biologists, and so on. In other words, the *percentage* of top scholars in each discipline should not depend on the size of the discipline. Of course the *number* of scholars in each discipline should be proportional to the size of that discipline.

Suppose each author is assigned a discipline  $d$ . For simplicity, let us assume that each author belongs to only one category. Selecting a fraction  $z$  of top scholars from the entire set according to a universal metric should be equivalent to sampling a fraction  $z$  of scholars at random. If we define  $f_{z,d}$  as the fraction of authors belonging to discipline  $d$  in the top  $z$ , the expected fraction is  $E[f_{z,d}] = z$ .

To understand the fluctuations in the numbers of authors from each category, consider a set of  $N$  authors in  $D$  categories. Let  $N_d$  be the number of authors in category  $d$ . Each author has a score calculated according to the rules of the particular

indicator we want to test. Imagine extracting the top fraction  $z$  of authors according to their scores. This list has  $n_z = \lfloor zN \rfloor$  authors. If the numerical indicator is fair, the selection of an author in category  $d$  should depend only on the category size  $N_d$ , and not on other features that may favor or hinder that particular category. Under these conditions, the number of authors  $n_d^z$  in category  $d$  that are part of the top  $z$  is a random variate obeying the hypergeometric distribution (Radicchi & Castellano, 2012):

$$P(n_d^z | n_z, N, N_d) = \binom{N_d}{n_d^z} \binom{N - N_d}{n_z - n_d^z} / \binom{N}{n_z} \tag{1}$$

where  $\binom{x}{y} = (x! / (y!(x - y)!))$  is a binomial coefficient that calculates the total number of ways in which  $y$  elements can be extracted out of  $x$  total elements. Eq. (1) describes a simple urn model (Mahmoud, 2008), where elements (authors in our case) are randomly extracted from the urn without replacement. Such a random process provides us with a null model for the values of  $f_{z,d}$ .

In Section 4 we estimate confidence intervals by simulating  $10^3$  times the null model leading to Eq. (1).

To obtain a quantitative criterion for the universality of  $m$  with respect to a set of  $D$  disciplines and a fraction  $z$  of top scholars, we compute the universality of metric  $m$  as

$$u_m(z) = 1 - \frac{1}{D} \sum_{d=1}^D \left| \frac{f_{z,d}^m}{z} - 1 \right|^\alpha$$

where the parameter  $\alpha$  tunes the relative importance given to small versus large deviations from the expected fractions. In Section 4 we use  $\alpha = 1$  and 2. If  $u_m(z)$  is high (close to one), the proportion of top scholars from each discipline is close to  $z$ , and therefore the impact measure  $m$  is able to compensate for discipline bias. This definition of universality satisfies the basic intuition that all metrics are unbiased in the limit  $z = 1$ .

Note that  $u_m(z) \leq 1$ ; it can take negative values in contrived biased scenarios. An alternative definition would normalize the deviations from the expected fractions by the variance within each discipline, however this approach would have decreasing universality as  $z \rightarrow 1$  due to the increasing variance. This would violate our basic intuition that all metrics are unbiased in the limit  $z = 1$ .

To eliminate the dependence of the universality assessment on a particular selectivity  $z$ , we can finally define the universality index of  $m$ :

$$\bar{u}_m = \int_0^1 u_m(z) dz.$$

We numerically approximate the integral as:

$$\bar{u}_m \simeq \sum_{q=1}^{99} u_m(q \cdot \Delta z) \Delta z,$$

where we set  $\Delta z = 0.01$ .

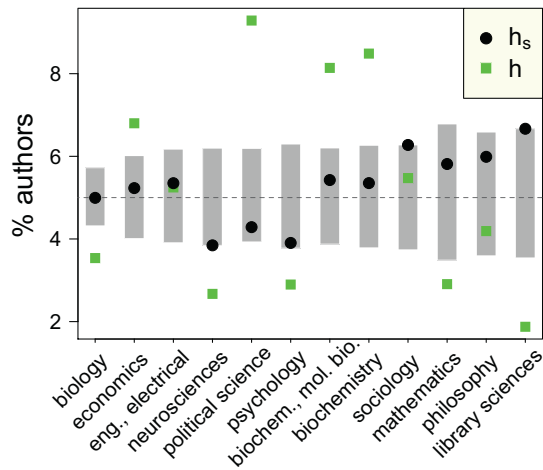
### 4. Results

To illustrate the usefulness of our index, let us analyze the universality for the ten impact metrics described in Section 2.2 across a set of scholarly disciplines. As evident in Fig. 5, some metrics are more universal than others. We first consider the disciplines from the Thomson–Reuters JCR classification (see Table 1) for the case  $\alpha = 1$ . To better appreciate the different biases, let us focus on just two impact metrics,  $h$  and  $h_s$  (Fig. 2). When we select the top 5% of all scholars,  $h_s$  yields close to 5% of scholars from each of the considered disciplines, consistently with the null model (gray area);  $h$  yields large fluctuations, favoring some disciplines and penalizing others.

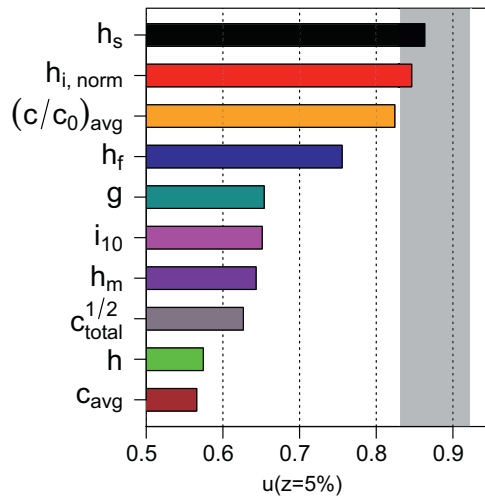
Fig. 3 shows that according to  $u(5\%)$ , two of the metrics appear to be least biased: Batista’s  $h_{i,norm}$  and our own  $h_s$ . These are consistent with the unbiased model at  $z = 5\%$ , while the other metrics are not.

Fig. 4(a) shows how the universality of each metric depends on the selectivity  $z$ . As we select more top scholars, the bias of all metrics decreases;  $u(z) \rightarrow 1$  as  $z \rightarrow 1$  by definition. For selectivity  $z < 40\%$ , the two best metrics display high universality, as illustrated by the overlap of the corresponding curves with the expectation of the null model (gray area).

Table 4 reports the values of the universality index  $\bar{u}$  integrated across  $z$ . To evaluate the statistical significance of differences in values of the universality index  $\bar{u}$  for different metrics, we need to estimate the fluctuations of this measure. Let us consider the variations in the values of  $\bar{u}_{null}$  obtained by simulating the null model for  $z \in (0, 1)$ . Running  $10^3$  simulations yields a standard deviation  $\sigma_{null} = 0.005$ . Therefore we do not consider differences in the third decimal digit statistically significant, and we round  $\bar{u}$  values to the second decimal digit. The differences shown are deemed significant. According to this summary,  $h_{i,norm}$  and  $h_s$  are the most universal among the impact metrics considered. Their universality indices are



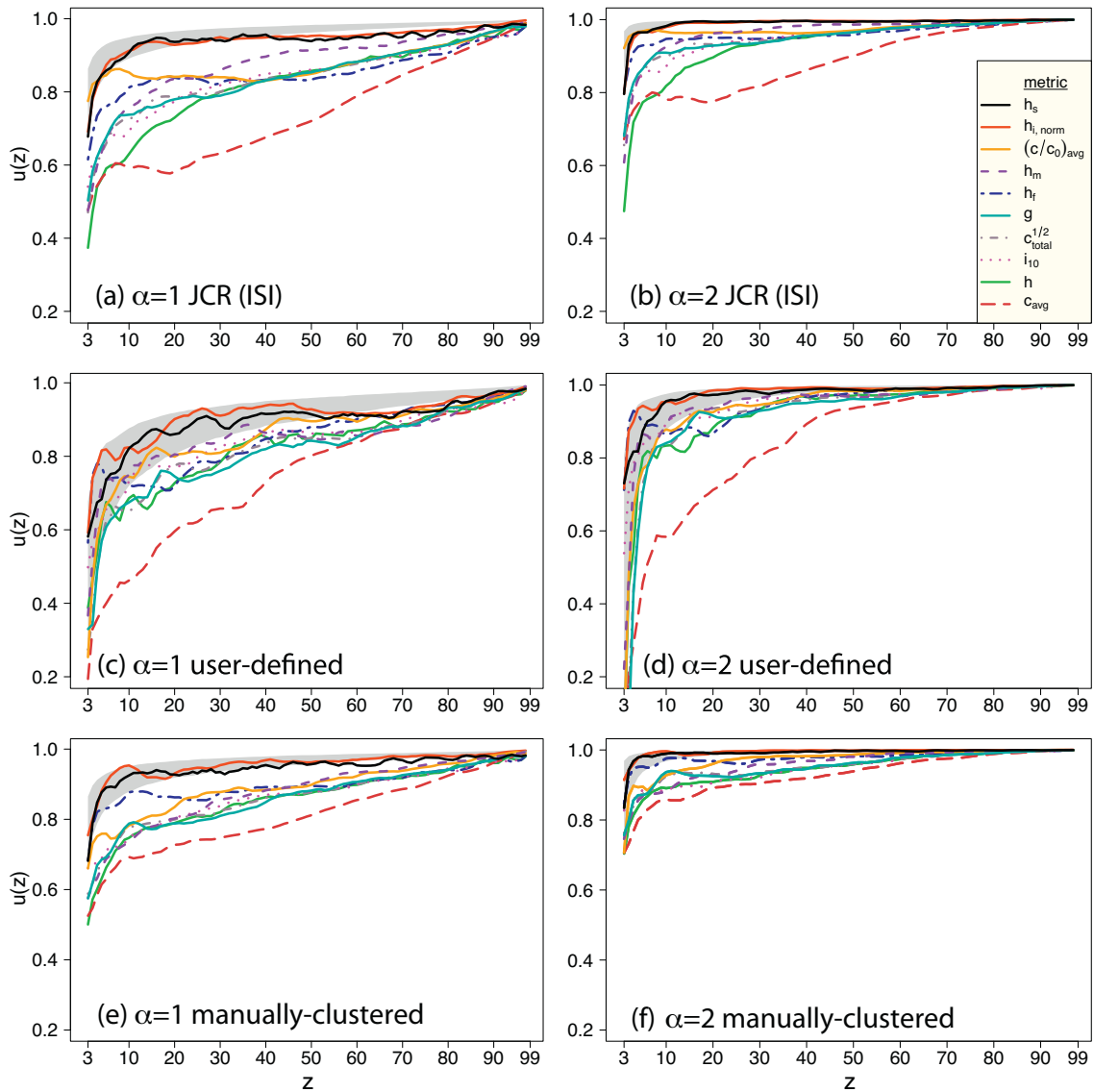
**Fig. 2.** Illustration of discipline bias. The analysis is based on empirical data from the Scholarometer system, which provides discipline annotations for scholars and associated citation material (Kaur et al., 2012). For legibility we display here only two impact metrics ( $h$  and  $h_s$ ) that are used to rank authors in the twelve top JCR disciplines spanning science, social sciences, and arts & humanities. Across these disciplines, we select the top 5% of authors according to each metric. We then measure the percentage of authors from this selection that belong to each discipline. The  $h$  index favors certain disciplines (e.g., political science) and penalizes others (e.g., library sciences). In this and the following plots, gray areas represent 90% confidence interval of unbiased samples, as discussed in Section 3.



**Fig. 3.** Universality  $u(z)$  for ten impact metrics and selectivity  $z = 5\%$ .

**Table 4**  
Universality index of the ten metrics for different discipline categorizations and exponent values.

Metric	JCR (ISI)		User-defined		Manual	
	$\alpha = 1$	$\alpha = 2$	$\alpha = 1$	$\alpha = 2$	$\alpha = 1$	$\alpha = 2$
$h_s$	0.94	0.99	0.90	0.98	0.95	0.99
$h_{i,norm}$	0.94	0.99	0.90	0.98	0.95	0.99
$(c/c_0)_{avg}$	0.88	0.98	0.86	0.95	0.89	0.97
$h_m$	0.88	0.97	0.85	0.95	0.88	0.96
$h_f$	0.86	0.96	0.85	0.95	0.90	0.98
$g$	0.85	0.95	0.81	0.92	0.86	0.96
$c^{1/2}_{total}$	0.85	0.95	0.82	0.93	0.87	0.96
$i_{10}$	0.84	0.95	0.85	0.95	0.87	0.95
$h$	0.83	0.94	0.82	0.93	0.86	0.95
$c_{avg}$	0.74	0.89	0.74	0.85	0.81	0.93



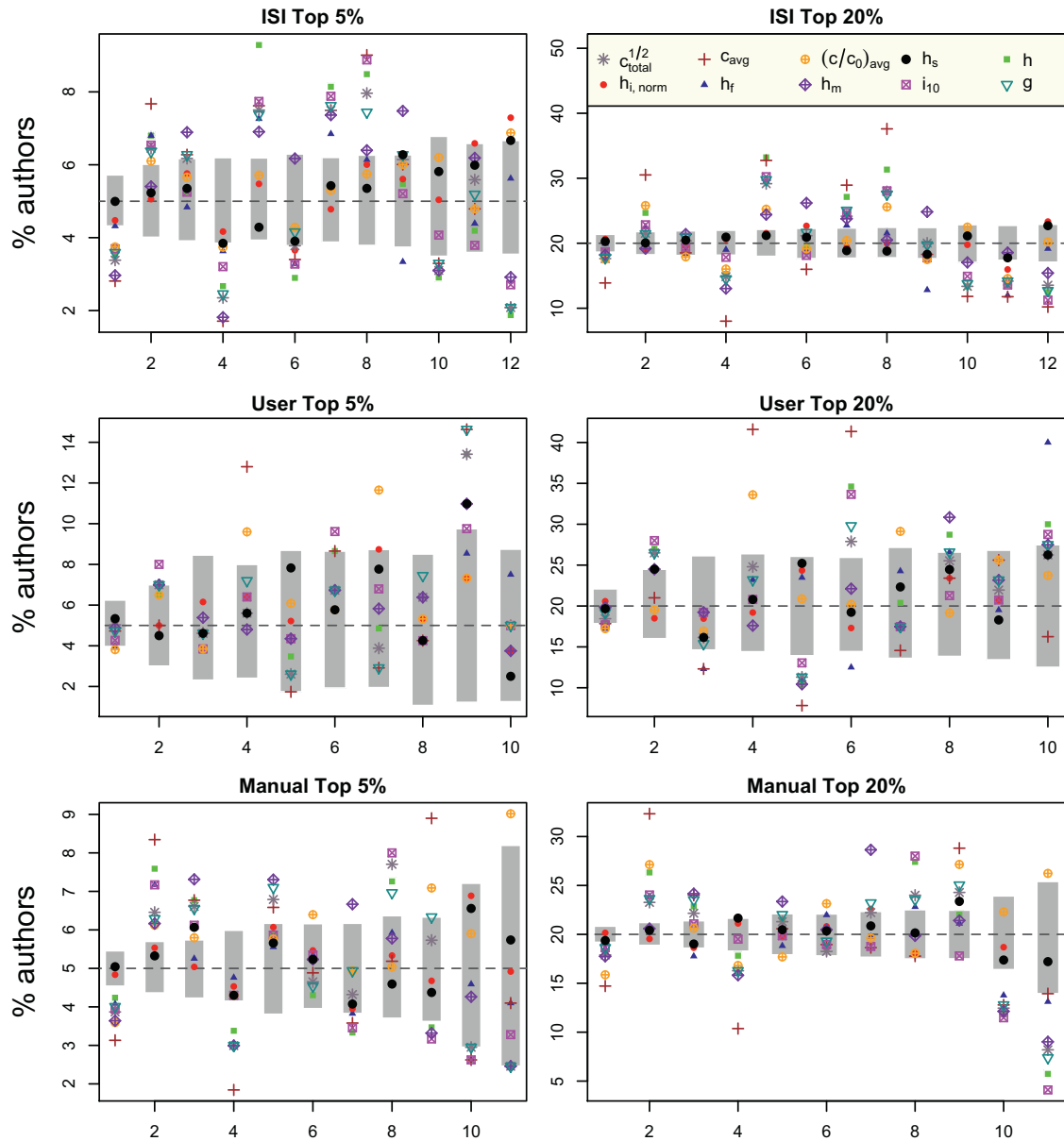
**Fig. 4.** Universality  $u(z)$  as a function of selectivity  $z$ . Different panels correspond to different categorizations and values for the exponent  $\alpha$ . Gray areas in the figure display 90% confidence intervals computed through the null model. The rankings are not particularly sensitive to different categorizations or exponent values (cf. Table 4).

statistically equivalent to each other. The computation of  $h_s$  is however much simpler, as it does not require co-author metadata.

Next we test the robustness of our findings with respect to several variations of our method: different ways to classify authors into disciplines, different selectivity values, and different exponents in the definition of universality.

#### 4.1. Sensitivity to discipline definitions

While our definition of universality assumes that authors are associated with disciplines, the results of our analysis are not dependent on the JCR classification. Fig. 5 extends Fig. 2 to the two additional discipline definitions (User and Manual, cf. Section 2.3). The results in all cases are similar. Fig. 4 and Table 4 show that with a few exceptions, the ranking of impact metrics is consistent across categorizations. In all cases,  $h_s$  and  $h_{i, norm}$  are the most universal (least biased) metrics.



**Fig. 5.** Percentage of authors belonging to different disciplines according to ISI/JCR (top), user-defined (middle), and manually clustered (bottom) disciplines listed in Section 2.3. The authors are ranked by each metric in the top  $z=5\%$  (left) and 20% (right). Gray areas bound the 90% confidence intervals obtained from the null model.

#### 4.2. Sensitivity to selectivity $z$

We repeated the analysis of Fig. 2 for two values of the selectivity parameter  $z$ . Fig. 5 shows that for each discipline categorization, the results of the cases  $z=0.05$  and  $z=0.20$  are statistically similar; the number of times that the measured values are inside the confidence intervals is not strictly depending on the choice of  $z$ .

#### 4.3. Sensitivity to exponent $\alpha$

Fig. 4 and Table 4 show that, with a few exceptions, the ranking of impact metrics is consistent for different exponents  $\alpha$ .

### 5. Conclusion

While discipline bias is quickly being recognized as a key challenge for objective assessment of impact, it has been problematic until now to evaluate the claims of universality for the multitude of proposed metrics. The index presented



here is the first quantitative gauge of universality that can be readily applied to any existing metric. The present analysis points to  $h_s$  as an impact metric that is intuitive, easy to compute, and universal.

The  $h_s$  metric does require that the disciplines associated with an author be known, something that can be a challenge because discipline boundaries are not sharp (Moed, 2010) and they are continually evolving as new fields emerge and old ones die (Sun, Kaur, Milojević, Flammini, & Menczer, 2013). The solution we have proposed and implemented in the Scholarometer system (Kaur et al., 2012) is that of crowdsourcing the discipline annotations of scholars. In this view, annotations are votes and a scholar is represented as a vector of disciplines. One may then compute the impact of interdisciplinary scholars with respect to any relevant discipline, or a combined metric based on the discipline weights in their vector representations. Further work is needed to verify that such a combination of universal metrics remains universal.

## Acknowledgements

Thanks to Santo Fortunato, Alessandro Flammini, Claudio Castellano, and Yong-Yeol Ahn for helpful feedback on earlier versions of this manuscript. Xiaoling Sun, Diep Thi Hoang, Mohsen Jafari Asbagh, and Lino Possamai provided technical support. We acknowledge financial support from the IU School of informatics and Computing, the Lilly Endowment, and NSF (award IIS-0811994) for funding the computing infrastructure that hosts the Scholarometer service.

## References

- Abbott, A., Cyranoski, D., Jones, N., Maher, B., Schiermeier, Q., & Van Noorden, R. (2010). Do metrics matter? *Nature*, 465(7300), 860–862.
- Batista, P., Campiteli, M., & Kinouchi, O. (2006). Is it possible to compare researchers with different scientific interests? *Scientometrics*, 68(1), 179–189.
- Bollen, J., Van de Sompel, H., Hagberg, A., & Chute, R. (2009). A principal component analysis of 39 scientific impact measures. *PLoS ONE*, 4(6), e6022.
- Bornmann, L., & Daniel, H. (2006). Selecting scientific excellence through committee peer review – A citation analysis of publications previously published to approval or rejection of post-doctoral research fellowship applicants. *Scientometrics*, 68(3), 427–440.
- Colwell, R., Blouw, M., Butler, L., Cozzens, S. E., Feller, I., Gingras, Y., et al. (2012). Informing research choices: Indicators and judgment. The expert panel on science performance and research funding. Council of Canadian Academies. Tech. rep., [http://www.scienceadvice.ca/uploads/eng/assessments%20and%20publications%20and%20news%20releases/science%20performance/scienceperformance\\_fullreport\\_en\\_web.pdf](http://www.scienceadvice.ca/uploads/eng/assessments%20and%20publications%20and%20news%20releases/science%20performance/scienceperformance_fullreport_en_web.pdf)
- Davis, P., & Papanek, G. (1984). Faculty ratings of major economics departments by citations. *The American Economic Review*, 74(1), 225–230.
- Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics*, 69(1), 131–152.
- Garfield, E. (2006). The history and meaning of the journal impact factor. *JAMA: The Journal of the American Medical Association*, 295(1), 90–93.
- Google scholar citations open to all <http://googlescholar.blogspot.com/2011/11/google-scholar-citations-open-to-all.html>
- Hirsch, J. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), 16569.
- Hoang, D. T., Kaur, J., & Menczer, F. (2010). Crowdsourcing scholarly data. In *Proc. web science conference: extending the frontiers of society on-line (WebSci)* <http://journal.webscience.org/321/>
- Kaur, J., Hoang, D. T., Sun, X., Possamai, L., JafariAsbagh, M., Patil, S., & Menczer, F. (2012). Scholarometer A social framework for analyzing impact across disciplines. *PLoS ONE*, 7(9), e43235. <http://dx.doi.org/10.1371/journal.pone.0043235>
- King, D. (2004). The scientific impact of nations. *Nature*, 430(6997), 311–316.
- Kinney, A. (2007). National scientific facilities and their science impact on nonbiomedical research. *Proceedings of the National Academy of Sciences of the United States of America*, 104(46), 17943–17947.
- Lane, J. (2010). Let's make science metrics more scientific. *Nature*, 464(7288), 488–489.
- Lehmann, S., Jackson, A., & Lautrup, B. (2006). Measures for measures. *Nature*, 444(7122), 1003–1004.
- Leydesdorff, L., & Bornmann, L. (2011). Integrated impact indicators compared with impact factors: An alternative research design with policy implications. *Journal of the American Society for Information Science and Technology*, 62(11), 2133–2146.
- Lundberg, J. (2007). Lifting the crown—citation z-score. *Journal of Informetrics*, 1(2), 145–154.
- Mahmoud, H. (2008). *Pólya urn models*, Vol. 76. Boca Raton, FL: Chapman & Hall/CRC.
- Moed, H. F. (2010). Measuring contextual citation impact of scientific journals. *Journal of Informetrics*, 4(3), 265–277.
- Radicchi, F., & Castellano, C. (2012). A reverse engineering approach to the suppression of citation biases reveals universal properties of citation distributions. *PLoS ONE*, 7(3), e33833. <http://dx.doi.org/10.1371/journal.pone.0043235>
- Radicchi, F., & Castellano, C. (2012). Testing the fairness of citation indicators for comparison across scientific domains: The case of fractional citation counts. *Journal of Informetrics*, 6(1), 121–130.
- Radicchi, F., Fortunato, S., & Castellano, C. (2008). Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences of the United States of America*, 105(45), 17268.
- Redner, S. (2010). On the meaning of the h-index. *Journal of Statistical Mechanics: Theory and Experiment*, 2010, L03005.
- Schreiber, M. (2008). To share the fame in a fair way,  $h_m$  modifies  $h$  for multi-authored manuscripts. *New Journal of Physics*, 10, 040201.
- Sun, X., Kaur, J., Milojević, S., Flammini, A., & Menczer, F. (2013). Social dynamics of science. *Scientific Reports*, 3, 1069, <http://dx.doi.org/10.1038/srep01069>
- Van Noorden, R. (2010). Metrics: A profusion of measures. *Nature*, 465(7300), 864–866.
- Waltman, L., van Eck, N., van Leeuwen, T., Visser, M., & van Raan, A. (2011). Towards a new crown indicator: Some theoretical considerations. *Journal of Informetrics*, 5(1), 37–47.