The *h*-index in Australian Astronomy

Kevin A. Pimbblet

School of Physics, Monash University, Clayton, VIC 3800, Australia.
Email: kevin.pimbblet@monash.edu

Abstract: The Hirsch *h*-index is now widely used as a metric to compare individual researchers. To evaluate it in the context of Australian astronomy, the *h*-index for every member of the Astronomical Society of Australia (ASA) is found using NASA's Astrophysics Data System Bibliographic Services. Percentiles of the *h*-index distribution are detailed for a variety of categories of ASA members, including students. This enables a list of the top ten Australian researchers by *h*-index to be produced. These top researchers have *h*-index values in the range $53 < h < 77$, which is less than that recently reported for the American Astronomical Society membership. We suggest that membership of extremely large consortia such as the Sloan Digital Sky Survey may partially explain the difference. We further suggest that many student ASA members with large *h*-index values have probably already received their Ph.D. and need to upgrade their ASA membership status. To attempt to specify the *h*-index distribution relative to opportunity, we also detail the percentiles of its distribution by years since Ph.D. award date. This shows a steady increase in *h*-index with seniority, as can be expected.

Keywords: sociology of astronomy — publications, bibliography — astronomical databases: miscellaneous

1 Introduction

The modern research academic is judged as never before: a large variety of metrics are now employed to determine the worth and merit of researchers, particularly when it comes to hiring. Anecdotally, one of the chief metrics used is the Hirsch index (*h*-index; Hirsch 2005). The *h*-index is formally defined as follows: 'A scientist has index *h* if *h* of his or her N_p papers have at least *h* citations each and the other $(N_p - h)$ papers have $\leq h$ citations each' (Hirsch 2005). Its modest simplicity is probably a prime factor in its rapid pick-up by major publishers (Anon. 2005; Ball 2005). Moreover, this index is particularly useful as it has superior predictive power (in terms of productivity) for the future of researchers compared to the total number of career citations, career publications and mean citations per paper (Hirsch 2007). Although other metrics and analyses exist (cf. Pearce 2004; Kurtz et al. 2005; Egghe 2006; Jin 2006; Kosmulski 2006; Blustin 2007; Jin et al. 2007; Bornmann, Mutz & Daniel 2008; Wu 2010; Zyczkowski 2010), the *h*-index remains as the most prominent of its class in the field.

Recently, Conti et al. (2011) presented work on the astronomer's H-R diagram (number of Google search results versus citations and *h*-index) for members of the American Astronomical Society (AAS). Contained within that presentation are a number of interesting concepts: a top-ten list of AAS members by *h*-index (spanning the range $94 < h < 118$) and the *h*-indices of all AAS members. This work is motivated by the Conti et al. (2011) presentation and seeks to determine the typical range of the *h*-index in Australian astronomy, which may be of use for future employers and employees in the community. The format of this work is as follows. In Section 2, we give an overview of the dataset that we use: the membership of the Astronomical Society of Australia. In Section 3 we determine percentiles of the *h*-index distribution for a variety of ASA membership categories, including students. To attempt to normalize relative to opportunity, we re-evaluate the *h*-index distribution as a function of time elapsed since Ph.D. award date in Section 4. Our conclusions are presented in Section 5.

2 Data

To determine the *h*-indices of Australian astronomers, we make use of the Astronomical Society of Australia (ASA) membership list. The membership list is a fair representation of the Australian astronomical community: the majority of professional astronomers are members. Membership of the ASA comes in several different categories, each of which we indicate with a single letter as detailed in Table 1. The advantage of the ASA membership list is that we can distinguish between different grades of members (i.e. amateurs and professional astronomers who actively publish) to better probe the *h*-index in these sub-categories.

For each ASA member, we then implement a search in NASA's Astrophysics Data System (ADS) to return a list of all refereed publications. We then sort this list

Table 1. Categories of ASA membership, showing the number of members (*N*) in each category and the fraction of each that were excluded (*X*) from the subsequent analysis due to name confusion (see Section 2). We ignore the extra categories of associate members (i.e. members of other learned societies who are likely not to possess a Ph.D. in astronomy) and corporate members. It is important to note that individual researchers can belong to multiple categories (e.g. retired, overseas fellows)

Category	*N*	*X*	Notes
M	262	0.10	Member: Full professional member with a Ph.D. in astronomy or related discipline
F	81	0.10	Fellow: Senior members of the community with potentially decades of experience
S	164	0.09	Student: Postgraduate students studying toward a Ph.D. in astronomy
H	15	0.00	Honorary: Elected by the ASA council for distinguished contributions
R	54	0.11	Retired
O	56	0.10	Overseas
A	14	0.00	Associate Members: Educators, communicators and amateur astronomers lacking a Ph.D.

according to citations to determine the *h*-index for each ASA member. We note for posterity that these searches were implemented on 24th–25th January 2011 and were correct on a best-efforts basis as of said date range.

A big issue in this methodology is attempting to tie down each individual to unique entries in ADS. Although the present author is blessed with a very rare surname, others in the community are not. For more common surnames, we use the first name and the middle initials to help determine the *h*-index of specific researchers, including attempting common substitutions for first names (e.g. 'Bill' for 'William'). However, for the very common surnames (e.g. Smith), this is not always possible. Therefore, the subsequent analysis in this work does not include any names for which we could not adequately differentiate a single individual in the literature in a reasonable amount of time. This affects ~10% of the membership list and the affected portion of each category is labelled *X* in Table 1. We caution that the subsequent analysis should therefore be regarded as incomplete: the inclusion of these names could increase or decrease the relative rankings of individuals within the ASA membership. We also note that we make no attempt to exclude self-citations in our analysis (e.g. Pimbblet 2011). Finally, it may be the case that some of the categories may not be up-to-date due to (for example) student members gaining their Ph.D. and either not upgrading to full membership status immediately or the list itself not being updated immediately.

3 *h*-index by ASA Membership Category

This simplest point of departure for the *h*-index analysis is to pull out the top ten: those people who could rightly be called academic giants in their own right in the community. To do this, we simply rank all professional members who are not based overseas (i.e. categories M + F + S + H + R; Table 1). This top ten list is presented in Table 2.

Although we refrain from commenting on individuals in this list, it is instructive to compare it to the list of Conti et al. (2011). The *h*-index values for the top ten AAS members are much higher than for the Australians ($94 < h < 118$ versus $53 < h < 77$). Examination of Conti et al.'s

Table 2. Top ten by *h*-index for Australian astronomy, excluding overseas professionals

Rank	Name	*h*
=1	Ken FREEMAN	77
=1	Jeremy MOULD	77
3	Karl GLAZEBROOK	71
4	Dick MANCHESTER	68
5	Michael DOPITA	64
6	Warrick COUCH	61
7	Matthew COLLESS	56
8	Brian SCHMIDT	54
=9	Mike BESSELL	53
=9	Joss BLAND-HAWTHORN	53

(2011) figures suggests that membership of very large observational programmes such as the Sloan Digital Sky Survey (SDSS; e.g. Abazajian et al. 2009) can boost researcher's *h*-index above mean values. It is certainly the case that the Australian top ten is dominated by non-SDSS professionals and we therefore suggest that most of the difference seen between the two samples could be due to this effect. Indeed, seven of the AAS's top ten (Conti et al. 2011) are contained in the author list of Abazajian et al. (2009). However, we do note that the Australian top ten does contain a number of members of other consortia (not as large or extensive as SDSS) such as the 2dF Galaxy Redshift Survey (e.g. Colless et al. 2001). Moreover, the majority of the listed researchers in Table 2 also feature in Thomson Reuters ISI's highly cited list[1] for space science.

But what of the rest of the community? In Figure 1 we display a histogram of *h*-index for all ASA members. This graph is dominated by those members having a zero *h*-index or slightly above, much as the AAS community is (Conti et al. 2011). The vast majority of these members are student members, many of whom are likely not to have published. Even if they have published, the duration of the Ph.D. may mean that sufficient time has not elapsed to gain large numbers of citations and that only the very exceptional papers produced by students garner large number of citations immediately. Clearly students in

---

[1] http://www.isihighlycited.com.

present-day collaborations such as WiggleZ (Drinkwater et al. 2010) will benefit from this effect in much the same way that SDSS members receive a boost.

To analyse the content of Figure 1 in a more in-depth manner, we now create sub-samples of the ASA membership according to grade and determine various percentiles of the h-index distribution. These percentiles are presented in Table 3. We do not present results for the individual categories H, O, R and A due to low numbers. This can be seen in the relatively tiny difference between the percentiles quoted for $M + F + R - O$ versus $M + F - R - O$ samples in Table 3.

We start by discussing the student membership result. At the upper echelons, students appear to have an h-index
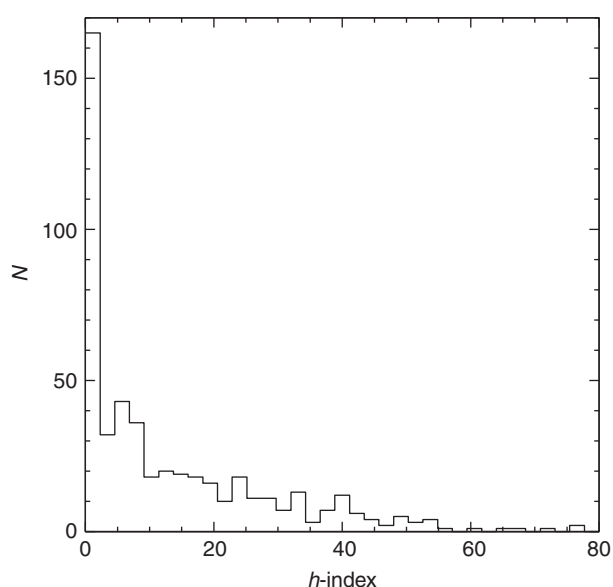
Figure 1 Histogram of h-index for all ASA members, excluding those for whom name confusion could not be resolved.

comparable to that of junior professionals. But a careful analysis of the membership list reveals that this is exactly what these students are: junior professionals who should be in the M category. We argue that anything above the 90th percentile for the S category should be regarded with suspicion.

Naturally, the fellows occupy much higher h-index values than the regular members do. The effect of adding or removing the retirees from the $M + F$ sample is modest: the most noticeable effect is at the upper echelons of the scale. However, the major problem of this analysis is that it does not specify the h-index relative to opportunity. To remedy this, we now try to divide up the ASA membership according to years since the award of a Ph.D.

4 h-index by Years Since Ph.D. Award

Even the award year of a Ph.D. must be regarded with healthy suspicion as a metric for performance relative to opportunity. This is especially true for early-career researchers who may complete their Ph.D. while undertaking their first post-doctoral placement and for the many researchers who have had significant time away from the profession (the present author included).

To determine the award date of the Ph.D., we use results from ADS where available. If the Ph.D. is not listed in ADS, then we use the date of the second first-author refereed publication by the member as a compromise proxy for this date, given the distribution of the S sample in Table 3. This date was determined for all ASA members in the $M + F + R - O$ category. Where no date could be determined by either method, the member was simply removed from the list. This may have the effect of meaning that the percentiles for this sample are upper limits as we have missed doctoral researchers who have few first author publications. We present the percentiles of this distribution in Table 4. The results show a fairly

Table 3. Percentiles of h-index distribution by ASA membership grade. N gives the total number of members for each row

Category	N	Percentile						
		25	50	75	90	95	97.5	99
M	235	6.0	12.0	21.0	32.5	43.0	47.8	52.4
F	73	21.0	30.0	39.8	50.0	59.2	69.8	75.4
S	149	0.0	0.0	2.0	4.0	5.0	6.7	9.5
$M + F - R - O$	227	7.0	15.0	28.0	41.0	45.0	53.0	63.7
$M + F + R - O$	266	6.0	15.0	28.0	40.0	45.4	53.3	68.8
$M + F + S - R - O$	374	1.0	5.0	19.0	33.6	42.3	46.7	59.7

Table 4. Percentiles of h-index distribution by years since Ph.D. award (explicitly: 2011 minus the award date). Only members in $M + R + F - O$ categories are included. N gives the total number of members for each row

Years since Ph.D. award	N	Percentile						
		25	50	75	90	95	97.5	99
0–5	32	5.0	6.0	9.0	10.0	13.2	15.8	16.3
6–10	42	8.5	14.0	19.0	22.0	25.0	27.8	29.7
11–20	74	12.5	19.0	28.5	39.6	43.4	53.1	58.4
>20	94	12.0	25.0	38.0	47.2	55.1	62.9	77.0

steady progression as one increases in seniority from Ph.D. award date without any obvious discrepancies, as may be expected. However, one comment to be made is that there seem to be many fewer young professionals in the samples than there perhaps should be (given the numbers in more senior years). This tentatively suggests that new recruits to Australian astronomy may not be joining the ASA immediately.

Further, not all areas and sub-disciplines of science and astronomy may be equal. Those researchers involved in (for example) instrumentation may have a very different *h*-index distribution to those researching observational cosmology (particularly those in larger-sized consortia).

5 Conclusions

This work has presented an analysis of the *h*-index distributions for present members of the ASA. As well as deriving a top ten (Table 2), we have presented the percentiles for various sub-samples of the ASA's membership, including student statistics (Table 3). We have also attempted to analyse the distribution relative to opportunity by detailing the percentiles by time elapsed since Ph.D. award date (Table 4).

Clearly the *h*-index is a crude estimator of the value of a researcher and should not be used in isolation from other metrics, even if it is a good predictor of future productivity (Hirsch 2007). It will be instructive to re-visit this analysis in future years or decades to determine how the field has changed.

We terminate this work with a *caveat emptor*: there are known deficiencies in this analysis, such as numerous missing persons (who are not ASA members) whose statistics may alter the results presented. We have tried to be up-front with various caveats throughout this work, but there may yet be unknown unknowns present as well. Further, there may exist transcription errors that went undetected during the data assembly stage. However, as far as possible, we believe the numbers quoted in this work are accurate.

References

Abazajian, K. N. et al., 2009, ApJS, 182, 543
Anon., 2005, Science, 309, 1181
Ball, P., 2005, Nature, 436, 900
Blustin, A., 2007, A&G, 48, 6.32
Bornmann, L., Mutz, R. & Daniel, H.-D., 2008, J. Am. Soc. Inf. Sci. Technol., 59, 830
Colless, M. et al., 2001, MNRAS, 328, 1039
Conti, A., Lowe, S., Accomazzi, A. & DiMilia, G., 2011, AAS, 217, #145.01 (see `http://dl.dropbox.com/u/473509/Astro-HR.pdf`)
Drinkwater, M. J. et al., 2010, MNRAS, 401, 1429
Egghe, L., 2006, Scientometr., 69, 131
Hirsch, J. E., 2005, PNAS, 102, 16569
Hirsch, J. E., 2007, PNAS, 104, 19193
Jin, B., 2006, Sci. Focus, 1, 8
Jin, B., Liang, L., Rousseau, R. & Egghe, L., 2007, Chin. Sci. Bull., 52, 855
Kosmulski, M., 2006, ISSIN, 2, 4
Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C. S., Demleitner, M., Murray, S. S., Martimbeau, N. & Elwell, B., 2005, J. Am. Soc. Inf. Sci. Technol., 56, 111
Pearce, F., 2004, A&G, 45, 2.15
Pimbblet, K. A., 2011, MNRAS, 411, 2637
Wu, Q., 2010, J. Am. Soc. Inf. Sci. Technol., 61, 609
Zyczkowski, K., 2010, Scientometr., 85, 301