Noname manuscript No.
(will be inserted by the editor)

P-score: a Reputation Bibliographic Index that Complements Citation Counts

João Mateus de Freitas Veneroso · Marlon Dias · Berthier Ribeiro-Neto · Nivio Ziviani · Edmundo de Souza e Silva

the date of receipt and acceptance should be inserted later

Conflict of Interest: The authors declare that they have no conflict of interest.

Abstract The notions of reputation and popularity in academia are critical for taking decisions on research grants, faculty position tenure, and research excellence awards. These notions are almost always associated with the publication track records of the researchers. Thus, it is important to assess publication track records quantitatively. To quantify a publication record, bibliographic indices are usually adopted. Among these, citation-based indices such as the H-index are frequently considered. In this paper we study the correlation between P-score, a publication record index and H-index, the most popular citation-based index. While H-indices reflect the popularity of a given publication or researcher in academia, P-scores reflect the reputation of a publication or researcher among its peers. Popularity and reputation are frequently considered to be equivalent properties, however they are not identical. Indeed, we first show that H-indices and P-scores are correlated with a Kendall-Tau coefficient that exceeds 0.5. However, we also notice that they have important differences. Particularly, we identify publication venues with high H-indices and low P-scores, as well as venues with low H-indices and high P-scores. We provide interpretations for these findings and discuss

João Mateus de Freitas Veneroso
Universidade Federal de Minas Gerais, Av. Pres. Antônio Carlos, 6627 - Pampulha, Belo Horizonte - MG, 31270-901, Departamento de Ciência da Computação, Sala 4304
Tel.: +55 (31) 99436-0909
E-mail: jmfveneroso@gmail.com

Marlon Dias
Universidade Federal de Minas Gerais

Berthier Ribeiro-Neto
Universidade Federal de Minas Gerais

Nivio Ziviani
Universidade Federal de Minas Gerais

Edmundo de Souza e Silva
Universidade Federal do Rio de Janeiro

how they can be used by research funding councils and committees to better support their funding decisions.

Keywords P-score · H-index · Bibliographic Index · Reputation Flows

1 Introduction

Academic research evaluation is a topic of interest to universities, research groups, research funding institutions, and the public at large. It is used to decide laboratory space allocation, research grants, tenure track, university choice by students, and more. It also has a direct impact on the reputation of the university or institution, which affects the ability of the institution to attract the best possible talent. While evaluation of academic research is complex, usually a function of multiple variables, a dominant variable in most evaluation procedures is the research output of the institution. In particular, considerable emphasis is given to the papers published by the professors and students of the institution, as well as the quality of these papers in terms of their impact on scientific knowledge. A rather popular approach to evaluate publication impact is to compare journals, conferences or papers through the usage of academic impact indicators. A popular and moderately effective approach is to compute an index based on the number of citations received by a given piece of research and assume that the number of citations is a good proxy for research quality. However, this approach has multiple shortcomings.

First, in the evaluation and bibliometrics research communities, citations are understood as a measure of attention rather than a proxy for impact or quality. Citations measure the attention to a paper of peers in related fields [24], not the quality of the work produced. Second, citations can take a long time to happen. To illustrate this, in a recent study that looked at first time to citation in a universe of more than a million papers, half the papers only received their first citation 20 months or more after they were published [26]. Third, citations are not simple to compute and are not always broadly available, particularly at the level of individual researchers.

Despite these limitations, citation-based indices continue to be a popular approach to academic research evaluation [20]. Nonetheless, it is desirable to employ complementary indicators that tackle the exposed problems without loss of effectiveness.

Academic reputation is an individual or group property strongly associated with academic impact, and while being similar it is not identical to academic popularity. An actor's popularity depends solely on the the total number of endorsements the actor receives from other actors, while its reputation or prestige depends on the reputations of the endorsing actors [6]. Reputable venues tend to concentrate the most relevant research papers because that is how they acquire and keep their good reputations. At the same time, reputable authors seek to publish on reputable venues because it gives their research more visibility and accreditation. Therefore, researchers convey reputation to a venue proportionally to their own reputation and the reputation of researchers is proportional to the reputation of the venues where they publish. This relationship between authors and venues constitutes a reputation network in which authors influence venue reputations and vice versa.

P-score [33] is a graph-based modeling index that attributes quantitative reputations to venues based on the publication patterns of a reference group of researchers, without relying on citation information. It deals with the three problems discussed previously.

First, instead of measuring the amount of researcher attention, P-score measures reputations which, under certain assumptions, provide a better proxy for academic impact as was shown by Ribas et al. [33]. Second, while citations can take years to happen, reputations tend to be established earlier on. As a consequence, the quality of new research can be promptly estimated using P-scores computed on the newest publication data. Third, P-scores do not require citation data to be computed, which means that they can be recomputed quickly whenever one finds fit or convenient.

Our interpretation of these results are based on the intuition that while H-indices provide a quantification of the popularity of a publication or researcher among their peers, P-scores provide a quantification of the reputation of a publication or researcher among peers. When P-scores and H-indices are strongly correlated, we have venues that are popular and of high reputation or unpopular with a low reputation. Second, when P-scores are high (better rank position) and H-indices are low (worse rank position) we have venues of good reputation that are associated with small research communities or subareas and are accordingly less popular. Third, when P-scores are low (worse rank position) and H-indices are high (better rank position) we have venues of good popularity (usually associated with large research communities) that do not have a high reputation (in the sense that the reference group of researchers does not publish frequently there).

2 Related Work

Quantitative measures of scientific impact have a long history. An influential approach is Garfield's Impact Factor [15], a journal-level measure that is defined by the mean number of citations received by articles published in a given journal over a 2-year period. Despite being one of the earliest approaches at measuring scientific impact, it has showed remarkable survivability. Pinski et al. [30] describe several limitations with the Impact Factor. Impact factors also suffered criticism for being misleading [27, 36] and for lacking reliable validation from an independent audit [35]. Riikonen et al. [34] showed that actual citation and publication counts were better predictors of a scientist's contribution than impact factors in a study that assessed the scientific contribution of Finnish researchers in biomedicine.

The acceptance rate is another important journal-level indicator that attempts to quantify scientific impact. It is defined by the proportion of accepted papers relative to the number of submitted papers. Chen et al. [11] have shown that highly selective conferences, the ones having low acceptance rates, have higher scientific impact measured in terms of the number of citations received by the published papers.

The H-index is an author-level index proposed by Hirsch [17]. Since its formulation, it has been widely adopted as a measure of individual scientific research output. The H-index takes into account both the number of publications and the number of citations per publication, achieving higher values when a researcher obtains a consistently high number of citations over multiple publications. The H-index has been criticized for not taking into account field specific citation statistics [38]. The G-index, proposed by Egghe [13], is a less strict variation of the H-index.

Citation counts and derived measures are straightforward to calculate, but they represent only a coarse estimate of academic impact, because they are essentially popularity measures. Some people are popular but not prestigious and vice versa. For example, an author of pulp detectives may sell many books, but may not have earned the respect of literary critics as pointed by Bollen et al. [6]. Citations from prestigious journals are more relevant than citations from peripheral journals, as noted by Pinski et al. [30]. To address this issue, some studies distinguish popularity from reputation or prestige with the usage of citation weighting strategies [12,39] or Page Rank based methods [6,37]. Furthermore, Piwowar [31] noted that citation-based indices are slow, since the first citation to a scientific article can take years to happen, concluding that the development of alternative indices to complement citation analysis is not only desirable, but a necessity. As discussed by Leydesdorff [23], each indicator has advantages and disadvantages deriving from their inherent biases.

Martins et al. [25] proposed a method of assessing the quality of scientific conferences through a machine learning classifier that makes use of multiple features in addition to citations. Gonçalves et al. [16] quantified the impact of various features on a scholar's popularity. They concluded that, even though most of the considered features are strongly correlated with popularity, only two features are needed to explain almost all the variation in popularity between different researchers: the number of publications and the average quality of the scholar's publication venues.

The idea of reputation, without the direct use of citation data, was discussed by Nelakuditi et al. [28]. They proposed a measure called *peers' reputation* for research conferences and journals, which ties the selectivity of the publication venue based upon the reputation of its authors' institutions. The proposed measure was shown to be a better indicator of the selectivity of a research venue than the acceptance ratio.

3 P-score: a network-based index

Contrary to citation counts such as the Impact Factor [15,36,1,3] and H-index [5,4,14,7,8], P-scores are a graph-based modeling index that takes into account the relations among researchers, papers they published and their publication venues. They are based on a framework of reputation flows we introduced previously [32]. Let us review it briefly.

A *reputation graph* in academia is a graph with three node types: (a) *reputation sources* representing groups of selected researchers, (b) *reputation targets* representing venues of interest, and (c) *reputation collaterals* representing entities we want to compare such as research groups and academic departments. Figure 1 provides a generic illustration of our reputation graph and introduces the following notation: S is the set of reputation sources, T is the set of reputation targets, and C is the set of reputation collaterals.

The reputation of source nodes influences the reputation of target nodes as much as the reputation of target nodes influences the reputation of source nodes. Note that the reputation of target nodes also influences the reputation of collaterals, but the reputation of collaterals has no impact in the reputation of sources and targets. The use of collaterals allows us to isolate the impact of a set of arbitrary nodes on the reputation graph if there is the need to do it, fixing reputation sources as the only set of nodes providing reputation.

Given that the reputation of collaterals has no effect on the reputation of nodes of other types, we can split the model in two phases. In the first phase, we propagate the reputation of the sources to the targets. In the second phase, we propagate the reputation of the targets to the collaterals.

The reputation flows framework allows for different configurations. In this work, we adopt research groups as reputation sources and publication venues as reputation targets, however collaterals (such as individual researchers) are absent from the ranking model, because we are only interested in calculating the P-scores of publication venues (weights of target nodes). Therefore, we do not need to perform the second phase when calculating venue P-scores.

Figure 1 Structure of the reputation graph.

The interaction between reputation sources and reputation targets is inspired by the notion of *eigenvalue centrality* in complex networks [10, 22, 22, 29]. In the reputation graph, if we consider only sources and targets, it is easy to identify reputation flows from sources to sources, from sources to targets, from targets to sources, and from targets to targets. These reputation flows can be modeled as a stochastic process. In particular, let P be a *right stochastic* matrix of size $(|S| + |T|) \times (|S| + |T|)$ with the following structure:

$$P = \left[\begin{array}{c|c} (d^{\langle S\rangle}).P^{\langle SS\rangle} & (1 - d^{\langle S\rangle}).P^{\langle ST\rangle} \\ \hline (1 - d^{\langle T\rangle}).P^{\langle TS\rangle} & (d^{\langle T\rangle}).P^{\langle TT\rangle} \end{array}\right] \qquad (1)$$

where each quadrant represents a distinct type of reputation flow, as follows:

$P^{\langle SS\rangle}$: right stochastic matrix of size $|S| \times |S|$ representing the transition probabilities between reputation sources;
$P^{\langle ST\rangle}$: matrix of size $|S| \times |T|$ representing the transition probabilities from reputation sources to targets;
$P^{\langle TS\rangle}$: matrix of size $|T| \times |S|$ representing the transition probabilities from reputation targets to sources;
$P^{\langle TT\rangle}$: right stochastic matrix of size $|T| \times |T|$ representing the transition probabilities between reputation targets.

The parameters $d^{\langle S\rangle}$, the fraction of reputation one wants to transfer among the source nodes themselves, and $d^{\langle T\rangle}$, the fraction of reputation one wants to transfer among the target nodes themselves, control the relative importance of the reputation sources and targets. Assuming that the transition matrix P is ergodic, we can compute the steady state probability of each node and use it as a reputation score, the P-score. More formally, we can write:

$$\gamma = \gamma P \qquad (2)$$

where γ is a row matrix with $|S|+|T|$ elements, where each row represents the transition probabilities of a node in the set $S \cup T$.

When there are collaterals, we can perform the second phase (propagation to collateral nodes) by further propagating the steady state probabilities of target nodes to the collateral set. In this case, we need a transition matrix $P^{\langle TC\rangle}$ of size $|T| \times |C|$ representing the transitions from reputation targets to collaterals.

The conceptual framework of reputation flows can be used in the academic context to model the transference of reputation between research groups and publication venues by associating each type of reputation flow with a specific quadrant of matrix P. That is:

$$P = \left[\begin{array}{c|c} Group \rightarrow Group & Group \rightarrow Venue \\ \hline Venue \rightarrow Group & Venue \rightarrow Venue \end{array}\right] \tag{3}$$

In this conceptual framework, research groups are aggregations of authors and publication venues are aggregations of papers. In the first quadrant, the framework represents the reputation flow from research groups to research groups, which can be expressed in terms of co-authorship relations. In the second and third quadrants, the framework represents group-venue and venue-group relations, respectively. An author who publishes a paper somehow transfers its own reputation to that paper or the converse, a paper may transfer its reputation or acceptance by the community to the authors who published it. In the fourth quadrant, the framework represents the reputation flow between venues, or between the papers in these venues. When a paper cites another, it is somehow transferring part of its reputation to the cited paper. This last quadrant is the focus of much more attention than the other ones by the academic community.

We should note that, while our network model allows modeling citations in the fourth quadrant, it is possible to compute steady state probabilities for the network without consideration to citations. This is accomplished by setting the parameter $d^{\langle T \rangle} = 0$. Thus, it should be clear that in all experiments described in this work, P-scores are computed without taking citations into account.

3.1 Reputation Sources

The choice of the reputation sources is an important part of the method since its composition has a direct impact on the final rankings. There is no definitive way to make it. This choice depends on what we want to measure.

One way to determine the top CS departments is through a simple randomization procedure. It starts with all 126 research groups evaluated by the NRC [1] in its 2011 evaluation of CS graduate programs in the USA. First, we need to instantiate the model described previously with a single difference, a subset of 10 groups from the NRC evaluation is chosen randomly and used as reputation sources, while the remaining 116 groups are used as reputation collaterals. This time we also need to execute the propagation of reputations from targets to collaterals with a transition matrix $P^{\langle TC \rangle}$ of size $|T| \times |C|$ that represents relations from venues to the collateral research groups. This matrix can be built the same way the $P^{\langle ST \rangle}$ (sources to targets) matrix is built, that is, by associating authors to research groups and their papers to publication venues and taking the relative number of papers published in each venue as a transition probability.

Observe that the randomization procedure is only needed to select the source nodes for the calculation of venue P-scores, it has no direct relation with the venue ranking model described in the last section. A run of that procedure works as follows:

[1] http://www.nap.edu/rdp/

1. Randomly select 10 departments from the set of top CS departments and use them as the set s of reputation sources
2. Compute steady state probabilities for all nodes using the method described in the last section
3. Using the steady state probabilities of reputation collaterals as a score, select the 10 entities with highest scores and use them as a new set s_{new} of reputation sources
4. If $s_{new} \not\equiv s$ then $s \leftarrow s_{new}$ and go back to step 1
5. $s_{auto} \leftarrow s_{new}$
6. Take s_{auto} as the set of automatically selected reputation sources
7. Exit

We repeated this procedure until the set of top 10 groups no longer changed. By applying this randomization procedure 100 times to a set of 126 USA graduate programs, we ended up with a subset of 12 CS programs that appeared among the top 10 at least once, after the process stabilized. These 12 CS programs are described in Table 1.

#	Department
1	Carnegie Mellon University
2	Georgia Institute of Technology
3	Massachusetts Institute of Technology
4	Stanford University
5	University of California-Berkeley
6	University of California-Los Angeles
7	University of California-San Diego
8	University of Illinois at Urbana-Champaign
9	University of Maryland College Park
10	University of Southern California
11	University of Michigan-Ann Arbor
12	Cornell University

Table 1 Reputation sources obtained with the randomization procedure.

All the 12 departments above are among the top 5th percentile in the ranking produced by NRC. Moreover, the first 8 listed groups appeared among the top 10 at every single run. This suggests that our recursive procedure is able to take advantage of patterns in the publication streams of the various CS departments to determine the most reputable ones in fully automatic fashion. We further observe that this was done while setting the parameter $d^{\langle T \rangle} = 0$. That is, we did not use information on citation counts in the model.

4 Correlation with H-indices

Bibliometric indicators can be roughly divided into two types. The attention indicators usually estimate researcher attention or the diffusion of a journal's articles through different forms of citation counts. The reputation indicators try to estimate a researcher's reputation based on the researcher's overall scientific production. What we argue in this paper is that both perspectives have a complementary nature, since indicators of different types are measuring similar but distinct aspects of a journal's scientific relevance.

We performed a combined analysis of conference rankings in Computer Science based on the H-index, an attention indicator, and P-score, a reputation indicator. We chose the H-index as an attention indicator because of its popularity and availability in bibliometric search engines such as Google Scholar, Web of Science and Scopus. We chose P-score as a reputation indicator because of its ease of calculation and for being a good proxy for academic impact as shown by Ribas et al. [33].

The H-index [17] is a composite index of lifelong scientific contribution that takes into account the productivity of researchers and the citation impact of their publications. A researcher has an H-index value of h if she has h publications that have been cited at least h times. For example, if a researcher has published 20 papers that received at least 20 citations each, then her H-index is 20, assuming that 20 is the highest number for which the definition of the H-index holds. A journal's H-index can be computed in the same way by considering all publications and their collective citation data over a definite period as suggested by Braun et al. [9]. Since the H-index was proposed in 2005, the original paper by [17] was cited 7,949 times, which attests to its popularity [2].

As any scientific impact index, the H-index has advantages and disadvantages. Some of its advantages are:

– The simplicity and intuitiveness of its formulation.
– The longer citation time window when compared to impact factors.
– H-index has a good predictive power regarding scientific achievement [7, 18].

H-index depends on citations to be calculated, thus it suffers from the same problems as any other citation-based indices. Some of its major disadvantages are:

– H-index does not distinguish citations from prestigious and peripheral journals, thus it measures popularity rather than quality.
– The time to first citation can be very long.
– It is not easy to collect all relevant information.
– H-index is field-dependent [38].
– Major sources do not always agree on its value [4].

We propose to use P-score as a complementary index to deal with some of the problems presented by citation-based indices such as the H-index. P-score is a index of reputation among peers based on the publication pattern of a set of reference groups of researchers, the reference seeds. It differs from other network-based indices because the venues' reputations derive exclusively from the reference seeds, which allows better control of the actual reputation flows. Additionally, to calculate P-scores we do not need citation data, which make P-scores easier to obtain than H-indices and avoids the problem of lack of data that arises from the long time to first citation.

We conducted experiments to investigate the relationship between P-scores and H-indices in conference rankings. In particular, we wanted to estimate the degree of correlation between both indices. The data set consists of 882 Computer Science [3] conferences for which the H5 index was available on Google Scholar [4] in 2012. P-scores were

[2] According to Google Scholar, up to the end of 2017.

[3] While our data set is entirely composed of Computer Science conferences, nothing in our method is particular to this field. That is, our index is applicable to any field of knowledge.

[4] https://scholar.google.com.br/

calculated with data obtained from DBLP [5]. The reference set of research groups was selected through the randomization process described in the previous section, yielding the 12 departments listed in Table 1.

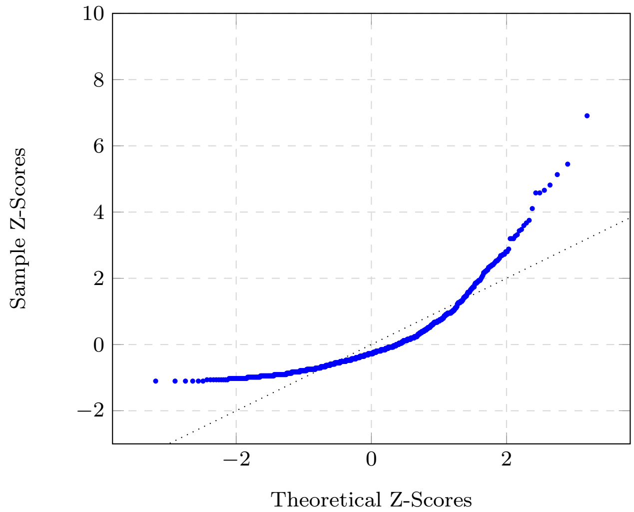

Figure 2 Normal probability plot: H-index

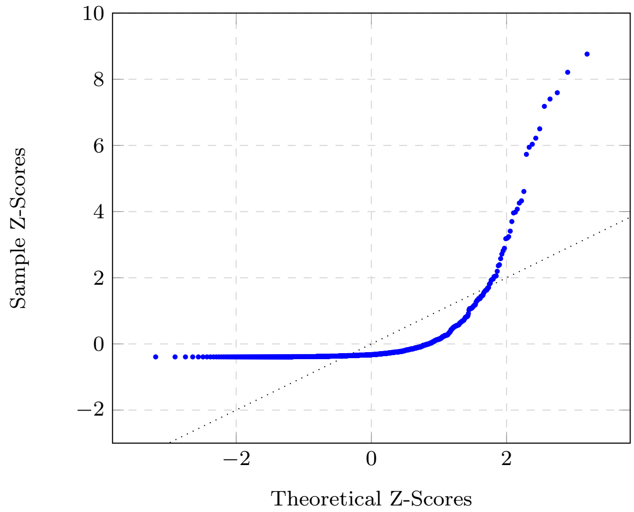

Figure 3 Normal probability plot: P-score

[5] http://dblp.uni-trier.de/

We first plotted normal probability plots for P-scores and H-indices in our conference data set to identify any departures from normality. A normal probability plot shows the theoretical Z-scores according to the standard normal quantile function on the horizontal axis. For example, at the 90th percentile (the value below which 90% of the observations will fall), the normal distribution has a Z-score of 1.28, meaning that the distribution value at this point is 1.28 standard deviations above the median value. On the vertical axis, if we plot the Z-scores according to the quantile function described by our sample, we can verify whether it is consistent with a sample from a normal distribution.

We compared theoretical and sample Z-scores for 882 quantiles (one for each data point) in Figures 2 and 3. The dotted line is the identity line for which sample and theoretical Z-scores converge. As becomes clear from the plots, both distributions are significantly skewed, indicating that P-scores and H-indices do not follow normal distributions. This restricts the application of correlation measures such as Pearson's Product Moment Correlation in our analysis, since it requires that variables be approximately normally distributed.

For this reason, we chose the Kendall-Tau coefficient to study the correlation between H-index and P-score. It is a measure of rank correlation that assumes a value between -1 and 1. It is higher when observations have similar ranks, assuming a value of 1 when two rankings are identical and -1 when they are the inverse of each other. A Kendall-Tau coefficient close to zero indicates that both rankings are uncorrelated [21, 2].

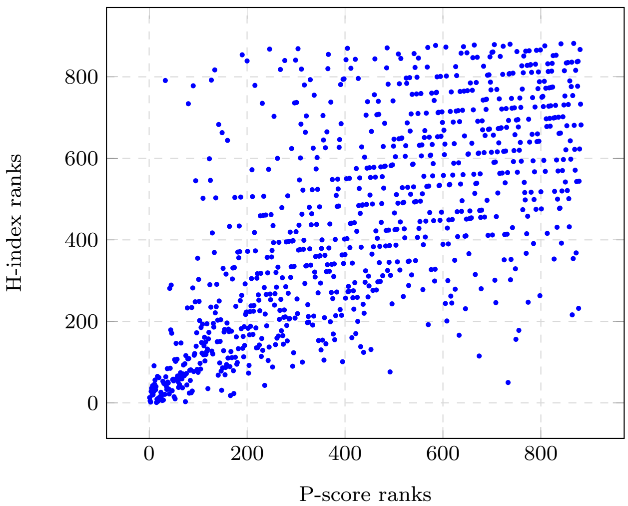

Figure 4 P-score and H-index rankings.

Consider two conference rankings. Figure 4 shows the ranks of our 882 conferences according to their P-scores on the X-axis and their H-indices on the Y-axis. Points that are closer to the origin have better ranks (i.e. a data point at position 1 is better ranked than a data point at position 200). The plot shows an evident correlation between the ranks that is confirmed by a high Kendall-Tau of 0.5200259445. With a p-value smaller than 0.000001, we can safely reject the null hypothesis (that the rankings are uncorrelated) with a reasonable margin.

5 Assessing conferences in CS

Since P-scores and H-indices are strongly correlated, most data points in our example are clustered close to the identity line, showing an apparent agreement between the indices. However, there are noticeable differences, particularly in those cases for which the P-score rank position is high and the H-index rank is low (top left corner of Figure 4), or for which the H-index rank is high and the P-score rank is low (bottom right corner of Figure 4). Therefore, we can distinguish between three groups of interest in the data set:

- *Center Group*: highly correlated data points. These are the venues that are popular (i.e. high H-index) and reputable (i.e. high P-score), or venues that are unpopular and not very reputable.
- *Top Group*: venues with a high P-score and low H-index. These are the venues that are reputable (i.e. endorsed by researchers at top CS departments), but not very popular.
- *Bottom Group*: venues with a high H-index and low P-score. There are the venues that are popular, but not very reputable.

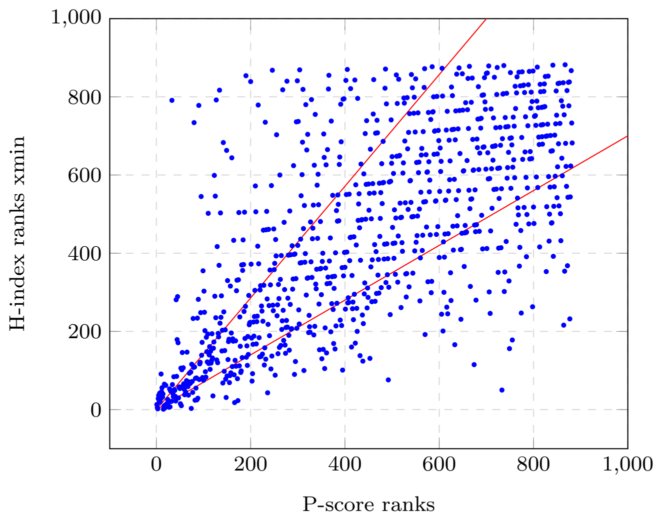

Figure 5 Angle strategy for delimiting our three groups of interest: Top, Center and Bottom - with pivot at the origin.

Next, we devise a strategy for delimiting the three sets presented above. We draw two line segments, starting from the origin, separated by an angle θ, and examine how the Kendall-Tau coefficient for the data points in the Center Group varies with respect to variations in the angle θ. It is what we call the "angle strategy", as illustrated in

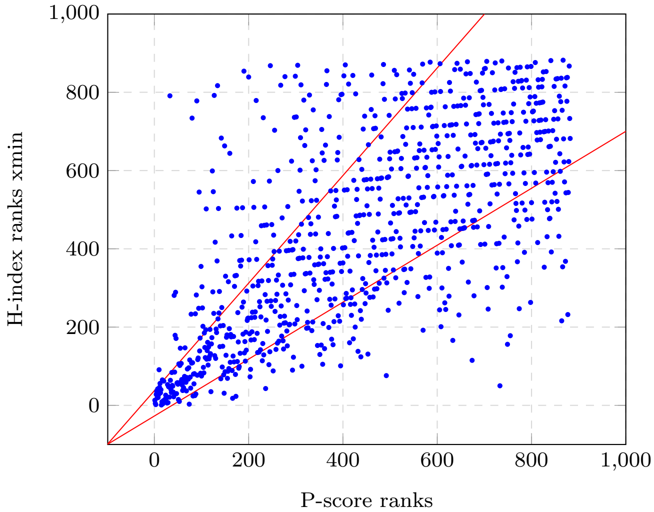

Figure 6 Angle strategy for delimiting our three groups of interest with pivot positioned behind the origin, at point (-100, -100).

Figure 5. We immediately notice several data points close to the origin which are not included in the Center Group. This is a problem, given that these points correspond to CS conferences that have a high H-index rank and a high P-score rank and thus, are strongly correlated. To correct this, we move the point where the lines meet, which we also refer to as the pivot, to a point behind the origin, one such as (-100, -100), as illustrated in Figure 6. By doing so, we ensure that all data points near the origin are included in the Center Group.

Angle	Kendall-Tau	Items in center group
90	0.5003	882
80	0.5003	882
70	0.5032	881
60	0.5154	876
50	0.5477	858
40	0.5780	829
30	0.6465	750
20	0.7321	616
10	0.8361	370

Table 2 Angle strategy Kendall-Tau in center group by varying the angle.

Table 2 describes how the Kendall-Tau coefficients vary for items in the Center Group, as we vary the angle θ, with the pivot positioned behind the origin as in Figure 6. We notice that for $\theta = 20°$, 616 conferences are included in the Center Group (close to 70% of all conferences in the data set) and, that is accomplished with a very high Kendall-Tau coefficient (i.e. above 0.7). Thus, in our subsequent analysis, we employed the angle strategy with $\theta = 20^\circ$, since it provides a good trade off between correlation and coverage in our data set.

#	Top	Center	Bottom
1	IROS	ICRA	ISCAS
2	ICMCS	CHI	HICSS
3	ISIT	CVPR	EUROCRYPT
4	ICCD	AAAI	IMC
5	CDC	NIPS	ISWC
6	ICPP	DAC	ASIACRYPT
7	SPAA	ICASSP	ISMB
8	COLT	STOC	TACAS
9	WCNC	FOCS	MOBIHOC
10	LCPC	INTERSPEECH	VTC
11	ISQED	ICCAD	ACSAC
12	APPROX	IJCAI	PSB
13	ISER	SODA	GECCO
14	WACV	INFOCOM	ICWS
15	ITS	ICML	ESOP
16	ICWSM	ICIP	ECAI
17	AIED	SIGMOD	AAMAS
18	GIS	ICSE	PKC
19	HRI	IPPS	CP
20	CCCG	ICCV	ISMIR
21	HOTNETS	ICDE	ICIS
22	SECON	ACL	ECRTS
23	HUMANOIDS	ECCV	FSE
24	ICCCN	ISCA	PIMRC
25	PPSC	KDD	GI
26	DGO	SIGCOMM	SACMAT
27	HIPC	SC	PG
28	WAFR	WWW	DIS
29	ISBI	ICDCS	P2P
30	SMC	DATE	MODELS

Table 3 Conferences (short names) ranked by P-score in the bottom, center and top groups.

With the pivot positioned at (-100, -100) and $\theta = 20^\circ$, the Top Group ended with 127 conferences, the Center Group with 616 conferences, and the Bottom Group with 139 conferences. Table 3 presents the top 30 conferences, ranked by P-score, for each of our three groups.

The Center Group consists of conferences that have a high correlation between reputation and popularity. The reputation of these conferences is well represented by the number of citations they receive. This group represents conferences that are relatively easy to classify regarding their scientific impact. Among them we find WWW, CVPR and KDD, to name a few.

The Top Group consists of conferences with a high P-score rank and low H-index rank. While these conferences receive a relatively low number of citations, reputable researchers (i.e. our reference set of top researchers in Computer Science) still publish in them consistently, as we can observe from their high P-score ranks. The primary reason is that conferences in the Top Group are usually venues associated with smaller subareas of Computer Science, and are therefore venues that receive a smaller number of citations. Because of that, their H-indices tend to be smaller and their H-index rank positions tend to be worse. Despite that, many of these venues might have a high reputation in their subareas and thus, high P-scores. For example, ICPP, SPAA, LCPC and PPSC are conferences related to parallel processing that capture considerable attention from the top CS departments, as we can observe from their high P-scores. However, due to their modest H-indices, they may end up being neglected by funding councils and committees in their assessment. The total number of publications in Parallel Processing is comparatively low compared to other subareas of Computer Science such as Algorithms or Database Management Systems [19]. So, it is reasonable to assume that conferences in this subarea are expected to receive fewer citations than their counterparts in more popular subareas. P-scores can assist human evaluators with funding decisions by shedding light on these conferences.

The Bottom Group consists of conferences with a low P-score rank and high H-index rank. These conferences receive a high number of citations, but their reputation among top computer scientists is comparatively low, as reflected by their P-scores. Conferences in this group are usually situated in the intersection between Computer Science and other areas of knowledge. Some examples of such venues are ISMB, GECCO and PSB, which are related to Biocomputing, ISMIR, which is related to Music Information Retrieval, and ISCAS, and VTC which are related to Engineering and Electronics. Because these are large research communities, venue H-indices tend to be relatively high (particularly with regard to Computer Science). In the case of these events, funding councils in Computer Science should consider not only the high H-indexes, but also the venue reputation among computer scientists. Indeed, an exceptionally popular and reputable conference in electronics may not be as popular or prestigious when considering it from the point of view of computer scientists. That is, P-scores may aid human evaluators with funding decisions by pinpointing conferences that intersect multiple areas of knowledge and by providing an estimate of the conference's reputation among peers in the field of interest.

6 Conclusions

We have compared the P-scores and H-indices of 882 conferences in Computer Science. We have found that P-scores and H-indices have a very high correlation reflected by a Kendall-Tau of approximately 0.52. However, we have also found important differences between the two indices. Our interpretation of these results are based on the intuition that while H-indices provide a quantification of the popularity of a publication or researcher among their peers, P-scores provide a quantification of the reputation of a publication or researcher among peers. We can then distinguish three separate cases. First, when P-scores and H-indices are strongly correlated, we have venues that are popular and of high reputation among computer scientists. Second, when P-scores are large (better rank position) and H-indices are small (worse rank position) we have venues of good reputation among computer scientists that are associated with small research communities or subareas. Third, when P-scores are small (worse rank position) and H-indices are large, we have venues of good popularity (usually associated with large research communities outside Computer Science) that do not count with high reputation among computer scientists (in the sense that they do not publish frequently there). These differences indicate a complementary aspect to the indices. P-scores, used in conjunction with H-indices, allow distinguishing the three distinct types of venues mentioned above. This provides useful information which can be employed by research funding councils and committees to better understand the scientific relevance of venues and researchers and thus, take more informed funding decisions.

References

1. Ahmed, K.M.: Thomson Reuters Impact Factor 2017 (2017)
2. Baeza-Yates, R., Ribeiro-Neto, B.: Modern Information Retrieval: the Concepts and Technology Behind Search (2nd ed.). Addison-Wesley Publishing Company (2011)
3. Balaban, A.T.: Positive and negative aspects of citation indices and journal impact factors. Scientometrics **92**(2), 241–247 (2012)
4. Bar-Ilan, J.: Which h-index? - A comparison of WoS, Scopus and Google Scholar. Scientometrics **74**(2), 257–271 (2008)
5. Benevenuto, F., Laender, A.H., Alves, B.L.: The H-index paradox: your coauthors have a higher H-index than you do. Scientometrics **106**(1), 469–474 (2016)
6. Bollen, J., Rodriquez, M.A., Van de Sompel, H.: Journal status. Scientometrics **69**(3), 669–687 (2006)
7. Bornmann, L., Daniel, H.D.: Does the h-index for ranking of scientists really work? Scientometrics **65**(3), 391–392 (2005)
8. Bornmann, L., Marx, W.: The h-index as a research performance indicator. European Science Editing **37**(3), 77–80 (2011)
9. Braun, T., Glänzel, W., Schubert, A.: A Hirsch-type index for journals. Scientometrics **69**(1), 169–173 (2006)
10. Brin, S., Page, L.: The anatomy of a large scale hypertextual Web search engine. Computer Networks and ISDN Systems **30**(1-7), 107–117 (1998)
11. Chen, J., Konstan, J.A.: Conference paper selectivity and impact. Communications of the ACM **53**(6), 79–83 (2010)
12. Ding, Y., Cronin, B.: Popular and/or prestigious? Measures of scholarly esteem. Information Processing and Management **47**(1), 80–96 (2011)
13. Egghe, L.: Theory and practise of the g -index. Scientometrics **69**(1), 131–152 (2006)
14. Egghe, L.: The Influence of Transformations on the h-Index and the g-Index. Journal of the American Society for Information Science and Technology **59**(8), 1304–1312 (2008)
15. Garfield, E.: Citation Indexes for Science. Science **122**(3159), 108–111 (1955)
16. Gonçalves, G.D., Figueiredo, F., Almeida, J.M., Gonçalves, M.A.: Characterizing scholar popularity: a case study in the computer science research community. Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries pp. 57–66 (2014)
17. Hirsch, J.E.: An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences **102**(46), 16569–16572 (2005)
18. Hirsch, J.E.: Does the h index have predictive power? Proceedings of the National Academy of Sciences **104**(49), 19193–19198 (2007)
19. Hoonlor, A., Szymanski, B.K., Zaki, M.J.: Trends in Computer Science Research. Communications of the ACM **56**(10), 74–83 (2013)
20. Kellner, A.W.A., Ponciano, L.C.M.O.: H-index in the Brazilian Academy of Sciences - comments and concerns. Anais da Academia Brasileira de Ciências **80**(4), 771–781 (2008)
21. Kendall, M.G.: Rank Correlation Methods (2nd ed.). Hafner Publishing Company (1955)
22. Langville, A.N., Meyer, C.D.: Google's pagerank and beyond: The science of search engine rankings. Princeton University Press (2006)
23. Leydesdorff, L.: How are new citation-based journal indicators adding to the bibliometric toolbox? Journal of the American Society for Information Science and Technology **60**(7), 1327–1336 (2009)
24. Loach, T.V., Evans, T.S.: Ranking journals using altmetrics. CoRR **abs/1507.00451** (2015)
25. Martins, W.S., Gonçalves, M.A., Laender, A.H., Pappa, G.L.: Learning to assess the quality of scientific conferences: a case study in computer science. Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries pp. 193–202 (2009)
26. Nane, T.: Time to first citation estimation in the presence of additional information. In: Proceedings of the 15th International Society of Scientometrics and Informetrics Conference, pp. 249–260 (2015)
27. Nature: Time to remodel the journal impact factor. Nature **535**(7613), 466 (2016)

28. Nelakuditi, S., Gray, C., Choudhury, R.R.: Snap judgement of publication quality: how to convince a dean that you are a good researcher. ACM SIGMOBILE Mobile Computing and Communications Review **15**(2), 20–23 (2011)
29. Newman, M.: Networks: An Introduction. Oxford University Press (2010)
30. Pinski, G., Narin, F.: Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics. Information Processing and Management **12**(5), 297–312 (1976)
31. Piwowar, H.A.: Value all research products. Nature **493**(7431), 159 (2013)
32. Ribas, S., Ribeiro-Neto, B., Santos, R.L., de Souza e Silva, E., Ueda, A., Ziviani, N.: Random walks on the reputation graph. In: Proceedings of the 2015 International Conference on The Theory of Information Retrieval, ICTIR '15, pp. 181–190. ACM, New York, NY, USA (2015)
33. Ribas, S., Ribeiro-Neto, B., de Souza e Silva, E., Ueda, A.H., Ziviani, N.: Using reference groups to assess academic productivity in computer science. In: Proceedings of the 24th International Conference on World Wide Web, WWW '15 Companion, pp. 603–608. ACM, New York, NY, USA (2015)
34. Riikonen, P., Vihinen, M.: National research contributions: A case study on Finnish biomedical research. Scientometrics **77**(2), 207–222 (2008)
35. Rossner, M., Van Epps, H., Hill, E.: Show me the data. The Journal of Cell Biology **179**(6), 1091–1092 (2007)
36. Saha, S., Saint, S., Christakis, D.A.: Impact factor: a valid measure of journal quality? Journal of the Medical Library Association **91**(1), 42–46 (2003)
37. Sun, Y., Giles, C.L.: Popularity weighted ranking for academic digital libraries. In: Proceedings of the 29th European Conference on IR Research, ECIR'07, pp. 605–612. Springer-Verlag, Berlin, Heidelberg (2007)
38. Wendl, M.C.: H-index: However ranked, citations need context. Nature **449**(7161), 403 (2007)
39. Yan, E., Ding, Y., Sugimoto, C.R.: P-Rank: An indicator measuring prestige in heterogeneous scholarly networks. Journal of the American Society for Information Science and Technology **62**(3), 467–477 (2011)