Scientometrics (2019) 121:1269–1291
https://doi.org/10.1007/s11192-019-03247-0

P-score: a reputation bibliographic index that complements citation counts

João Mateus de Freitas Veneroso[1] · Marlon Dias[1] · Alberto Ueda[1] · Sabir Ribas[1] · Berthier Ribeiro-Neto[1] · Nivio Ziviani[1] · Edmundo de Souza e Silva[2]

Received: 30 November 2018 / Published online: 28 September 2019

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 researchers. Thus, it is important to assess publication track records quantitatively. To quantify publication records, bibliographic indices are usually adopted and, 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, a very popular citation-based index, in the setting of conference ranking. While *H*-indices reflect the popularity of a given publication or researcher in academia, *P*-scores can reflect the reputation of a publication or researcher among its peers, considering a reference set of reputable researchers. Popularity and reputation are frequently considered to be equivalent properties in the formulation of citation based indices, however these properties 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 show 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 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

Introduction

Academic research evaluation is a topic of interest to universities, research groups, research funding institutions, and the public at large. An effective evaluation helps in the decisions regarding laboratory space allocation, research grants, tenure track, and university choice

✉ João Mateus de Freitas Veneroso
jmfveneroso@gmail.com

[1] Departamento de Ciência da Computação Sala 4304, Universidade Federal de Minas Gerais, Av. Pres. Antônio Carlos, 6627 - Pampulha, Belo Horizonte, MG 31270-901, Brazil

[2] Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

by students. It also has a direct impact on the reputation of universities and research institutions, which affects their ability 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 an institution. In particular, considerable emphasis is given to the papers published by professors and students in this institution, as well as the quality of these papers in terms of their impact on scientific knowledge. A popular approach to the evaluation of scientific impact consists in comparing journals, conferences or papers with the usage of academic impact indices. Most often, these indices are based on the number of citations received by scientific publications, assuming that citation counts act as a good proxy for research quality. However, this method has multiple shortcomings.

First, in the evaluation and bibliometrics research communities, citations are understood as a measure of popularity rather than a proxy for scientific impact, which is a similar but not identical concept. 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 (Nane 2015). 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 (Kellner and Ponciano 2008). 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 total number of endorsements the actor receives from other actors, while its reputation or prestige depends on the reputations of the endorsing actors (Bollen et al. 2006). 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 Ribas et al. (2015a) 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. Reputation is a subjective concept that can be quite difficult to define and quantify in the form of a bibliometric index. Thus, instead of relying on a precise definition of reputation, *P*-score takes an agnostic view, exploiting the transference of this quantity called reputation among entities in order to identify the most reputable ones. The concept of reputation is determined by the choice of a reference group, which is presumed to be composed of reputable entities that act as sources of reputation. *P*-score can deal 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. (2015a). 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.

In this work, we show that the Pscores of conferences in Computer Science show a significant correlation with citation based indices, such as the *H*-index. Yet, there are important cases where the indices diverge. 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).

Related work

Quantitative measures of scientific impact have a long history. An influential approach is Garfield's Impact Factor (Garfield 1955), 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 and Narin (1976) describe several limitations with the Impact Factor. Impact factors also suffered criticism for being misleading (Nature 2016; Saha et al. 2003) and for lacking reliable validation from an independent audit (Rossner et al. 2007). Riikonen and Vihinen (2008) 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 and Konstan (2010) 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 (2005). 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 (Wendl 2007). The *G*-index, proposed by Egghe (2006), 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. (2006). Citations from prestigious journals are more relevant than citations from peripheral journals, as noted by Pinski and Narin (1976). To address this issue, some studies distinguish popularity from reputation or prestige with the usage of

Fig. 1 Structure of the reputation graph

citation weighting strategies (Ding and Cronin 2011; Yan et al. 2011) or Page Rank based methods (Bollen et al. 2006; Sun and Giles 2007). Furthermore, Piwowar (2013) 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 (2009), each indicator has advantages and disadvantages deriving from their inherent biases.

Martins et al. (2009) 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. (2014) 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. (2011). 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.

P-score: a network based index

Contrary to citation counts such as the Impact Factor (Garfield 1955; Saha et al. 2003; Balaban 2012) and *H*-index (Benevenuto et al. 2016; Bar-Ilan 2008; Egghe 2008; Bornmann and Daniel 2005; Bornmann and Marx 2011), *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* Ribas et al. (2015b) which we describe in this section.

Let a *reputation graph* be a graph with three node types of academic entities: (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 to produce a conference ranking, where *P*-scores are the weights of target nodes. Additionally, we can include research groups in the model as collaterals to rank them in terms of the publication patterns of the reputation sources. In fact, this formulation can be used to produce the reputation sources automatically in a procedure described in Sect. 3.3. However, once we have determined the reputation sources, there is no need to add collaterals to the model and we can skip the second phase altogether if we only wish to produce conference rankings.

The interaction between reputation sources and reputation targets is inspired by the notion of *eigenvalue centrality* in complex networks (Brin and Page 1998; Langville and Meyer 2006; Newman 2010). 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:

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

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. If we do not want to consider reputation flows between nodes of the same type, it is sufficient to set both parameters to zero. If, instead, we want to consider reputation flows between nodes of the same type, we may increase these parameters according to the desired relative importance. Note that, as (i) the sub-matrices $P^{\langle SS\rangle}$ and $P^{\langle TT\rangle}$ are *right stochastic*, (ii) each of the rows of matrices $P^{\langle ST\rangle}$ and $P^{\langle TS\rangle}$ sums to 1, and (iii) the parameters $d^{\langle S\rangle}$ and $d^{\langle T\rangle}$ are both in the range [0,1), then P defines a Markov chain. 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:

$$\gamma = \gamma P \tag{2}$$

where γ is a row matrix with $|S| + |T|$ elements, where each row represents the transition probabilities of a node in the set $S \cup T$. This system of linear equations can be solved with standard Markov chain techniques. Then, from Eq. (2), we obtain the steady state probabilities of all nodes in $S \cup T$, that is, reputation sources and reputation targets.

Flow equations

We recursively define the reputation of sources in terms of the reputation of targets, and the reputation of targets in terms of the reputation of sources. Specifically, the reputation γ_s of a source s is defined as:

$$\gamma_s = \sum_{t \in T} (1 - d^{\langle T \rangle}) \cdot P_{ts}{}^{\langle TS \rangle} \gamma_t + \sum_{s' \in S} (d^{\langle S \rangle}) \cdot P_{s's}{}^{\langle SS \rangle} \gamma_{s'} \tag{3}$$

where $P_{ts}{}^{\langle TS \rangle}$ is the transition probability from t to s, given by $P_{ts}{}^{\langle TS \rangle} = n_{ts}/n_t$, where n_{ts} is the number of edges running from t to s and n_t is the total number of edges running from t. Finally, γ_t is the reputation of target t, defined recursively as:

$$\gamma_t = \sum_{s \in S} (1 - d^{\langle S \rangle}) \cdot P_{st}{}^{\langle ST \rangle} \gamma_s + \sum_{t' \in T} (d^{\langle T \rangle}) \cdot P_{t't}{}^{\langle TT \rangle} \gamma_{t'} \tag{4}$$

where, similarly, $P_{st}{}^{\langle ST \rangle}$ is the transition probability from s to t, given by $P_{st}{}^{\langle ST \rangle} = n_{st}/n_s$, where n_{st} is the number of edges running from s to t and n_s is the total number of edges running from s.

In the context of this work, the reputation graph is a bipartite graph. Therefore, the transition matrix P is reduced to a periodic Markov chain with the following structure:

$$P = \left[\begin{array}{c|c} \mathbf{0} & P^{\langle ST \rangle} \\ \hline P^{\langle TS \rangle} & \mathbf{0} \end{array} \right] \tag{5}$$

From decomposition theory (Meyer 1989), we can obtain values for ranking the set of reputation *sources* by solving:

$$\boldsymbol{\gamma}^{\langle S \rangle} = \boldsymbol{\gamma}^{\langle S \rangle} P' \tag{6}$$

where $P' = P^{\langle ST \rangle} \times P^{\langle TS \rangle}$ is a stochastic matrix and $\boldsymbol{\gamma}^{\langle S \rangle}$ is a row matrix with $|S|$ elements, where each one represents the probability of a node in the set S of reputation sources. Note that matrix P' has dimension $|S| \times |S|$ only and can be solved with standard Markov chain techniques. Then we obtain the reputation of all reputation *targets* linked by the reputation sources:

$$\boldsymbol{\gamma}^{\langle T \rangle} = \boldsymbol{\gamma}^{\langle S \rangle} \times P^{\langle ST \rangle} \tag{7}$$

By modeling the problem as a bipartite reputation graph instead of a general reputation graph, we reduce the network from a graph of size $(|S| + |T|) \times (|S| + |T|)$ to a graph of size $|S| \times |S|$, which allows us to compute the steady state probabilities more efficiently.

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.

More generally, we can define the reputation score of an entity e according to:

$$P\text{-score}(e) = \begin{cases} \sum_{t \in T} P_{te}{}^{\langle TC \rangle} \gamma_t & \text{if } e \in C, \\ \gamma_e & \text{otherwise,} \end{cases} \tag{8}$$

where $P_{te}{}^{\langle TC \rangle}$ is the transition weight from a target node t to a collateral node $e \in C$. The P-score of all candidate entities (targets or collaterals) can then be used to produce an overall reputation-oriented ranking of these entities.

Instantiation example

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:

$$P = \left[\begin{array}{c|c} \text{Group} \rightarrow \text{Group} & \text{Group} \rightarrow \text{Venue} \\ \hline \text{Venue} \rightarrow \text{Group} & \text{Venue} \rightarrow \text{Venue} \end{array}\right] \tag{9}$$

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. Similarly, we do not consider co-authorship relations in this work (first quadrant), i.e., we use $d^{\langle S \rangle} = 0$ in all experiments. The use of academic relationships among nodes of the same type in P-score is an interesting direction of research but it is not the focus of this work.

Figure 2 shows an example with two research groups used as reputation sources, Group 1 and Group 2, and three venues used as reputation targets, venues v_1, v_2 and v_3. Group 1 published 3 papers in venue v_1, 2 papers in venue v_2, and 1 paper in venue v_3. The number of publications of Group 1 is 6. Venue v_1 receives 3 papers of Group 1, and 2 papers of Group 2. The fractions of publications from groups to venues and from venues to groups are the edge weights. We have:

$$P = \left[\begin{array}{cc|ccc} 0 & 0 & 3/6 & 2/6 & 1/6 \\ 0 & 0 & 2/8 & 4/8 & 2/8 \\ \hline 3/5 & 2/5 & 0 & 0 & 0 \\ 2/6 & 4/6 & 0 & 0 & 0 \\ 1/3 & 2/3 & 0 & 0 & 0 \end{array}\right]$$

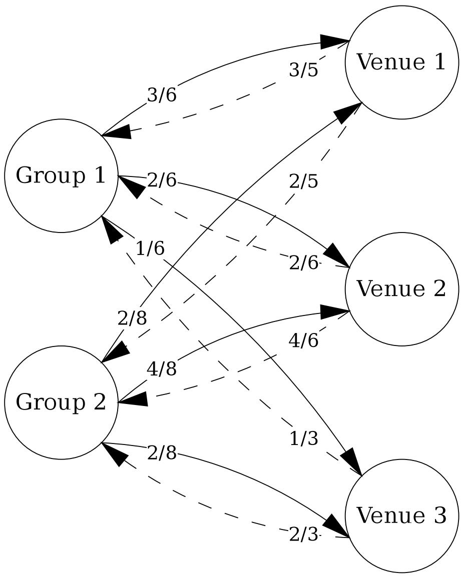
This stochastic matrix corresponds to the Markov chain displayed in Fig. 2, which can be immediately aggregated to a two-state Markov chain, yielding:

$$P' = \begin{bmatrix} 0.467 & 0.533 \\ 0.400 & 0.600 \end{bmatrix}$$

which is the stochastic matrix we use in the solution of Eq. (6). Recall that the dimension of P' is $T \times T$ and, as such, much smaller than that of P for real size problems. Solving Eq. (6) and applying Eq. (7), we obtain the ranking for the three venues:

$$\gamma^{\langle T \rangle} = \langle 0.36, 0.43, 0.21 \rangle \tag{10}$$

Fig. 2 Markov chain for an example with 2 research groups and 3 publication venues



Venue v_2 has the highest rank, followed by v_1, and then by v_3. We remark that the individual values give the *relative reputation* of each publication venue. When we are insterested in computing the *P*-scores of collateral nodes in the reputation graph, we use the Equation (8).

Reputation sources

The choice of 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 this choice, because it depends on what we want to measure. If we select a group of researchers that is generally considered to be prestigious, then we will be transferring this group's prestige across nodes in the reputation graph by following their publication patterns. If we select a group of researchers that share an interest in a research topic, then we will be privileging conferences related to this topic in the final ranking. A possible interpretation of *P*-scores in the context of conference ranking is the amount of attention devoted to a conference by a reference group of researchers.

To provide a general ranking of CS conferences, we can use top CS departments as a reference group. One way to determine the top CS departments is through a simple randomization procedure. This procedure 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

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

Table 1 Reputation sources obtained with the randomization procedure

#	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

$|T| \times |C|$ 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. 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 nodes with highest scores and use them as a new set s_{new} of reputation sources
4. If $s_{\text{new}} \not\equiv s$ then $s \leftarrow s_{\text{new}}$ and go back to step 1
5. $s_{\text{auto}} \leftarrow s_{\text{new}}$
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.

All 12 departments listed 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.

Summary

The basic idea of the *P*-score index is to capture the transference of reputation among source, target, and collateral nodes through their relationships in academia. In particular, in the instantiation used in this work, *P*-score only takes into account researchers and their publication records (i.e., papers in publication venues): given a pre-selected set of reference research groups (reputation sources), *P*-score associates weights to venues according to the publications of these research groups, producing a ranking of publication venues (reputation targets). Furthermore, these weights could also be used to rank other research groups or authors (reputation collaterals).

The reputation of a research group is strongly influenced by the reputation of its members, which is largely dependent on their publication records. We assume that:

1. A research group conveys reputation to a publication venue proportionally to its own reputation.
2. A publication venue conveys reputation to a research group proportionally to its own reputation.

Once a reference group is selected, the reputation of its members is transferred to the venues. Recursively, since the reputation of research groups is correlated with the reputation of the venues in which they published, the venues transfer reputation to the groups. A score for venues can then be computed by solving a system of linear equations relating publication venues and research groups in the reputation graph, as exemplified in Sect. 3.2.

Correlation between *P*-score and citation based indices

Bibliometric indicators can be roughly divided into two types. Popularity indicators estimate the diffusion of a venues' articles by measuring the number of endorsements received by that venue (usually in the form of citation counts). Reputation indicators are similar, but they also try to account for the relative reputations of each endorsing agent, although the concept of reputation may be hard to define. We argue in this paper that both perspectives have a complementary nature, since indicators of different types are measuring similar but distinct aspects of a venue's scientific relevance.

We performed a combined analysis of conference rankings in Computer Science based on *P*-score, a reputation indicator, and the *H*-index, a popularity indicator, taking into consideration that the concepts of reputation and popularity are partially related, but also considering that our interest lies exactly in the points where they diverge more dramatically. Additionally, we provide comparisons between *P*-score, Total Citations and the Impact Factor to shed light on the relation between *P*-score and citation based indices. In computer-related fields, conferences are important channels of dissemination for cutting-edge research findings (Lee 2019). Therefore, conference rankings can be valuable to help identifying promising research topics and good funding opportunities. We chose the *H*-index as the principal indicator of popularity because of its acceptance and availability for conferences in bibliometric search engines such as Google Scholar, Web of Science and Scopus.

The *H*-index (Hirsch 2005) 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. (2006). Since the *H*-index was proposed in 2005, the original paper (Hirsch 2005) 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 (Bornmann and Daniel 2005; Hirsch 2007).
- Its availability in conference rankings.

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, although the concepts can have a significant overlap.
- The time to first citation can be very long.
- It is not easy to collect all the relevant information.
- *H*-index is field-dependent (Wendl 2007).
- Major sources do not always agree on its value (Bar-Ilan 2008).

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 an index of reputation among peers based on the publication pattern of a set of reference groups of researchers, the reputation sources. It differs from other network-based indices because the venues' reputations derive exclusively from the reputation sources, which allows better control of the actual reputation flows. Additionally, to calculate *P*-scores we do not need any citation data, what makes *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.

The empirical validation of *P*-score as a measure of reputation was discussed in the research article by Ribas et al. (2015b). The effectiveness of *P*-score as a method of venue ranking was attested by comparing its performance to the *H*-index baseline in terms of their normalized discounted cumulative gains (nDCG) (Järvelin and Kekäläinen 2002) at various ranking cutoffs using ground truth data obtained from the Qualis system maintained by CAPES[3], a foundation within the Brazilian Ministry of Education. Qualis is an official Brazilian system that annually classifies publication venues in a 1-8 scale. The grades are assigned by a committee of experts in each field of knowledge following a set of criteria, such as: the number of issues, the number of publications, the number of repositories that list the venue, citation information, among others. Using the reference group of departments listed in Table 1 and no citation information, *P*-score scored consistently higher than

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

[3] http://www.capes.gov.br/

the H-index in every ranking cutoff, showing that the ranking produced with P-score more closely resembles the proxy reputation measure provided by the Qualis classification system. Naturally, to understand this evidence as a proof of P-score's ability to capture a venue's reputation, we must first accept that the Qualis ranking provides a reliable measure of venue reputation. The Qualis rankings are carefully produced by a committee of specialists in each field of knowledge, yet we can expect that a different committee would probably produce a slightly different ranking.

While most researchers can probably agree on the relative reputations of some top conferences in their fields of expertise, the distinction becomes blurrier as we get to lower positions in the ranking. In fact, any ground truth that tries to capture venue reputations can be questioned to some extent on the basis of the subjective nature of the concept of reputation. P-score does not impose a particular view on reputation, but it relies on two assumptions: (1) a research group conveys reputation to a publication venue proportionally to its own reputation; (2) a publication venue conveys reputation to a research group proportionally to its own reputation. Considering that these assumptions are valid, the framework of reputation flows assigns reputations to venues based on the publication patterns of the reputation sources. In the context of this work, all P-score rankings were produced using a reference group of CS departments that is generally considered to be reputable, so these P-score rankings capture the reputations of Computer Science conferences according to the publication patterns of these reputable groups.

We conducted experiments to investigate the relationship between P-scores, H-indices, Total Citations and Impact Factors in conference rankings. In particular, we wanted to estimate the degree of correlation between P-scores and citation based indices. The data set consists of 794 Computer Science[4] conferences for which the H-index was available on Google Scholar[5] in 2012. P-scores were calculated with data obtained from DBLP.[6] The number of publications, total number of citations and the Impact Factors were obtained from the Microsoft Academic Graph.[7] The reputation sources were selected with the randomization procedure described in the previous section, yielding the 12 departments listed in Table 1. The Impact Factors were calculated over a 5 year period and they correspond to the yearly average number of citations received by articles published in each conference during this period.

P-score and *H*-index behavior

We plotted normal probability plots for P-scores and H-indices in the CS conference data set to identify 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.

[4] 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.

[5] http://scholar.google.com.br

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

[7] http://academic.microsoft.com

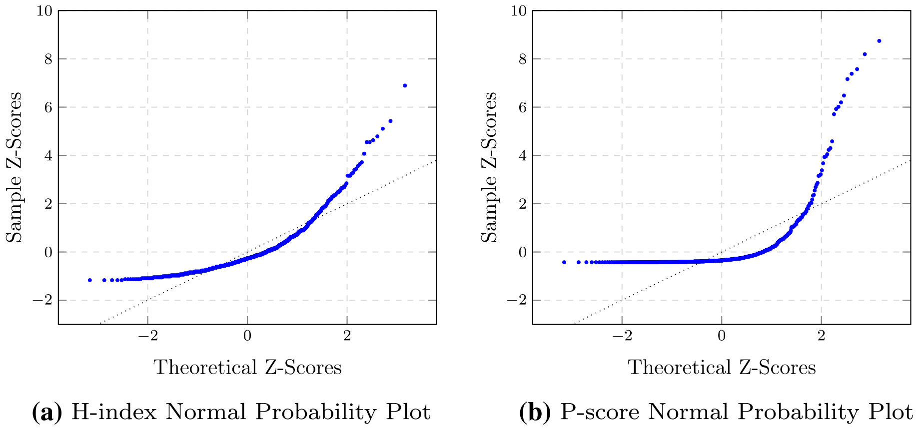


Fig. 3 Normal probability plots

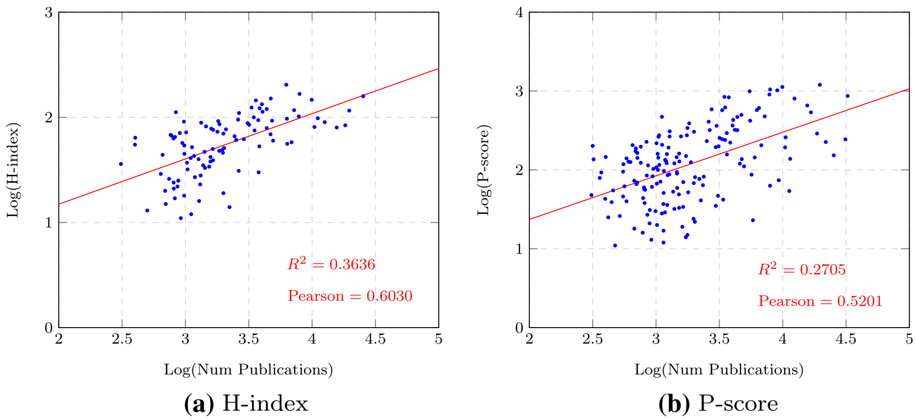


Fig. 4 Log(*H*-index) versus Log(Num Publications) for the CS conferences only considering conferences with more than 1000 publications and averaging the number of publications for conferences with identical indices

We compared theoretical and sample *Z*-scores for 794 quantiles (one for each data point) in Fig. 3a and b. 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. In fact, the values of *H*-indices and *P*-scores tend to increase with the number of publications N and their behavior is better described by a power-law model:

$$X_i = C \times N_i^a \tag{11}$$

where C and a are index-specific constants, X_i is the index value and N_i is the number of publications for observation i. Figure 4 presents the log-log plots for *P*-scores and *H*-indices versus the number of publications. A linear relationship between the logarithms indicate a power-law behavior. The estimated parameters for the constants of the power-law model in the best linear fit are $C = 2.0591$ and $a = 0.4303$ for the H-index, $C = 1.8498$ and $a = 0.5525$ for P-score.

The tendence of P-scores and H-indices to increase with the number of publications may hurt the comparison of venues with a largely different size, obfuscating less prolific conferences. To account for this fact, we also plot correlations for the normalized P-scores and H-indices using a measure called the Scientific Performance QuaLity (SPQL) level (Babić et al. 2016). Originally, the SPQL level only handles H-indices, but we also use it as a normalization procedure for P-score. The SPQL level can be defined as:

$$\mathrm{SPQL}(i) = 100 \cdot \frac{\tilde{X}_i}{X_i} \tag{12}$$

where $\tilde{X}_i$ is the observed value for the index in point i and X_i is the corresponding value in the best fit line. In our case, $X_i = 2.0591 \times N_i^{0.4303}$ for the H-index and $X_i = 1.8498 \times N_i^{0.5525}$ for P-score. This way, we have a measure that indicates the degree of deviation from average that is independent of the number of publications and allows the comparison of venues with different sizes.

P-score correlations

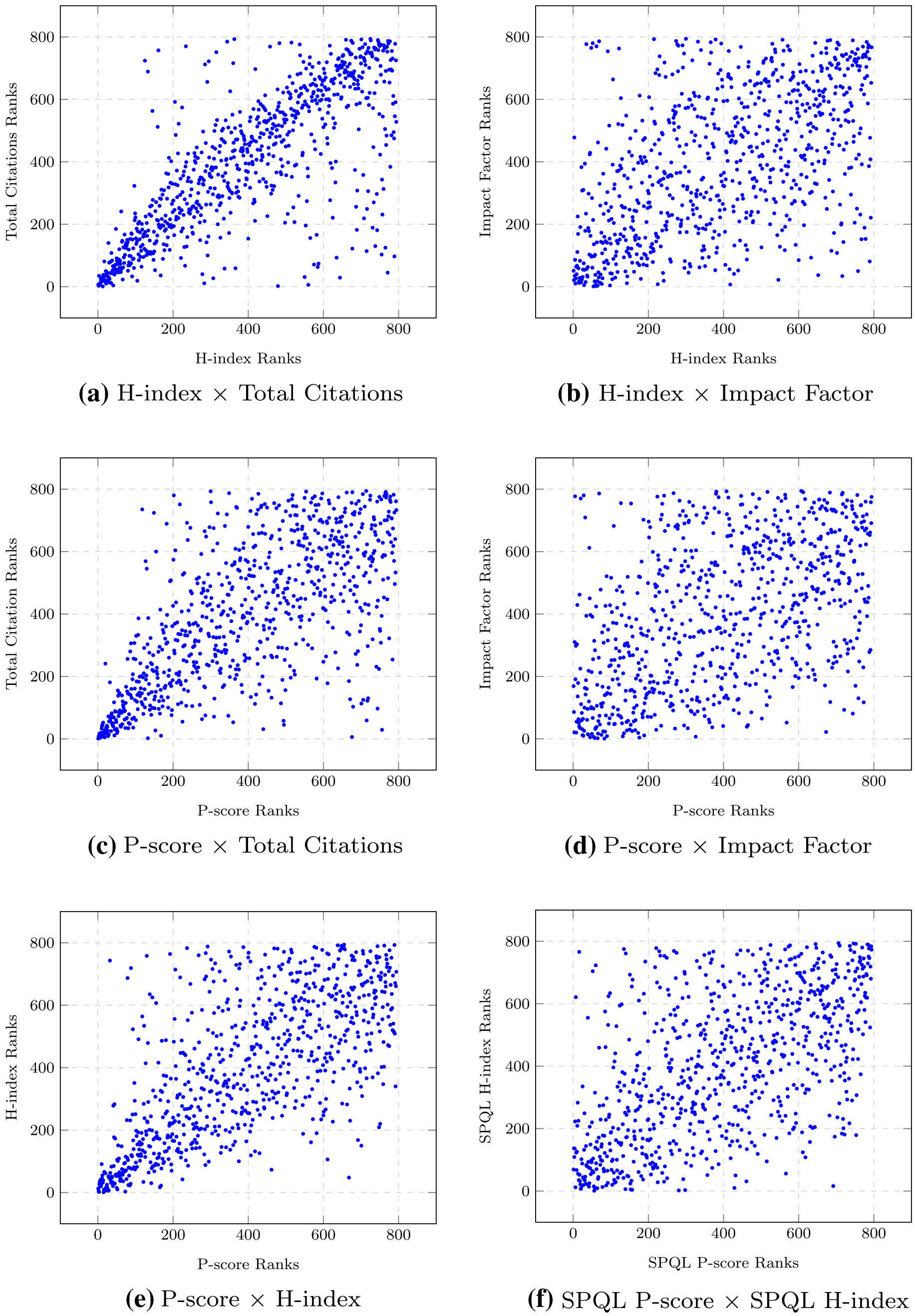
We studied the correlation between P-score and various citation-based indices in the context of Computer Science conference ranking to understand how reputation and popularity indices diverge, and to understand if the combined assessment of reputation and popularity indices can reveal interesting conferences that are not perceived in the rankings produced with a single index.

Figure 5 compares rankings of the 794 CS conferences in terms of their P-score, H-index, Total Citations and Impact Factors. 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). Also, Table 2 presents the correlations between indices in terms of their Kendall-Tau, Spearman and Pearson coefficients. The Spearman coefficient is essentially a Pearson Correlation of ranking positions. The Kendall-Tau is another correlation measure based on ranks that is deemed less sensitive to errors than the Spearman coefficient (Kendall 1955; Baeza-Yates and Ribeiro-Neto 2011). These correlation coefficients assume a value between − 1 and 1. A value of 1 indicates that two rankings are identical and a value of − 1 indicates that rankings are the inverse of each other, with values inbetween indicating partial correlation. A coefficient close to zero indicates no correlation.

The first fact to point out is that H-indices are significantly correlated to the Total Citations received by a conference but not as much to the Impact Factor, what is also observed in P-score. Both indices seem to be somewhat reliant on the number of publications, so it is not unexpected that they are less correlated to Impact Factors than to Total Citation counts.

Also, P-scores seem to be quite correlated to H-indices, assuming a Kendall-Tau of 0.5153. The Kendall-Tau is defined in terms of concordant and discordant pairs of ranks. P-scores agree aproximately 76% of the time with H-indices on which conference should be ranked above the other. However, normalized P-scores and H-indices in the form of the SPQL measure have a significantly smaller degree of correlation (a Kendall-Tau of 0.4148) showing that, once we control for the number of publications, P-scores and H-indices seem to be more clearly capturing different aspects of a conference's scientific impact.

In our assessment, we are especially interested in conferences for which popularity and reputation measures are most disagreeing. These would be conferences with high P-scores



(a) H-index × Total Citations

(b) H-index × Impact Factor

(c) P-score × Total Citations

(d) P-score × Impact Factor

(e) P-score × H-index

(f) SPQL P-score × SPQL H-index

Fig. 5 Scatter plots comparing the rankings produced with *P*-scores, *H*-indices, total citations and impact factors in a set of 794 conferences in computer science

Table 2 Correlations between P-scores, H-index, total citations and impact factors in a set of 794 conferences in computer science

First	Second	Kendall-Tau	Spearman	Pearson
H-index	Total citations	0.6141	0.7623	0.6848
H-index	Impact factor	0.3786	0.5306	0.5276
P-score	Total citations	0.5028	0.6751	0.7798
P-score	Impact factor	0.3353	0.4851	0.3120
P-score	H-index	0.5153	0.6939	0.7089
SPQL Pscore	SPQL H-index	0.4148	0.5822	0.4957

and low H-indices or high H-indices and low P-scores. These interesting outliers can both be captured by plotting H-indices against P-scores or the normalized version of the measures. We will proceed in the next Section with the assessment of P-scores versus H-indices.

Assessing conferences in CS

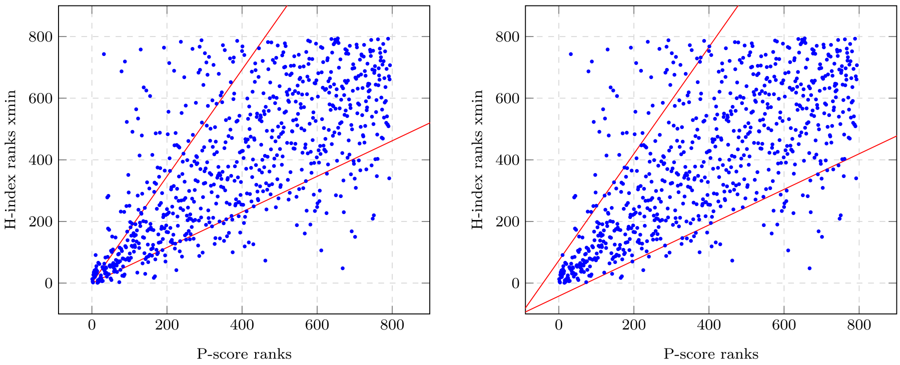
Since P-scores and H-indices are fairly correlated, most data points in our example are clustered close to the identity line, showing an apparent agreement between the indices. However, there are noticeable differences, particularly in those cases for which the P-score rank position is high and the H-index rank is low, or for which the H-index rank is high and the P-score rank is low. 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. These are the venues that are popular, but not very reputable in the sense that they are not endorsed by the reputation sources.

Delimitation of interest groups

We devised a strategy for delimiting the Top, Center and Bottom sets. It consists of drawing two line segments, starting from the origin, separated by an angle θ, and examining how the correlation between 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 Fig. 6a. 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 Fig. 6b. By doing so, we ensure that all data points near the origin are included in the Center Group.

Table 3 describes how the Kendall-Tau and Spearman coefficients vary for items in the Center Group, as we vary the angle θ, with the pivot positioned behind the origin as



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

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

Fig. 6 Scatter plots comparing *P*-score and *H*-index ranks, with lines describing the angle strategy for delimiting groups of interest

Table 3 Angle strategy Kendall-Tau in center group by varying the angle

Angle	Kendall-Tau	Spearman	Items in center group
90	0.5153	0.6939	794
80	0.5153	0.6939	794
70	0.5186	0.6989	793
60	0.5269	0.7094	790
50	0.5523	0.7341	777
40	0.5884	0.7611	749
30	0.6510	0.7998	689
20	0.7290	0.8295	576
10	0.8434	0.8481	345

in Fig. 6b. We notice that for $\theta = 30°$, 689 conferences are included in the Center Group (close to 87% of all conferences in the data set). What leaves only a handful of conferences to be analyzed individually outside the Center Group. Thus, in our subsequent analysis, we employed the angle strategy with $\theta = 30°$, since it provides a good trade off between correlation and coverage in our data set.

With the pivot positioned at $(-100, -100)$ and $\theta = 30°$, the Top Group ended with 56 conferences, the Center Group with 689 conferences, and the Bottom Group with 49 conferences. Table 4 presents the top 30 conferences, ranked by *P*-score, for each of the three groups and Fig. 7 shows the *P*-score $\times$ *H*-index scatter plots showing all conference names for the Bottom and Top 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

Table 4 Conferences (short names) ranked by *P*-score in the bottom, center and top groups

#	Top	Center	Bottom
1	IROS	ICRA	IPTPS
2	ICMCS	CHI	MOBIHOC
3	ISIT	CVPR	ACSAC
4	ICCD	AAAI	ICWS
5	CDC	NIPS	ISM
6	ICPP	DAC	ISMIR
7	WCNC	ICASSP	PIMRC
8	LCPC	STOC	BIBE
9	ISQED	FOCS	DFT
10	ISER	INTERSPEECH	FSKD
11	WACV	ICCAD	ICIDS
12	ITS	IJCAI	MMM
13	ICWSM	SODA	ECIS
14	AIED	INFOCOM	ISVLSI
15	GIS	ICML	IHI
16	CCCG	ICIP	MMVR
17	DCC	SIGMOD	ARC
18	HUMANOIDS	ICSE	VNC
19	LREC	IPPS	IAT
20	WAFR	ICCV	ICAI
21	WSDM	ICDE	ICA3PP
22	CICC	ACL	ANALCO
23	MMSP	ECCV	CNSR
24	CLUSTER	ISCA	ICEBE
25	CONLL	KDD	PAAP
26	CLOUD	SIGCOMM	PROMAS
27	MFCS	SC	WI
28	FSR	WWW	WKDD
29	ICAC	ICDCS	APSCC
30	WABI	DATE	CATA

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 research groups 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, and LCPC are conferences related to parallel processing that capture considerable attention from the top CS departments, as we can observe from their high *P*-scores. Similarly IROS, WAFR, and HUMANOIDS are conferences in Robotics that have high *P*-scores and relatively low *H*-indices. These conferences

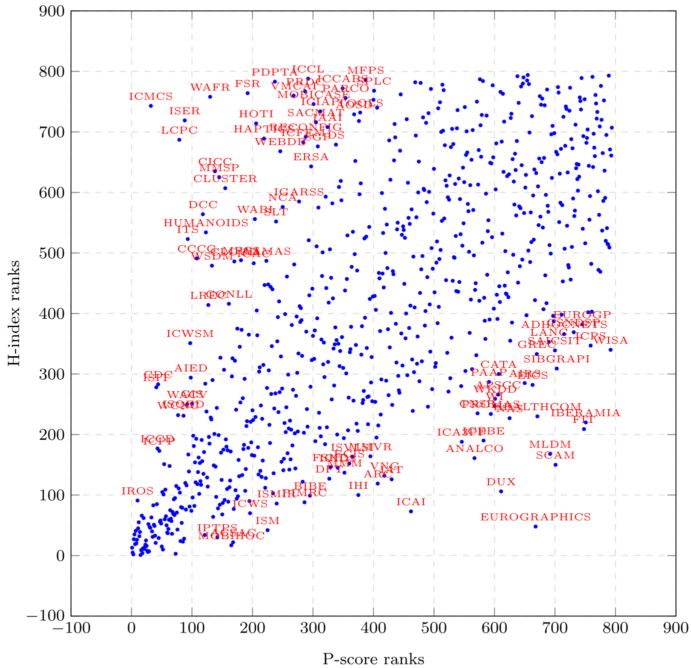


Fig. 7 P-score ranks × H-index ranks scatter plot with markings showing conference names

may end up being neglected by funding councils and committees in their assessment if we look only at the H-index, because these subareas do not receive as many citations as others.

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 BIBE, related to Biocomputing, IHI, MMVR, related to Health, and DFT, and VNC 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 being considered 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.

P-score on a per subarea basis

In the previous experiments, we showed how to use P-score as a complementary index to citation-based indicators, such as H-index, to assess the reputation of conferences in Computer Science. However, it is important to note that the index P-score captures the

Table 5 Conferences (short names) selected by *P*-score in three subareas of computer science

Natural language processing	Machine learning	Information retrieval
NIPS	NIPS	SIGIR
ACL	ICML	CIKM
EMNLP	CVPR	WWW
ICML	ECCV	SIGMOD
NAACL	ICCV	VLDB
INTERSPEECH	ICASSP	ICDE
CONLL	BMVC	TREC
ICASSP	IJCAI	WSDM
ICANN	AAAI	ECIR
IJCNN	ICANN	KDD

transference of reputation *from* source nodes *to* target and collateral nodes. Therefore, the choice of reputation sources defines the context where the reputation flows will be assessed and it can be adjusted to specific information needs. In this article, we used *P*-score to investigate Computer Science conference reputations from the point of view of a general reference set of computer scientists. But, one could be interested, for example, in determining which are the most relevant conferences in a given *subarea* of Computer Science. This is an important piece of information for funding councils when evaluating the productivity of individual researchers or research groups. In contrast to *H*-index, *P*-score is capable of producing results on a per subarea basis.

In Table 5, we show different sets of conferences in Computer Science highly ranked by *P*-score in three subareas: Natural Language Processing (NLP), Machine Learning (ML), and Information Retrieval (IR). In this experiment, we used as reputation sources the first ten individual researchers in each subarea according to Microsoft Academic.[8] Indeed, although the information available in Microsoft Academic was used in this experiment, there are other effective strategies to choose the reputation sources, such as: (1) to ask experts in each subarea to provide a list of the most reputable researchers and then aggregate the results using a voting method, (2) to use curated lists of impactful researchers in each subarea (when available), (3) to build a list only with awarded researchers in each subarea.

With a few reputation sources, *P*-score can correctly find reputable conferences on each subarea of knowledge, a problem that has been the subject of study in the literature (Wainer et al. 2013; Leydesdorff et al. 2013; Waltman and Eck 2013). Moreover, we can find conferences with low popularity but high reputation, such as CONLL in NLP, ICANN in ML, and WSDM in IR. On the other hand, well-known conferences that have both popularity and reputation within their respective subareas are also presented, as expected — e.g., ACL and INTERSPEECH in NLP, ICML and CVPR in ML, NIPS both in NLP and ML, SIGIR and CIKM in IR. This information may be useful to support decisions of research funding councils focused on specific subareas instead of broad areas of knowledge by assessing the subarea-specific *P*-scores in comparison with conference *H*-indices, as presented in Sect. 5.1. It is noteworthy that these subarea-specific rankings were produced with only

[8] http://academic.microsoft.com/authors

a small subset of reputable researchers to demonstrate the flexibility of *P*-scores, however they should not be taken authoritatively.

Conclusions

We have compared the *P*-scores and *H*-indices of 794 conferences in Computer Science to understand how *P*-score, a reputation based index that does not rely on citation data, can be used in conjunction with citation based indices to provide insight on scientific impact assessment. We found that *P*-score and *H*-index have a significant correlation reflected by a Kendall-Tau of approximately 0.5153. However, there are 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 distinguish three separate cases. First, when *P*-scores and *H*-indices agree, we find venues that have proportionate amounts of popularity and reputation among computer scientists. Second, when *P*-scores are high (better rank position) and *H*-indices are low (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 low (worse rank position) and *H*-indices are high, we have venues of good popularity (usually associated with large research communities outside Computer Science) that do not stand with a high reputation among computer scientists, that is, computer scientists do not publish frequently there. These differences indicate a complementary aspect to these indices. *P*-scores, used in conjunction with *H*-indices, allow us to distinguish the three 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.

There are several directions for further research using *P*-score. Given the flexibility of *P*-score to grasp the relative importance of different fields and subareas, an interesting topic for future research would be to analyze the development of research interest over time in different fields and countries. Also, considering that citation counts for newly published articles take at least a couple of years to develop and, since *P*-scores can be promptly obtained, future work could study the capacity of *P*-score to predict the future scientific impact of research articles, venues and authors. This capacity would be especially useful to help identifying promising research in advance.

Compliance with ethical standards

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

References

Babić, D., Kutlača, Đ., Živković, L., Štrbac, D., & Semenčenko, D. (2016). Evaluation of the quality of scientific performance of the selected countries of southeast europe. *Scientometrics*, *106*(1), 405–434.

Baeza-Yates, R., & Ribeiro-Neto, B. (2011). *Modern information retrieval: The concepts and technology behind search* (2nd ed.). Boston: Addison-Wesley Publishing Company.

Balaban, A. T. (2012). Positive and negative aspects of citation indices and journal impact factors. *Scientometrics*, *92*(2), 241–247.
Bar-Ilan, J. (2008). Which h-index?—A comparison of WoS, Scopus and Google Scholar. *Scientometrics*, *74*(2), 257–271.
Benevenuto, F., Laender, A. H., & Alves, B. L. (2016). The H-index paradox: Your coauthors have a higher H-index than you do. *Scientometrics*, *106*(1), 469–474.
Bollen, J., Rodriquez, M. A., & Van de Sompel, H. (2006). Journal status. *Scientometrics*, *69*(3), 669–687.
Bornmann, L., & Daniel, H. D. (2005). Does the h-index for ranking of scientists really work? *Scientometrics*, *65*(3), 391–392.
Bornmann, L., & Marx, W. (2011). The h-index as a research performance indicator. *European Science Editing*, *37*(3), 77–80.
Braun, T., Glänzel, W., & Schubert, A. (2006). A Hirsch-type index for journals. *Scientometrics*, *69*(1), 169–173.
Brin, S., & Page, L. (1998). The anatomy of a large scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, *30*(1–7), 107–117.
Chen, J., & Konstan, J. A. (2010). Conference paper selectivity and impact. *Communications of the ACM*, *53*(6), 79–83.
Ding, Y., & Cronin, B. (2011). Popular and/or prestigious? Measures of scholarly esteem. *Information Processing and Management*, *47*(1), 80–96.
Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics*, *69*(1), 131–152.
Egghe, L. (2008). 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.
Garfield, E. (1955). Citation indexes for science. *Science*, *122*(3159), 108–111.
Gonçalves, G. D., Figueiredo, F., Almeida, J. M., & Gonçalves, M. A. (2014). Characterizing scholar popularity: A case study in the computer science research community. In *Proceedings of the 14th ACM/IEEE-CS joint conference on digital libraries* (pp. 57–66).
Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, *102*(46), 16569–16572.
Hirsch, J. E. (2007). Does the h index have predictive power? *Proceedings of the National Academy of Sciences*, *104*(49), 19193–19198.
Järvelin, K., & Kekäläinen, J. (2002). Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems*, *20*(4), 422–446.
Kellner, A. W. A., & Ponciano, L. C. M. O. (2008). H-index in the Brazilian academy of sciences—Comments and concerns. *Anais da Academia Brasileira de Ciências*, *80*(4), 771–781.
Kendall, M. G. (1955). *Rank correlation methods* (2nd ed.). New York: Hafner Publishing Company.
Langville, A. N., & Meyer, C. D. (2006). *Google's pagerank and beyond: The science of search engine rankings*. Princeton: Princeton University Press.
Lee, D. H. (2019). Predictive power of conference-related factors on citation rates of conference papers. *Scientometrics*, *118*(1), 281–304. https://doi.org/10.1007/s11192-018-2943-z.
Leydesdorff, L. (2009). 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.
Leydesdorff, L., Zhou, P., & Bornmann, L. (2013). How can journal impact factors be normalized across fields of science? An assessment in terms of percentile ranks and fractional counts. *Journal of the American Society for Information Science and Technology*, *64*(1), 96–107.
Martins, W. S., Gonçalves, M. A., Laender, A. H., & Pappa, G. L. (2009). Learning to assess the quality of scientific conferences: A case study in computer science. In *Proceedings of the 9th ACM/IEEE-CS joint conference on digital libraries* (pp. 193–202).
Meyer, C. (1989). Stochastic complementation, uncoupling Markov chains, and the theory of nearly reducible systems. *SIAM Review*, *31*(2), 240–272.
Nane, T. (2015). 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).
Nature (2016). Time to remodel the journal impact factor. *Nature* *535*(7613), 466.
Nelakuditi, S., Gray, C., & Choudhury, R. R. (2011). 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.
Newman, M. (2010). *Networks: An introduction*. Oxford: Oxford University Press.
Pinski, G., & Narin, F. (1976). Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics. *Information Processing and Management*, *12*(5), 297–312.

Piwowar, H. A. (2013). Value all research products. *Nature*, *493*(7431), 159.

Ribas, S., Ribeiro-Neto, B., de Souza e Silva, E., Ueda, A. H., & Ziviani, N. (2015a). 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). New York, NY, USA: ACM.

Ribas, S., Ribeiro-Neto, B., Santos, R. L., de Souza e Silva, E., Ueda, A., & Ziviani, N. (2015b). Random walks on the reputation graph. In *Proceedings of the 2015 international conference on the theory of information retrieval, ICTIR '15* (pp. 181–190). New York, NY, USA: ACM.

Riikonen, P., & Vihinen, M. (2008). National research contributions: A case study on Finnish biomedical research. *Scientometrics*, *77*(2), 207–222.

Rossner, M., Van Epps, H., & Hill, E. (2007). Show me the data. *The Journal of Cell Biology*, *179*(6), 1091–1092.

Saha, S., Saint, S., & Christakis, D. A. (2003). Impact factor: A valid measure of journal quality? *Journal of the Medical Library Association*, *91*(1), 42–46.

Sun, Y., & Giles, C. L. (2007). Popularity weighted ranking for academic digital libraries. In *Proceedings of the 29th European conference on IR research, ECIR'07* (pp. 605–612). Berlin: Springer.

Wainer, J., Eckmann, M., Goldenstein, S., & Rocha, A. (2013). How productivity and impact differ across computer science subareas. *Communications of the ACM*, *56*(8), 67–73.

Waltman, L., & Eck, N. (2013). Source normalized indicators of citation impact: An overview of different approaches and an empirical comparison. *Scientometrics*, *96*(3), 699–716.

Wendl, M. C. (2007). H-index: However ranked, citations need context. *Nature*, *449*(7161), 403.

Yan, E., Ding, Y., & Sugimoto, C. R. (2011). P-rank: An indicator measuring prestige in heterogeneous scholarly networks. *Journal of the American Society for Information Science and Technology*, *62*(3), 467–477.