

Using Reference Groups to Assess Academic Productivity in Computer Science

Sabir Ribas
CS Dept, UFMG
Belo Horizonte, Brazil
sabir@dcc.ufmg.br

Berthier Ribeiro-Neto
CS Dept, UFMG & Google Inc
Belo Horizonte, Brazil
berthier@dcc.ufmg.br

Edmundo de Souza e Silva
COPPE, UFRJ
Rio de Janeiro, Brazil
edmundo@land.ufrj.br

Alberto Ueda
CS Dept, UFMG
Belo Horizonte, Brazil
ueda@dcc.ufmg.br

Nivio Ziviani
CS Dept, UFMG & Zunnit Tech
Belo Horizonte, Brazil
nivio@dcc.ufmg.br

ABSTRACT

In this paper we discuss the problem of how to assess academic productivity based on publication outputs. We are interested in knowing how well a research group in an area of knowledge is doing relatively to a pre-selected set of reference groups, where each group is composed by academics or researchers. To assess academic productivity we adopt a new metric we propose, which we call P-score. We use P-score, citation counts and H-Index to obtain rankings of researchers in Brazil. Experimental results using data from the area of Computer Science show that P-score outperforms citation counts and H-Index when assessed against the official ranking produced by the Brazilian National Research Council (CNPq). This is of our interest for two reasons. First, it suggests that citation-based metrics, despite wide adoption, can be improved upon. Second, contrary to citation-based metrics, the P-score metric does not require access to the content of publications to be computed.

Categories and Subject Descriptors

H.4 [Information Systems Applications]: Miscellaneous

Keywords

Academic productivity; Similarity; Reputation

1. INTRODUCTION

The assessment of academic productivity usually involves the association of metrics with the researchers or groups of researchers one wants to evaluate. Funding agencies, university officials, and department chairs are examples of entities interested in these metrics, as these have application in a variety of practical situations. There are also cases in

which one needs to compare researchers working on a same sub-area of knowledge, some examples are finding review peers, constructing program committees or compiling teams for grants.

Today, the most reliable and complete way to compare researchers is by compiling information on their academic output such as number of publications, citation based metrics, number of undergraduate and graduate students under supervision, number of advised masters and PhD theses, and participation in conferences and in technical committees. Some councils also use extensive surveys to compile qualitative information on features associated with the programs.

However, as compiling this information is not a simple task and takes a long time, it is a common procedure to use just citation data to gain quick insights into the productivity of research groups and academics. But, given that computing citation counts requires access to the contents of a large pool of publications, which is not always available, new and complementary metrics, such as P-score, are a necessity.

The notion of academic productivity is intrinsically associated with the notion of reputation. And although the concept of reputation lacks on definition, we can see it as a simple property of an individual or group which measures their academic impact in the world and which we can associate metrics with, as shown in [4, 12]. To measure the reputation of researchers, it is a common procedure to use the publication venues they publish in. Higher the impact of a venue, higher is considered the reputation of the researchers who publish in it. We use this idea of transferring reputation through publications to introduce a new metric called P-score (see [14] for details on the theoretical aspects of P-score).

The basic idea of P-score is to associate a reputation with publication venues based on the publication patterns of a reference group of researchers, in a given area of knowledge. Reference groups are composed of highly acclaimed academics of areas or sub-areas of knowledge, which then transfer reputation to the venues they publish in.

The evaluation procedure used in this work is based on two steps: (i) adopting as ground-truth the official evaluation of brazilian researchers given by the Brazilian National Research Council (CNPq); and (ii) use a well known metric in information retrieval to evaluate ranking approaches,

the DCG. Using this evaluation procedure, we present here experimental results where P-score outperforms the citation counts and H-Index metrics in the area of Computer Science, when assessed against the official ranking produced by the Brazilian National Research Council (CNPq).¹

The paper is organized as follows. In Section 2 we discuss the related work to assess academic productivity. In Section 3, we formally describe the P-score approach. In Section 4, we describe the dataset we built to work with and the evaluation procedure used to obtain our results. In Section 5, we discuss how to obtain a set of reference groups in Computer Science and apply it to make a comparison between P-score and two baselines — citation counts and H-Index. In Section 6 we discuss our conclusions and directions for future research on P-score.

2. RELATED WORK

One of the earliest metrics that aims to quantify the academic impact was the Garfield's Impact Factor [3]. Despite its widely usage, since it was proposed, in 1955, it has been criticized [15]. Many alternatives have been proposed in the literature, such as other citation-based metrics like H-Index [5], PageRank-like measures [16], and download-based measures [1]. There is a large amount of related works available on literature and existing metrics proposed in an attempt to alleviate possible issues of previous ones. But, as argued in [8], each metric has its own bias and there are both advantages and disadvantages associated with each one.

A recently proposed concept called Altmetrics [13] motivates the development of complementary metrics to evaluate research. Authors claim that citation-based metrics are useful, but not sufficient. One of the reasons is that metrics like the H-Index are slow, a work's first citation can take years.

In [4], the authors investigate the importance of various academic features to scholar popularity (defined as citation counts) and concluded that only two features are needed to explain all the variation in popularity across different scholars: (i) the number of publications and (ii) the average quality of the scholar's publication venues. In our work we use exactly these two features to formulate the model.

In [9], the authors propose a metric called ca-index, which present a novel approach for ranking researchers across multiple research areas. They argue that productivity indices should account for the singularities of the publication patterns of different research areas, in order to produce an unbiased assessment of the impact of academic output. The researcher's relative performance in multiple areas is aggregated into a unified ranking, the ca-index. The main difference between our work and this approach is that ca-index depends on citation data, while P-score does not.

The idea of reputation, instead of citations, was discussed in [11], where the authors propose a metric called peers' reputation. The metric ties the selectivity of a publication venue with the reputations of authors' institutions and argue that this metric is a better indicator of selectivity than acceptance ratio, and many conferences have similar or better peers' reputation than journals.

The main difference between our work and the aforementioned ones is the introduction of a set of reference groups, which none of them applies. The problem of finding comparable researchers, presented in [2], has many motivations.

¹<http://cnpq.br>

The work in [2] ranks authors by computing the similarity between a list of authors and a single reference. Our work is distinct given that (i) their method is based on a single reference author while ours is based on a set of reference groups, and (ii) they compute the similarity between authors through string distance while we employ a Markov model to obtain the metric of interest.

3. THE P-SCORE APPROACH

In this section we introduce a new metric called P-score. The basic idea of P-score is to associate a reputation with publication venues based on the publication patterns of a set of *reference groups* of researchers in a given area or sub-area of knowledge.

The reputation of a research group is strongly influenced by the reputation of its members, which is largely dependent on their publication records. P-score is based on the following assumptions:

1. A researcher or a group member conveys reputation to a venue proportionally to its own reputation.
2. The reputation of a researcher is proportional to the reputation of the venues in which he/she publishes.

Once a reference group in a given area is selected, the reputation of members in this group is transferred to the venues. A Markov chain can then be built from these ideas.

Figure 1, left side box, illustrates a Markov chain in which two research groups, ω_1 and ω_2 , publish papers in three venues, v_1 , v_2 and v_3 . The numbers in the arcs are the relative frequencies of the publications (e.g., fraction of the total number of papers published). The chain can be solved by a stochastic computation which associates steady state probabilities to each node in the chain. These probabilities are taken as weights associated with venues, such as ν_1 , ν_2 , and ν_3 , which we refer to as P-scores. The venue P-scores are used to compute a rank for each research author a_j we want to consider for evaluation purposes. Notice that usually there is no intersection between the set of authors in the reference groups ω_i and the set of researchers a_j we are comparing.

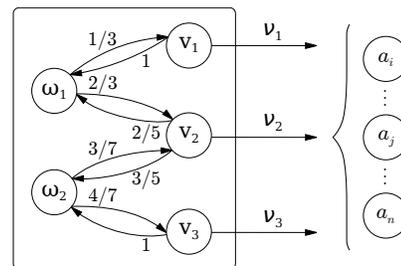


Figure 1: Example of a small P-score model

Before developing the model, we introduce some notation. Table 1 summarizes the notation and definitions used in this work. We use ω and j as indexes for research groups and the venues where they publish, respectively. The research groups used as reputation sources are referred to jointly as the *reference groups*. Consider a chosen set \mathcal{T} of reference groups, and let T be its cardinality. Let \mathcal{V} be the set of all venues v_j where the groups in \mathcal{T} publish, and V the total

number of venues in the set \mathcal{V} . Members of research group ω publish in subset $\mathcal{V}_\omega \subseteq \mathcal{V}$ with cardinality $V_\omega = |\mathcal{V}_\omega|$.

Table 1: Notation

\mathcal{T}	set of reference groups
T	cardinality of \mathcal{T}
ω	a research group in \mathcal{T}
\mathcal{V}	set of venues where the researchers in \mathcal{T} publish
V	cardinality of \mathcal{V}
\mathcal{V}_ω	set of venues where the researchers of group ω publish
V_ω	cardinality of \mathcal{V}_ω
v_j	the j^{th} venue where members of a group in \mathcal{T} publishes at
$N(\omega, v_j)$	total number of distinct papers published by group ω in venue v_j
$N(v_j)$	total number of papers published in venue v_j
$N(w)$	total number of publications of group w
$D(v_j)$	number of distinct authors publishing in venue v_j
γ_w	reputation of group $w \in \mathcal{T}$
ν_j	reputation of venue $v_j \in \mathcal{V}$
A	set of authors we want to compare

We define a function N that counts the papers published by research groups and the papers published at venues. Let $N(\omega, v_j)$ be the total number of *distinct papers* published by research group ω in venue v_j and let $N(v_j)$ and $N(w)$ be the total number of papers published in venue v_j and the total number of publications of group w during the observation period, respectively. That is:

$$N(w) = \sum_{j=1}^V N(\omega, v_j)$$

$$N(v_j) = \sum_{w=1}^T N(\omega, v_j)$$

From Assumption 1, the reputation of group w is defined as:

$$\gamma_w = \sum_{j=1}^V \nu_j \times \alpha_{wj} \quad (1)$$

where

$$\alpha_{wj} = \frac{N(\omega, v_j)}{N(v_j)} \quad (2)$$

is the fraction of publications of venue v_j that are from research group ω and V is the number of venues.

Let $D(v_j)$ be the number of distinct authors that publish in venue v_j . From Assumption 2, the reputation of venue v_j is defined as:

$$\nu_j = \sum_{w=1}^T \gamma_w \times \beta_{wj} \quad (3)$$

where

$$\beta_{wj} = d \times \frac{N(\omega, v_j)}{N(w)} + (1-d) \times \frac{D(v_j)}{\sum_k D(v_k)} \quad (4)$$

combines the fraction of publications of group ω that are from venue v_j and the fraction of distinct authors that publish in v_j . The intuition for this formulation is venues that receive publications from a small set of authors are most likely to have lower reputation, e.g. local workshops may receive a large amount of publications but the total number of distinct authors tend to be small. The parameter d

($0 \leq d \leq 1$) controls the relative importance between the volume of publications that v_j receives from a group ω and the total number of authors publishing there.

If $d = 1$ then the reputation of the publication venues is totally derived from the reference groups. If $d = 0$ then the reputation of the publication venues is totally derived from the amount of distinct authors (from reference groups or not) publishing there. We noticed that varying d does have a significant impact on venue weights, particularly when we considered all venues in Computer Science. In this work, we use $d = 0.75$ since it provides a good balance between reference groups and total number of distinct authors in a venue.

Let \mathbf{P} be a $(T+V) \times (T+V)$ square matrix such that element $p_{mn} = 0$ if either $m, n \leq T$ or $m, n \geq T$. In addition, $p_{mn} = \beta_{m, n-T}$ for $m \leq T, n > T$ and $p_{mn} = \alpha_{m-T, n}$ for $m > T, n \leq T$. Note that, since $\sum_{w=1}^T \alpha_{wj} = 1$ for all $1 \leq j \leq V$ and $\sum_{j=1}^V \beta_{wj} = 1$ for all $1 \leq w \leq T$ then \mathbf{P} defines a Markov chain. In addition, the Markov chain is periodic and has the following structure:

$$\mathbf{P} = \left[\begin{array}{ccc|ccc} \mathbf{0} & & & \mathbf{P}_{12} & & \\ & & & & & \\ & & & & & \\ \mathbf{P}_{21} & & & \mathbf{0} & & \end{array} \right] = \left[\begin{array}{ccc|ccc} 0 & \dots & 0 & \beta_{11} & \dots & \beta_{1V} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \beta_{T1} & \dots & \beta_{TV} \\ \hline \alpha_{11} & \dots & \alpha_{T1} & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \alpha_{1V} & \dots & \alpha_{TV} & 0 & \dots & 0 \end{array} \right]$$

From decomposition theory, see [10], we can obtain values for ranking the reference groups by solving:

$$\gamma = \gamma \mathbf{P}' \quad (5)$$

where $\mathbf{P}' = \mathbf{P}_{12} \times \mathbf{P}_{21}$ is a stochastic matrix and $\gamma = \langle \gamma_1, \dots, \gamma_T \rangle$. Note that matrix \mathbf{P}' has dimension $T \times T$ only and can be easily solved by standard Markov chain techniques. Then, from Equation (1) we obtain the reputation of all venues where the reference groups publish.

$$\nu = \gamma \times \mathbf{P}_{12} \quad (6)$$

Once the vector ν of P-scores has been computed, we can easily compute a rank \mathcal{R} for each author a in a set of authors A we want to compare as:

$$\mathcal{R}(a \in A) = \frac{S_a}{\max_{i \in A} \{S_i\}} \quad (7)$$

where S_a ($a \in A$) is a weighted sum of P-scores associated with author a in set A , computed as:

$$S_a = \sum_{j=1}^V \nu_j \times N(a, v_j) \quad (8)$$

where ν_j is the weight (or P-score value) of venue v_j according to ν and $N(a, v_j)$ is the total number of publications from author a in venue v_j .

4. METHODOLOGY

In this section, we discuss the methodology used to obtain the experimental results described in Section 5 for evaluating the P-score approach to rank researchers. In this case, we aim to answer the following research question:

RQ: *Can we assess the productivity of academics using their similarity with a reference set of pre-selected well-known researchers? How does it compare with classic citation-based metrics?*

4.1 Dataset Description

To answer the research question stated in the previous section, we built our own dataset² based on publication lists from DBLP. The citation data used in this paper for validation purposes were collected from Google Scholar.

The data collection process consists of four steps: (i) collect the list of research groups to compose the analysis; (ii) retrieve the list of members/authors of each research group; (iii) match each author to the corresponding entity(ies) in a repository of publications; (iv) identify the publication venues of each publication of the authors. The list of researchers and their corresponding publications are sufficient to compute the P-score metric. There is no need to access the contents of the publications.

In here, we focus on the area of Computer Science basically because the publication patterns in the area are well described by good sources such as DBLP. However, we notice that P-score can be applied to any area of knowledge, as long as listings of the publications of the groups we use as reference are made available.

4.2 Evaluation Procedure

The evaluation procedure used in this work is based on the evaluation procedure adopted in [9]. The ground-truth is an official evaluation of Brazilian researchers and the metric used is the normalized DCG, as discussed in this section.

4.2.1 Ground-Truth: CNPq Productivity Levels

There is no official world wide evaluation of researchers in Computer Science. Thus, we have adopted as ground-truth an official evaluation from Brazil, the CNPq productivity levels of researchers. The CNPq is a well established agency dedicated to the promotion of scientific and technological research in Brazil. One of the roles of CNPq is to provide productivity grants to researchers. These grants aim at stimulating high quality research in the country.

CNPq classifies the researchers who have received grants in different levels of productivity, which are 1A, 1B, 1C, 1D, and 2, in descending order of prestige. This classification follows a set of specific criteria, such as academic output, contribution to formation of human resources, academic leadership, among others. To illustrate, CNPq currently groups researchers in Computer Science by productivity levels as shown in Table 2.

In our work, these productivity levels attributed by CNPq are used as relevance weights of the evaluation metric we adopted, as we now discuss.

4.2.2 Metric: Discounted Cumulative Gain

As in [9], to compare our results with citation counts and H-Index we use the Discounted Cumulative Gain (DCG) metric [6]. DCG adopts a non-binary notion of relevance, by assessing a given ranking based upon a graded scale, from less relevant to more relevant. This metric also uses a log-based discount factor that reduces the impact of the gain as we move lower in the ranking. Let g_i be the non-binary

²The P-score framework and all data used in our experiments are available at <http://pscore.latin.dcc.ufmg.br>

Table 2: Distribution of researchers in CNPq productivity levels in Computer Science (CS)

Level	Researchers	
	CS	All Areas
1A	23	1320
1B	22	1308
1C	31	1376
1D	70	2386
2	244	7933
Total	390	14323

relevance grade associated with the item ranked at the i -th position. The DCG at a rank position k is:

$$DCG@k = \sum_{i=1}^k \frac{2^{g_i} - 1}{\log_2(i + 1)} \quad (9)$$

To apply this metric to evaluate our results, as the procedure in [9], we use a graded relevance scale based upon each researcher’s classification according to CNPq, as shown in previous section. To bind the result within the interval $[0,1]$, we use the normalized version of DCG, denoted nDCG, which is obtained by dividing the DCG@ k value given in previous equation by the maximum possible value at the same rank cutoff k . In Table 3 we present the map we used.

Table 3: CNPq productivity levels and respective relevance weights in nDCG

CNPq Level	1A	1B	1C	1D	2
Relevance in nDCG	5	4	3	2	1

5. EXPERIMENTAL RESULTS

In this section, we investigate how P-score performs when compared to existing metrics. It specifically addresses the research question RQ stated in the beginning of Section 4 on the assessment of academics productivity using their similarity with reference groups.

5.1 Reference Groups in Computer Science

We now discuss the reason for using reference groups and how to choose them. We use reference groups because they provide a natural way to produce relative comparisons. By computing the similarity of the research output of a group of authors with reference groups, we can get an insight about the productivity of these authors in a certain area or sub-area of knowledge.

The choice of a reference group depends on what ones want to compare. In here, our objective is to compare Brazilian researchers in Computer Science among themselves but using as reference the top researchers in Computer Science. To determine the top researchers we rely on the rankings from Microsoft Academic Search (MAS).

MAS distinguishes 24 sub-areas in Computer Science ranging from Algorithms & Theory to the World Wide Web, as illustrated in Table 4. Because of this, we model the problem as a Markov chain in which each of the 24 sub-areas is a reference node. Thus, the reference set is composed of

Table 4: Sub-areas of Computer Science according to Microsoft Academic Search

#	Subarea
1	Algorithms & Theory
2	Artificial Intelligence
3	Bioinformatics & Computational Biology
4	Computer Vision
5	Data Mining
6	Databases
7	Distributed & Parallel Computing
8	Graphics
9	Hardware & Architecture
10	Human-Computer Interaction
11	Information Retrieval
12	Machine Learning & Pattern Recognition
13	Multimedia
14	Natural Language & Speech
15	Networks & Communications
17	Operating Systems
18	Programming Languages
19	Real-Time & Embedded Systems
20	Scientific Computing
21	Security & Privacy
22	Simulation
23	Software Engineering
24	World Wide Web

24 reference groups, one for each sub-area. The publication output of each reference group is the union of the publications of the top 10 researchers in that area according to MAS. By doing so, we guarantee that the reference set is all of high reputation.

To illustrate, Table 5 presents the reference groups for three sub-areas, which are Databases, Information Retrieval and Networks & Communications. By using P-score and a set of reference groups, like those discussed in this section, we can sort publication venues and authors in any area or sub-area of knowledge.

Table 5: Reference groups for the sub-areas of Databases, Information Retrieval and Networks

Databases reference group
Hector Garcia-Molina, Alon Halevy, Jennifer Widom, David Dewitt, Michael Stonebraker, Jeffrey D. Ullman, Michael J. Carey, Dan Suciu, Rakesh Agrawal, Serge Abiteboul
Information Retrieval reference group
W. Bruce Croft, Gerard Salton, Ellen Voorhees, Chris Buckley, Stephen E. Robertson, Jamie Callan, Susan Dumais, James Allan, Hsinchun Chen, Justin Zobel
Networks and Communications reference group
Deborah Estrin, Scott J. Shenker, Donald F. Towsley, David E. Culler, Sally Floyd, Hari Balakrishnan, Mario Gerla, Randy H. Katz, Ion Stoica, Ian F. Akyildiz

5.2 Comparison with Citations and H-Index

In this section we experiment with the problem of ranking Brazilian researchers in Computer Science. The dataset used for experimentation is composed of 655 professors associated

with the top 25 graduate programs in CS in Brazil, according to CAPES (see [7]). Of these, 390 receive individual grants classified in 5 levels as shown in Table 2 and are assigned relevance scores as illustrated in Table 3. The remaining 265 researchers are not considered here.

We investigate how P-score performs when compared to existing metrics. Specifically, we compare a P-score ranking of Brazilian researchers with analogous ranking produced using citation counts and H-Index. We collected the citation-data from Google Scholar.

Figure 2 displays DCG curves for three metrics: P-score, H-Index and citation counts. The ground-truth is the classification of researchers presented in Table 3. For simplicity only the top 100 positions in the rankings are shown. The results indicate that P-score consistently outperforms both H-Index and citation counts.

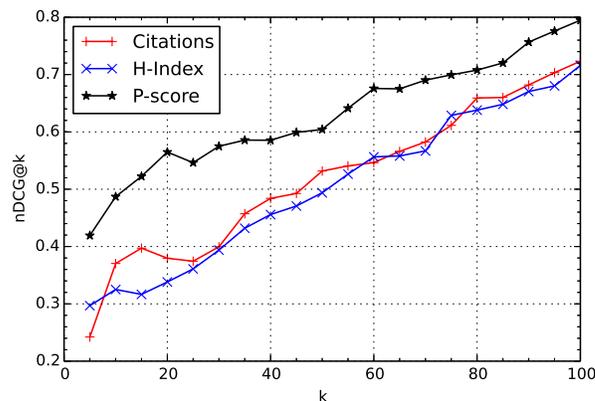


Figure 2: Comparison between researcher rankings based on Citations, H-Index and P-score.

To better appreciate these results, let us look in more detail at the ground-truth. To classify researchers in productivity levels, the Brazilian National Research Council (CNPq) committee makes a deep evaluation of their academic profile. This evaluation is based on a set of specific criteria such as academic output, contribution to formation of human resources, academic leadership, among others. Further, citation-based metrics have a high weight in the evaluation process. Thus, it was expected that citation-based metrics, such as citation counts or H-Index, would produce good rankings when compared with CNPq classifications. What was not expected is that P-score would produce better results.

While it requires further investigation, our interpretation is that the ranking of venues produced by P-score is better than the rankings of venues produced by both citation counts and H-Index, probably because reputation frequently overlaps with citation counts, plus the combination P-score does with information on publication distributions. This combination leads to improved results.

Given that P-score depends on a set of reference groups, a question may arise about the stability of the produced rankings. For instance if we consider a slightly different set of reference groups, how much will the outcome change? We investigated this question by running experiments to analyze rankings of publication venues and the results shows that the

venue rankings are stable when we slightly vary reference sets. Given that researcher ranks are computed from the venue weights, their ranks should also be stable. This is a valid question that needs further investigation.

Another question is whether the minimum size of the reference groups stands for a sub-area. While we do not present experiments here, we have experimentation that shows that 10 researchers in each sub-area is enough.

We recognize that our results were produced in a limited context (which is the context of Brazilian researchers). However, we believe that there is nothing in particular in the CNPq evaluation, our ground-truth, that seems to be different from other governmental research councils evaluations in other countries. Therefore, nothing suggests that our method can not be applied in other contexts.

6. CONCLUSIONS

In this work we investigated the problem of assessing the academic productivity of a group of Brazilian researchers in light of the productivity of well-known researchers from 24 sub-areas of Computer Science ranging from Algorithms & Theory to Software Engineering and the World Wide Web. From each sub-area we selected the top 10 most productive researchers according to Microsoft Academic Search, comprising a total of 240 researchers — our reference set. Using a stochastic model we proposed, called P-score, we transferred reputation from these 240 researchers to the venues they publish in, a process that led to a vector of venue weights. These weights were then used to rank the Brazilian researchers based on their distributions of publications (in the venues weighted with P-scores). We compared the results with rankings of the same researchers based on H-Index and citation counts.

Our experimental results indicated that P-score outperforms both H-Index and citation counts throughout the whole range of ranking. This is a bit surprising given our ground-truth, a ranking of Brazilian researchers by the funding agency CNPq, relies heavily on citation counts. At this point, our interpretation is that P-score led to better results because it combines information on reputation with information on the distribution of publications.

While our results are preliminary and restricted to a set of Brazilian researchers, there is nothing in our approach that is specific to the dataset used for experimentation and we did not tune the P-score model in any particular way to fit the data. Thus, it is reasonable to expect that the model might be of value in other contexts, other datasets, other countries — a hypothesis we intend to explore in short term future works.

ACKNOWLEDGEMENTS

This work was partially sponsored by the Brazilian National Institute of Science and Technology for the Web (MCT/CNPq 573871/2008-6) and the authors' individual grants and scholarships from CNPq, FAPEMIG and FAPERJ.

7. REFERENCES

- [1] J. Bollen, H. van de Sompel, J. Smith, and R. Luce. Toward alternative metrics of journal impact: A comparison of download and citation data. *IP&M*, 41(6):1419–1440, 2005.
- [2] G. Cormode, S. Muthukrishnan, and J. Yan. People like us: mining scholarly data for comparable researchers. In *WWW, Companion Volume*, pages 1227–1232, 2014.
- [3] E. Garfield. Citation indexes for science. *Science*, 122(3159):108–111, 1955.
- [4] G. D. Gonçalves, F. Figueiredo, J. M. Almeida, and M. A. Gonçalves. Characterizing scholar popularity: A case study in the computer science research community. In *JCDL*, pages 57–66, 2014.
- [5] J. Hirsch. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, pages 16569–16572, 2005.
- [6] K. Järvelin and J. Kekäläinen. Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems*, 20(4):422–446, 2002.
- [7] A. H. F. Laender, C. J. P. de Lucena, J. C. Maldonado, E. de Souza e Silva, and N. Ziviani. Assessing the research and education quality of the top brazilian computer science graduate programs. *ACM SIGCSE Bulletin*, 40(2):135–145, 2008.
- [8] L. Leydesdorff. 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.
- [9] H. Lima, T. H. P. Silva, M. M. Moro, R. L. T. Santos, W. M. Jr., and A. H. F. Laender. Aggregating productivity indices for ranking researchers across multiple areas. In *JCDL*, pages 97–106, 2013.
- [10] C. Meyer. Stochastic complementation, uncoupling Markov chains, and the theory of nearly reducible systems. *SIAM Review*, 31(2):240–272, 1989.
- [11] S. Nelakuditi, C. Gray, and R. R. Choudhury. Snap judgement of publication quality: how to convince a dean that you are a good researcher. *Mobile Computing and Commun. Review*, 15(2):20–23, 2011.
- [12] J. P. Ostriker, P. W. Holland, C. V. Kuh, and e. James A. Voytuk. *A Guide to the Methodology of the National Research Council Assessment of Doctorate Programs*. The National Academies Press, 2009.
- [13] H. Piwowar. Altmetrics: Value all research products. *Nature*, 493(7431):159–159, 2013.
- [14] S. Ribas, B. Ribeiro-Neto, E. de Souza e Silva, A. Ueda, and N. Ziviani. P-score: A publication-based metric for academic productivity. *CoRR*, 2015.
- [15] S. Saha, S. Saint, and D. Christakis. Impact factor: a valid measure of journal quality? *Journal of the Medical Library Association*, 91(1):42–46, 2003.
- [16] S. Yan and D. Lee. Toward alternative measures for ranking venues: a case of database research community. In *JCDL*, pages 235–244, 2007.