C³-index: Revisiting Author's Performance Measure

Dinesh Pradhan, Partha Sarathi Paul, Umesh Maheswari, Subrata Nandi
Department of CSE, NIT Durgapur, India
{dineshkrp,mtc0113,umeshmaheswari7,subrata.nandi}@gmail.com
Tanmoy Chakraborty
University of Maryland, College Park, USA
tanchak@umiacs.umd.edu

ABSTRACT

Author ranking indices, like h-index and its variants, fail to resolve ties while ranking authors with low index values (major volume including the young ones). In this work we leverage the citations as well as collaboration profile of an author in a novel way using a weighted multi-layered network and propose a page-rank variant to obtain a new author performance measure, C³-index. Experiments on a massive publication dataset reveal several interesting characteristics of our metric: (i) we observe that C³-index is consistent over time, (ii) C³-index has high potential to break ties among low rank authors, (iii) C³-index can be used to predict future achievers at the early stage of their career.

CCS Concepts

• Information systems → Rank aggregation; • Human-centered computing → Social networks; • Applied computing → Digital libraries and archives;

Keywords

Author Ranking; Citation Network; h-index

1. INTRODUCTION

Easy access of publications via web has increased their visibility, leading the volume of authors and their publications increased exponentially, especially in computer science (CS) domain, during last decade [5] and made ranking authors harder. An effective index may help nominating an outstanding researcher for award, allocating research grants, etc. One may ask – Is it possible to design an evaluation index for authors by combining multiple features, such as citation count, impact of coauthors, citing authors’ profiles, etc such that not only high performers but also performance of the beginners (including young researchers) can be quantified unambiguously?

Most popular way of ranking authors is based on citations their works receive from peers. Based on citation count, several elegant yet simple indices exist: h-index, g-index, rank-citation index [1, 4], and so on. A closer look in Fig. 1(a) reveals a heavily skewed distribution of CS author count corresponding to the h-index they achieve as observed over a decade spanning from 1998 to 2008. In 2008, ≈80% authors have h-index only upto 2; max-scale 43. The drift of h-index of 1998 author set observed in 2008 is also very small (Fig. 1(b)). The same holds true for other derivatives of h-index. Hence, such indices are unable to provide sufficient resolution to bulk authors (having low index) including young promising researchers who possibly have published few good papers and would receive enough citations afterwards [2].

The limitations of citation count can be compensated by considering other features like (a) co-authors’ profile - here an author’s score is directly proportional to her co-authors’ score, assuming that high-performers will publish with promising ones, (b) citing authors’ profile - here an author’s score is directly proportional to citing authors’ score, assuming that high-profile authors refer to quality papers. In contrast to h-index and its derivatives, there are approaches which model above features as author-author collaboration and author-author citation networks and applied page-rank based algo-

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

WebSci ’16 May 22-25, 2016, Hanover, Germany
© 2016 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-4208-7/16/05.
DOI: http://dx.doi.org/10.1145/2908131.2908185
rithms [3] to rank authors. Obtaining a scoring function by modeling and combining multiple features is tricky; and unfortunately, multi-layered modeling has not been tried much. Moreover, existing literature broadly concerned on the dynamics of high and medium profile authors [4], and strangely ignored the bottom-liners who represent the bulk.

In this work we model citations of papers and authors as well as coauthorship profile of authors in a novel way using a weighted multi-layered graph and propose a page-rank variant to rank authors. Our score is consistent over years and can be used for ranking authors and predict early risers. Next section explains the dataset, the network model and the ranking algorithm followed by the results.

2. EXPERIMENT SETUP AND RESULTS

Dataset. We use the dataset available in Arnetminer Project [5]. After preprocessing, we consider 1,421,121 papers and 833,306 authors respectively spanning years 1960 – 2008. While considering the impact of an author till year T we only consider the evidences from our dataset till year T.

Network Model. To utilize all the above features in parallel, we use network model shown in Fig 1(c).

Measuring $C^3$-index. The strategy we propose is called $C^3$-index (abbreviation of paper-paper Citations and author-author Citations and author-author Collaborations), where the author score $C(t)$ at iteration $t$ is obtained as the normalized sum described by: $C(j) = (1 - \theta) + \theta \times (ACI_j(t) + AAI_j(t) + PCI_j(t))$. Here, $\theta$, set to 0.5 in our experiments, is the damping factor for the page-rank based strategy. The component scores, Author Citation Index (ACI), Author coAuthorship Index (AAI) and Paper Citation Index (PCI) are obtained in iteration $t$ using:

$$ACI_j(t) = (1 - \theta) + \theta \times \sum_{A_k \in C(A_j)} \frac{ACI_k(t-1)}{\text{outdeg}(A_k)}$$

$$AAI_j(t) = \sum_{A_k \in C(A_j)} \frac{AAI_k(t-1)}{\text{deg}(A_k)}$$

$$PCI_j(t) = \left( C^3(t-1) \right)^\alpha \sum_{P_k \in P(A_j)} \frac{PQI_k(t-1)}{\sum_{A_k \in A(P_k)} \left( C^3(t-1) \right)^\alpha}$$

where $PQI$, refers to as Paper Quality Index for a paper, is obtained as: $PQI(i) = (1 - \theta) + \theta \times \sum_{P_k \in C(P)} \frac{PQI_k(t-1)}{\text{outdeg}(P_k)}$

Here, $P(A_j)$ is set of papers of author $A_j$, $C(A_j)$ is set of authors citing author $A_j$, and $C(A_i)$ is set of authors coauthoring with $A_i$. Parameter $\alpha$ decides the way credit a paper being shared among coauthors. We set $\alpha$ to 0, meaning that all the coauthors will receive equal share. But other values of $\alpha$ may be tried, the credit then will be shared based on current $C^3$-Index of the coauthors.

Results. Scatter-plots in Fig. 2(a) show $C^3$-index for authors against their h-index for citation dataset relevant till 1998 (Inset: till 2008). We observe large pile of authors for each h-index values; however, one can notice that the authors bearing same h-index are sufficiently dispersed along the vertical scale. This indicate that $C^3$-index may be used to break the ties between authors having same h-index. Moreover, large cluster close to the diagonal indicate strong reasonable correlation among $C^3$-index and h-index.

To understand the inconsistent points of the diagram we choose eight points from Fig. 2(a) and find their h-index and $C^3$-index component scores (refer to Table 1). We observe h-index and $C^3$-index component scores of eight authors from 1998 author pool (In- h-index from 1998 to 2008: same for 2008)

| Author       | h-index | ACI | AAI | PCI |
|--------------|---------|-----|-----|-----|
| B. Reilly    | 1       | 0.49 | 0.68 | 0.74 |
| B Shomstein (A) | 13 | 23.12 | 18.12 | 13.42 |
| G. Rosenberg | 4       | 2.94 | 14.44 | 9.21 |
| H. V. Jagellish (B) | 11 | 0.50 | 5.70 | 9.24 |
| M. Shetano    | 4       | 0.78 | 0.64 | 0.58 |
| Ronald L. Rivera (C) | 9 | 39.58 | 28.07 | 11.17 |
| Tova Milo (D) | 2       | 0.44 | 8.29 | 6.24 |

3. CONCLUSION

The proposed score successfully resolves ambiguity among low profile authors which we believe is a major contribution. The impact of growth/failure/saturation of an author score needs to be studied more systematically with respect to features like topical influence in which the author primarily works. A thorough investigation is required to show the universality of our findings for other domains. One can think of categorizing the authors based on their future prospect.

4. REFERENCES

[1] S. Alonso et al. h-index: A review focused in its variants, computation and standardization for different scientific fields. JOI, 3(4):273 – 289, 2009.
[2] T. Chakraborty et al. On the categorization of scientific citation profiles in computer science. Comm. ACM, 58(9):82–90, Aug. 2015.
[3] M. Nykl et al. Pagerank variants in the evaluation of citation networks. JOI, 8(3):683 – 692, 2014.
[4] A. M. Petersen et al. Statistical regularities in the rank-citation profile of scientists. Sci. Rep., 1:181, 2011.
[5] J. Tang et al. Arnetminer: Extraction and mining of academic social networks. In SIGKDD, pages 990–998, 2008.