Research Contribution as a Measure of Influence

[Extended Abstract]

Lais M. A. Rocha, Mirella M. Moro
Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
{laismota,mirella}@dcc.ufmg.br

ABSTRACT

We propose the 3c-index that measures the influence degree of researchers by evaluating the links they establish between communities. We evaluate its performance against well known metrics. The results show 3c-index outperforms them in most cases and can be employed as a complementary metric to assess researchers' productivity.

CCS Concepts

•Information systems → Content ranking; •Human-centered computing → Social network analysis;

Keywords

Bibliometry; Research performance; Ranking strategy

1. INTRODUCTION

Studies have analyzed academic social networks to uncover information about their participants (researchers, professors, etc.) and relationships (advirsorship, coauthorship, etc.). They have considered performance of each researcher individually as well as measured the influence degree among research areas and contributions from researchers, providing meaningful visions of academia [1, 6, 7, 8].

We focus on the propagation of influence through the concept of *communities* formed by researchers who share common interests. Specifically, each researcher has a base community where he/she presents greater influence. Then, we follow the ideas of diversity and novelty [2] and consider that when a researcher works on a community (besides the base one), he/she transfers new knowledge to that community, then increasing the overall research quality. Our goal is to measure the *influence degree* by evaluating such links between different communities exploring two social concepts: closure as the potential knowledge acquired and brokerage as the potential for sharing information. We then propose an indicator, the **3c-index** (*Cross-Community Contribution*), to aggregate the score of the researcher in the base community and the influence transfer among communities.

SIGMOD/PODS'16 June 26 - July 01, 2016, San Francisco, CA, USA

ACM ISBN 978-1-4503-3531-7/16/06.

DOI: http://dx.doi.org/10.1145/2882903.2914834

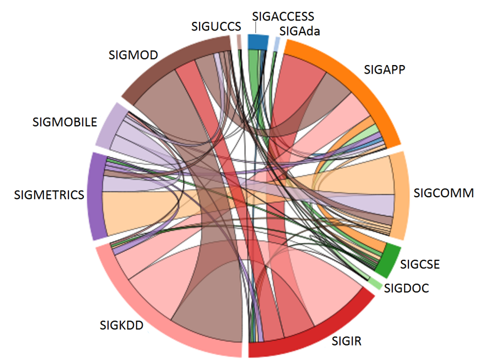


Figure 1: Knowledge transfer among SIGs showing, for instance, a significant transfer between SIGMOD-SIGKDD and a smaller one between SIGIR-SIGMOD.

2. RELATED WORK

In academic networks, ranking researchers on their importance is hard. One of the first studies on such importance is Granovetter's [3], which introduced the idea of *weak ties* as links that join different groups by building bridges within the network. Then, Newman [7] measures the influence on the information flow among individuals through high scores of *brokering*. Likewise, Burt [2] calls *brokers* the people who build social capital through structural holes in the networks. For ranking, Haveliwala [4] measures domain-specific importance using biased PageRank vectors, and Lima et al. [5] create a generic strategy for researchers from multiple areas by projecting productivity under a single perspective.

Overall, the aforementioned studies (among many others) emphasize the importance of building bridges and connecting distant network nodes. In academic social networks, researchers who connect different groups should bring more influence to those groups. Hence, this work aims at measuring such influence by considering his/her specific area and the knowledge transfer to other areas.

3. TECHNICAL OVERVIEW

In general, researchers have one or a group of areas of expertise. The knowledge transfer among research areas is fundamental for improving Science, because it enables the application of well-known and proved ideas from one context to solve problems on other domains of knowledge. Here, we define the **3c-index** as a metric that relates the knowledge transfer within different communities. This new index measures the influence of researchers according to their specialities (defined as influence degree) to other contexts.

The *3c-index* quantifies the influence degree in terms of percentile ranks and by rewarding important links that act as bridges. Formally, let p_i^c be the percentile ranking for a researcher i in the community c defined by: $p_i^c = \frac{l_i^c + 0.5e_i^c}{N^c}$, where N^c is the number of researchers in community c, l_i^c and e_i^c are the number of such researchers with a score lower than or equal to that of researcher i, respectively. For example, in a community with 100 researchers, the researcher at position 10^{th} in the rank has $l_i^c = 89$ and $e_i^c = 1$, then producing a percentile rank of 90%. After calculating the p_i^c for all communities in which a researcher publishes, we define the *base community* b_i as the one with the researcher's largest percentile, i.e., $b_i = \arg\max_{c \in C} p_i^c$. Then, the influence degree $infd_i^c$ is defined as the difference between the percentiles attained, i.e., $infd_i^c = b_i - p_i^c$. Here, a community corresponds to all authors who published in a SIG (ACM Special Interest Groups) (Figure 1), and $infd_i^c = 0$ for the base community because it has no knowledge *tranfer*.

Finally, *3c-index* combines the scores from the base community to the contribution across communities. For a set of communities, a normalization factor accounts for their distinct profiles (e.g., size, number of citations, etc). As Lima et al. [5], we consider the percentile ranks of researchers *mapped* to their base community. The *3c-index* of a researcher i is then defined as $3c\text{-}index(i) = f_b(b_i) + \sum_{c \in C}(b_i - p_i^c) f_b(p_i^c)$, where $f_b(x)$ is the projection function that maps the percentile x to the respective value ranking on the researcher's base community b. The score considers both base community information and the scores on other communities. Also, the ranking strategy allows different metrics in the projection function, such as number of citations, volume of publications and h-index, giving flexibility to 3c-index.

4. RESULTS AND CONCLUSION

We built a dataset from DBLP and Google Scholar in January 2015, with 18,511 authors and about 100 thousand papers with more than a half million citations. Our evaluation validates 3c-index against a ground truth of 137 (out of 18,511) ACM Awards winners, i.e., a perfect top-50 ranking has 50 winners of ACM Awards. We compare the ranking results to those by citations, volume, h-index and ca-index (cross-area) [5]. We use the normalized Discounted Cumulative Gain (nDCG), which measures the ranking quality according to a graded scale with a log-based discount factor to penalize relevant items in lower positions. We consider the *nDCG@k* for a rank cutoff k varying from 1 to 50.

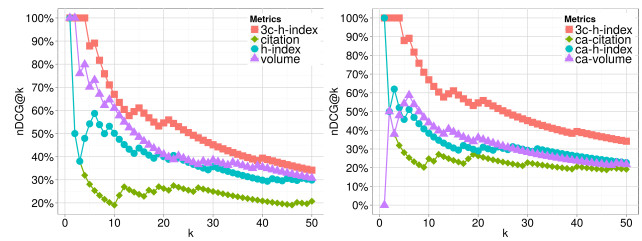
Influential Researchers. Table 1 shows the researchers ranked by 3c-index. The 3c-volume and 3c-h-index rank first researchers (names in bold) that are recognized in their communities: five for 3c-volume and four for 3c-h-index. Instead, 3-citation ranks two outstanding researchers. The 3c-index ranks well-known researchers at the first positions.

3c-index against Baselines. Figure 2a compares 3c-index (with h-index as its score function) and three standard metrics. It summarizes the good performance of our ranking strategy for almost the entire range of k.

3c-index against ca-index. ca-index follows three properties [5]: *plurality* as researcher's productivity is assessed in all areas of her publications; *diversity* as each research area profile is considered; and *equality* as all areas are regarded as equally important. In fact, the *equality* property puts SIGUCCS with equally importance as SIGMOD, which

Table 1: Researchers best ranked according to 3x-index in SIGs. In bold, the winners of ACM Awards, ♯*ACM fellow*, ‡*ACM distinguished scientist*, †*ACM senior member.*

Position	3c-citation	3c-volume	3c-h-index
1^{st}	**Scott Shenker♯**	**W. Bruce Croft♯**	**Jiawei Han♯**
2^{nd}	Ion Stoica♯	**Christos Faloutsos♯**	**Christos Faloutsos♯**
3^{rd}	M. Frans Kaashoek♯	**Surajit Chaudhuri♯**	**Scott Shenker♯**
4^{th}	David R. Karger♯	**Jiawei Han♯**	**W. Bruce Croft♯**
5^{th}	Sylvia Ratnasamy	**Scott Shenker♯**	ChengXiang Zhai‡
6^{th}	Mark Handley	ChengXiang Zhai‡	**Surajit Chaudhuri♯**
7^{th}	Paul Francis	Philip S. Yu♯	Wei-Ying Ma‡
8^{th}	Richard M. Karp♯	Zheng Chen†	Zheng Chen†
9^{th}	**Jon M. Kleinberg♯**	Divesh Srivastava♯	Philip S. Yu♯
10^{th}	Dina Katabi♯	Leif Azzopardi	Divesh Srivastava♯



(a) 3c-index versus Baselines (b) 3c-index versus ca-index

Figure 2: Results for 3c-index versus baselines and ca-index.

may have negatively affected its behavior. Hence, ca-index is not indicated to be used in a *closed community perspective*; however, it is still good for assessing research in multiple area scenario. Figure 2b shows the results of the ranking produced by 3c-index and ca-index using score given by the same baselines. It emphasizes the good performance of our proposed index comparing with the ca-index.

Our approach has best results to measure research contribution across communities, surpassing previous metrics. Overall, *3c-index* successfully measured the impact of researchers' contributions based on the publications by community. As future work, we plan to expand *3c-index* by considering the researchers' contributions through areas.

Acknowledgments

Work in collaboration with Thiago H. P. Silva and Ana Paula C. Silva, funded by CNPq. A complete version of this work is available at [8].

5. REFERENCES

[1] M. A. Brandão and M. M. Moro. Analyzing the strength of co-authorship ties with neighborhood overlap. In *DEXA*, pages 527–542, 2015.

[2] R. S. Burt. Structural Holes and Good Ideas. *American Journal of Sociology*, 110(2):349–399, 2004.

[3] M. S. Granovetter. The Strength of Weak Ties. *American Journal of Sociology*, 78(6):1360–1380, 1973.

[4] T. H. Haveliwala. Topic-sensitive pagerank. In *WWW*, pages 517–526, 2002.

[5] H. Lima et al. Aggregating Productivity Indices for Ranking Researchers Across Multiple Areas. In *JCDL*, pages 97–106, Indianapolis, USA, 2013.

[6] S. L. Montolio, D. Dominguez-Sal, and J. L. Larriba-Pey. Research endogamy as an indicator of conference quality. *ACM SIGMOD Record*, 42(2):11–16, 2013.

[7] M. Newman. Who is the best connected scientist? a study of scientific coauthorship networks. *Complex networks*, pages 337–370, 2004.

[8] T. H. P. Silva, L. Mota, A. P. C. Silva, and M. M. Moro. 3c-index: Research Contribution across Communities as an Influence Indicator. *JIDM*, 6(3):192–205, 2015.