Not so Snap a Judgement: Discussing the Peer Reputation Metric

Orestis-Stavros Loizides and Polychronis Koutsakis
Department of Electronic and Computer Engineering,
Technical University of Crete, Greece
Email: oloizidis@isc.tuc.gr, polk@telecom.tuc.gr

Abstract *- The lack of a clear method to judge a researcher's contribution has recently [1] led to the proposal of a new metric, called Peer Reputation (PR) metric. PR ties the selectivity of a publication venue with the reputation of the first author's institution. In [1], the authors compute PR for a number of networking research publication venues and argue that PR is a better indicator of selectivity than a venue's Acceptance Ratio (AR). We agree that PR is an idea towards the right direction and that it offers substantial information that is missing from AR. Still, we argue in this paper that PR is not adequate by itself in giving a solid evaluation of a researcher's contribution. In our study, we discuss and evaluate quantitatively the points on which PR does not sufficiently serve its purpose. To evaluate PR, we have gathered data for 11 conferences belonging to different research fields (networking, informatics and electronics), between 2008-2011. We also use three different rankings of doctoral programs in USA and two world university rankings, to study how they influence the PR results.*

I. AR versus PR

The authors in [1] argue that a new metric for the qualitative assessment of a researcher's contribution in a specific discipline is needed. To make their point, they refer to the shortcomings of the use of conference AR (authors self-filtering can lead to highly selective conferences having similar ARs with less selective ones) and of the use of the citation number a publication receives (citations of conferences are not well-tracked, new publications in conferences and journals take time to attract citations). The authors in [1] argue that the selectivity of a publication venue is a function of the reputations of the authors' affiliating institutions, and for simplicity they represent each paper by the affiliation of its first author, which is mapped to a rank. Then, the PR of the conference is a function of this rank set, and defined as PR(P)=gist({rank(inst(first(p))): $\forall p \in P$}), where P is a publication venue, p is paper, first() gives the first author, inst() maps an author to the affiliation, and rank() is the given function for mapping an institute to a rank. Their proposed PR metric, which is proposed for making "snap judgements" on a publication quality, conveys the selectivity of a conference with a tuple, say <1/3, 20>, indicating that 1/3 of the papers at that conference are from the top 20 universities. PR is evaluated over 18 venues, in [1], 16 of which are conferences and 2 are journals. Our work in this paper focuses only on conferences.

It is also stated in [1] state that although PR is not a perfect metric to assess the quality of a publication, it provides a coarse-grain measure of the selectivity of a conference or a journal, and can potentially be more helpful than AR. We agree with the latter comment, not only because of the aforementioned shortcomings associated with AR, but also because AR does not give the extent of competition among peers for a conference publication, since there is no indication of the quality of rejected or accepted papers (this point is not made in [1]). Hence, PR can indeed be very helpful in highlighting the work of researchers from lower ranked universities who are publishing in top venues. However, we believe that, despite the improvement it offers over AR, PR also has significant room for improvement itself. We conducted an extensive study to show how it can improve.

II. Ways of Improving PR

As mentioned above, we intuitively found PR and the results in [1] to have significant room for improvement. Before going into detail about the practical issues we found, there is also a "philosophical" issue regarding the use of PR, that we wish to briefly mention:

PR adopts the currently prevalent bias in academia, that an institution "makes" a researcher, i.e., a researcher working in a top-ranked university is expected to be better than a colleague affiliated with a lower-ranked university. Of course, many top-class researchers work in lower ranked universities. The number of IEEE/ACM Fellows in universities ranked below the top-20 is enough to question the "aloofness" of this bias. However, we will not delve further into this point in this paper, as it is not a problem of PR but of the prevalent conception, as already mentioned.

The practical issues that we found concerning PR are the following:

1. The evaluation of PR is based, in [1], only on networking conferences, and is made only over two years (2008, 2009). This is not only limiting in itself, but also because the universities' rankings used in [1] are based on the evaluation of *Computer Science* graduate schools, hence the rankings reflect the quality of a graduate program on a much larger set of disciplines than the single discipline (networking) that [1] focuses on. Hence, these rankings may not even be fully representative in terms of the networking field (e.g., excellence in another field may give one university the edge in Computer Science rankings over other universities which might have stronger networking graduate programs).
2. PR disregards the importance of the location of a conference. Depending on the continent where a conference takes place, submissions from far-away countries might be discouraged as travel could be too time- and energy-consuming, or too expensive.
3. PR is based only on the ranking of the university that the first author is affiliated with. Given the fact that many papers stem from the collaboration of authors from different universities, or from different departments of the same university, which are differently ranked in their respective fields, the choice of using just the first author seems to be an oversimplification.
4. A limitation that the authors themselves mention, but do not consider as important, is that they base PR solely on the ranking of US universities. The authors explain, in [1], that in their view this is not a serious limitation for popular networking conferences as these venues receive a high fraction of papers from USA universities. Our study, which includes rankings of world universities, shows that PR results can significantly change when universities outside the USA are taken into account.

All of the above limitations are discussed in Section III.

III. An Extensive Evaluation of PR

In order to address the first of the practical issues mentioned above, so that we could get a more representative set of results from more fields of Computer Science, we studied 11 top conferences between 2008 and 2011. In total, we studied 30 venues, 6 of which focus on Informatics (I), 8 on Electronics (E) and 16 on Networking (N), of which 4 are the networking symposiums of flagship IEEE conferences (ICC, WCNC)[1] but have a much higher AR than the rest of the conferences under study (note that the ARs for ICC and WCNC refer to the whole conference, not just the networking symposiums). A total of 1885 papers were used in our study.

The respective data for each venue are presented in Table I. The AR values for the conferences used in our study have been taken from [6-10] and from numerous other sources on the web, in order to double-check the AR values reported. In the very few cases that our results differ from those in [1], this is caused by the fact that we round up to the higher integer while the authors in [1] round up to the lower (e.g., for the 31 accepted papers in Mobihoc 2009, our 1/3 PR results use the 11th highest ranked authors' affiliation, while [1] used the 10th).

#	Conference	#papers	AR	PR: 1/4	PR: 1/3	Field
1	MOBIHOC 2008, HONG KONG	44	14.7	28	44	N
2	MOBIHOC 2009, NEW ORLEANS	31	17.7	44	44	N
3	MOBIHOC 2010, CHICAGO	26	16.6	28	35	N
4	MOBIHOC 2011, PARIS	25	19.7	13	20	N
5	IMC 2008, ATHENS	31	17.3	10	11	N
6	IMC 2009, CHICAGO	41	22.4	13	20	N
7	IMC 2010, MELBOURNE	47	22.3	14	20	N
8	IMC 2011, BERLIN	42	19.1	14	35	N
9	SENSYS 2008, RALEIGH	25	16.3	5	5	N
10	SENSYS 2009, BERKELEY	21	17.6	1	17	N
11	SENSYS 2010, ZURICH	25	17.6	14	20	N
12	SENSYS 2011, SEATTLE	24	19.5	20	28	N
13	ICC 2010, CAPE TOWN	138	39.5	122	122	N
14	ICC 2011, KYOTO	165	38.5	122	122	N
15	WCNC 2010, SYDNEY	180	37.2	122	122	N
16	WCNC 2011, CANCUN	142	48.1	122	122	N
17	ISCA 2010, SAINT-MALO, FRANCE	44	18	10	11	E
18	ISCA 2011, SAN JOSE	40	19	8	11	E
19	MICRO 2010, ATLANTA	45	18.1	8	13	E
20	MICRO 2011, PORTO ALEGRE,	44	21	11	13	E

[1] For ICC, we used the Wireless Networking Symposium and the Next Generation Networking and Internet Symposium, for both 2010 and 2011. For WCNC, we used the Networks Track for both years.

	BRAZIL					
21	ASPLOS 2010, PITTSBURGH	32	17.7	7	8	E
22	ASPLOS 2011, NEWPORT BEACH, USA	32	21.1	8	11	E
23	FCCM 2010, CHARLOTTE	24	18.2	39	122	E
24	FCCM 2011, SALT LAKE CITY	21	17.7	53	99	E
25	ICDE 2010, LONG BEACH, USA	69	12.5	20	39	I
26	ICDE 2011, HANNOVER	98	19.8	122	122	I
27	VLDB 2008, AUCKLAND, NEW ZEALAND	101	17.1	14	39	I
28	VLDB 2009, LYON, FRANCE	108	17	28	122	I
29	VLDB 2010, SINGAPORE	120	20.1	20	58	I
30	VLDB 2011, SEATTLE	100	18.1	20	53	I

Table I. Conference Data

The second limitation of PR (not taking into account the conference location) is evident by the results for the SenSys conference, between the years 2008 and 2010. The results reveal that when the conference took place in the USA, in 2008 and 2009, the 1/4 and 1/3 PR results showed a significantly larger selectivity in comparison to the results for 2010, when the conference took place in Europe. It should also be emphasized that the AR fluctuated very slightly for all three of these venues of SenSys; this agrees with the results in [1] and the rest of our results and indicates that there is no correlation between PR and AR, other than the obvious one that a much higher AR usually leads to lower selectivity. The results for SenSys 2011, which again took place in the USA, are slightly worse, but the AR was quite higher, therefore a direct comparison cannot be made. Another indication of the influence that the location of a venue can have on PR results is given by comparing the results for the four VLDB venues, which took place in four different continents. The VLDB 2009 conference, which took place in France and had the lowest AR among the four venues (therefore based on AR it would seem to be the *most* selective), appears to have been the *least* selective one based on the PR metric. A possible explanation for this result is the attractiveness of France as a location. This may have led to the increased number of submissions (highest number among the four VLDB editions studied), thus leading to the lowest AR but also to the acceptance of many papers written by non-US-based authors, hence the worse PR values (this result is therefore also tied with the fourth issue of PR, which we discuss below).

Still, location is not always a critical factor. Counter-examples include the IMC and ISCA conference venues, where the change in location led to rather negligible changes in the PR results; the slightly better results for IMC 2008, in comparison to IMC 2009-2011, are connected to the much lower AR and number of accepted papers for that venue.

The above results seem to indicate that location is a factor, but its influence can be mitigated by other facts (e.g., a paper may be co-authored by researchers working in different locations, hence it is easier for one of them to present it). The efficient incorporation of location into a metric is a quite complex task which we intend to tackle in future work.

The co-authorship of papers leads us to the third limitation of PR (taking into account the affiliation of just the first author). To give an example of how this issue, combined with the inherent "aloofness" of the metric, can lead to incorrect judgements regarding the quality of a venue, consider the following case: let's assume that a conference contains only papers sole-authored by IEEE/ACM fellows from universities that are ranked between 50-70. Suppose also that another conference contains papers where the first authors are graduate students or post-docs from top-20 ranked universities. This second conference will be ranked by PR as a top-class venue, whereas the first conference will be considered a rather mediocre venue. This is obviously a wrong conclusion, but this is what someone using PR will conclude by making a "snap judgement". The above example takes the specific limitation of PR to the extreme, of course, but the limitation is clear. Therefore, a finer grain evaluation than that which is offered by PR is needed, in our view.

To alleviate this problem (which is of course a policy decision having to do with the concept of "snap judgement" and not a problem with the idea of PR itself) we used the mean ranking of the universities that all authors of each paper are affiliated with. Figures 1 and 2 show, once again, that the results vary depending on the conference. It should be noted that the conference numbers on the x-axis follows that of Table 1, is completely ad-hoc and plays no role in the results other than allowing a visual comparison between the ranking of the first author and the mean ranking for the same venue.

As shown in the Figures, in quite a few venues the 1/4 and 1/3 PR first author and average results are almost identical. This is explained by the fact that a large number of papers were written from researchers who had the same affiliation. However, the inclusion of all authors can also lead to significant changes in the PR

values and hence to very different "snap judgements", as shown in the Figures for some of the other conferences. This was true for 23% of the venues in terms of the 1/4 PR results, and for 40% of the venues in terms of the 1/3 PR results. In all but one of the venues these changes were related to a *decrease* in the 1/4 and 1/3 PR values (i.e., an increase in selectivity). This indicates that at least for these conferences, the first authors were on average affiliated with lower-ranked universities than their co-authors.

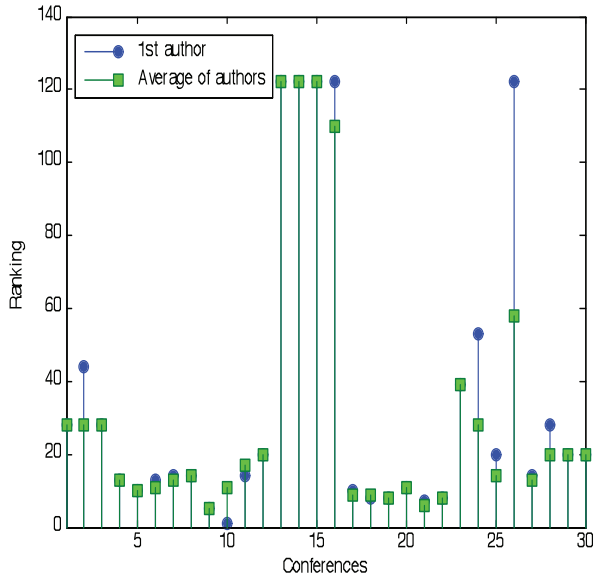

Figure 1. 1/4 PR results for the first authors and for the average over all authors

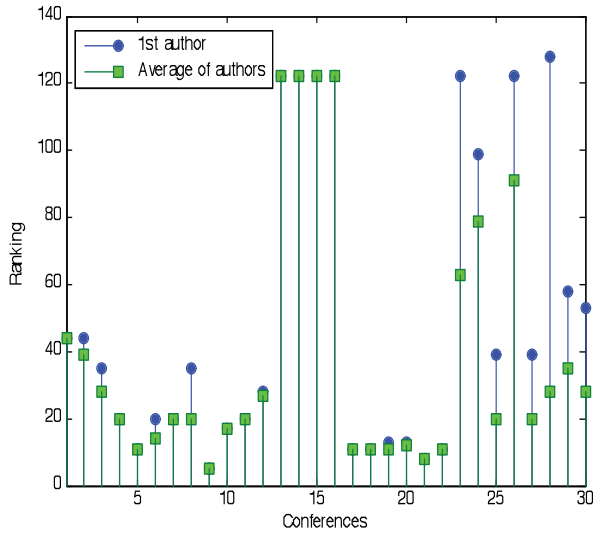

Figure 2. 1/3 PR results for the first authors and for the average over all authors

All of the above results have been produced, similarly to [1], with the US News and World Report rankings of graduate programs in Computer Science [2]. As it is well-known, university rankings are the subject of much controversy, and the rankings methods used are often questioned by universities and individual scholars. In order to address the fourth issue of PR and get a more clear view of how different rankings influence our results, we have used four additional rankings:

a) the latest (2010) National Research Council (NRC) rankings for the top 127 Computer Sciences graduate programs in the USA [3]

b) the NRC rankings for the top 136 Electrical and Computer Engineering graduate programs in the USA [3].

c) the specialized Academic Ranking of World Universities (ARWU) in Engineering/Technology and Computer Sciences - 2010 [4].

d) the specialized latest ARWU ranking in Computer Science for 2012 [5].

Over all the venues we studied, we found that 55% of the papers were written by authors who did not have an affiliation with a university in the USA (therefore, they are completely disregarded by PR, if they are first authors; otherwise they would be disregarded by PR anyway). This indicates the importance of taking world university rankings into account to ensure the accuracy and fairness of the results, and is further supported by the results presented in Figures 3-6, on the PR metric. These reveal that, despite the fact that the results show more or less similar *behavior* for all rankings, the PR values differ significantly depending on the ranking, hence very different "snap judgements" will be made by using PR when including non-US universities that in the case of ignoring them. The results also reveal, similarly to Figures 1-2, that for over 30% of the results, for each ranking, the use of the average PR leads to significantly different results than those when only the first author's affiliation is used.

We need to clarify that, similarly to [1], we treat all research institutions that are not included in a specific ranking as ranked lower than the lowest ranking (e.g, for the NRC Computer Science rankings, they get a rank equal to 128). This is the reason that the ICC and WCNC conferences (conferences 13-16) have an almost constant PR value in all the Figures, i.e., the 1/4 and 1/3 PR values correspond to research institutions that are not included in the rankings.

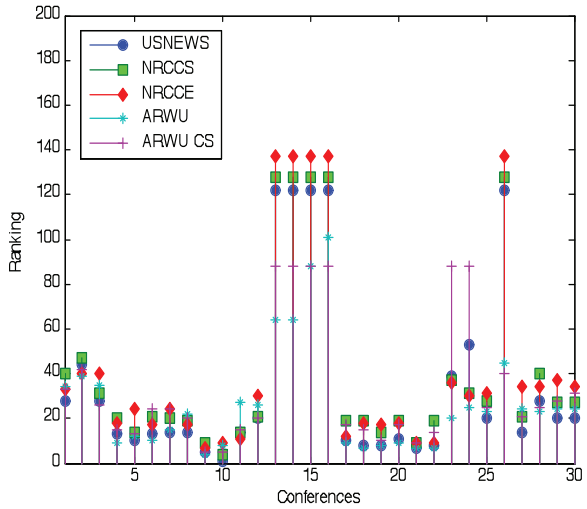

Figure 3. 1/4 PR results for the first author, with the use of 5 different rankings

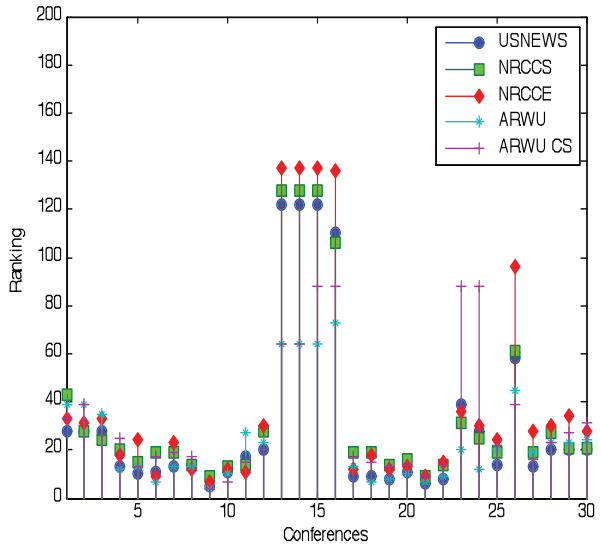

Figure 4. Average 1/4 PR results, with the use of 5 different rankings

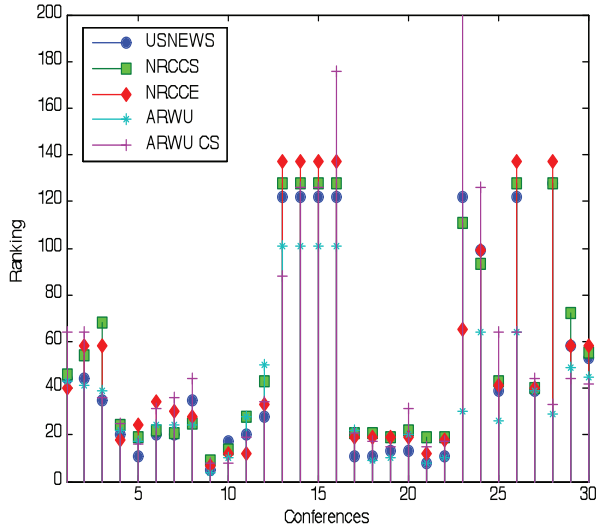

Figure 5. 1/3 PR results for the first author, with the use of 5 different rankings

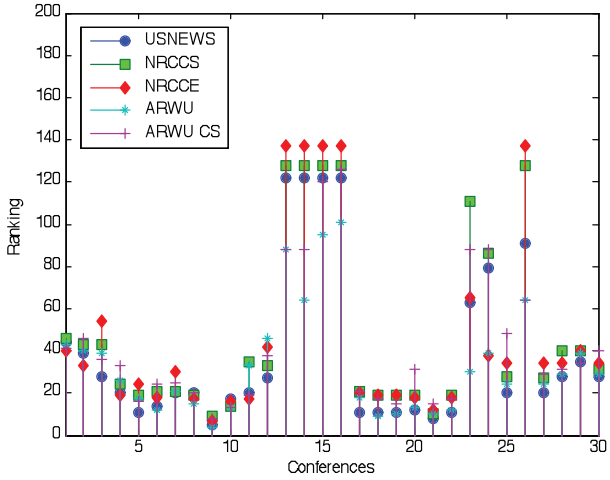

Figure 6. Average 1/3 PR results, with the use of 5 different rankings

Tables II and III show the average 1/4 and 1/3 PR results derived from each one of Figures 3-6, for all 30 venues. Table II includes the average results for just the first author, while Table III includes the average results for the mean PR for all authors. We present the results for all venues, and for the N, E, and I venues separately. The results presented in the Tables show once again the differences when using different rankings and that, on average, first authors are affiliated with lower ranked institutions than the rest of the authors in the accepted papers. These results also reveal that, when using any of the specialized rankings (i.e., with the exception of the more general ARWU one), the inclusion of all authors has minor effects on the results for networking conferences, larger effects on the results for electronics conferences and major effects on the results for informatics conferences.

Finally, another way of judging the influence of including all authors in the PR results is the relative ordering of conferences based on their PR values for the first author and for all authors, respectively. Our results once again confirm the significant effect of taking all authors into consideration, for some of the conferences: the SENSYS2009 conference, for example, which is the best among all 30 conferences in terms of its 1/4 PR value for the first author (US News ranking), falls to the 8th place when all authors are considered; another example is the VLDB09 conference, which is tied with 5 other conferences in the last place in terms of its 1/3 PR value (equal to 122) for the first author, but climbs to the 18th place (with a strong 1/3 PR value equal to 28) when all authors are considered.

IV. Conclusions and Future Work

The use of the paper acceptance ratio of a conference is not a trustworthy metric to evaluate a researcher's contribution in his/her field. The PR metric, proposed in [1], is a rough coarse-grain metric proposed for making quick assessments of a researcher's contribution. Although PR is an idea towards the right direction in our view, it still has room for improvement in ways that we tried to point out in this work.

Our future work will be targeted towards proposing a new, finer-grain metric that will partially incorporate PR, but will alleviate its limitations based on our findings in this paper. More specifically, our metric will take into account not just an author's affiliation (as PR does) but will also utilize data regarding an author's recognition in his/her research field (e.g., h-index), a paper's efficiency (number of citations in recent years) and the location of a conference. Our work will try to classify conferences into categories, in terms of their quality, instead of focusing on the individual ranking of each conference, but it will also allow for such an individual conference ranking.

	USNEWS		NRCCS		NRCCE		ARWU		ARWU_CS	
	1/4PR	1/3PR	1/4PR	1/3PR	1/4PR	1/3PR	1/4PR	1/3PR	1/4PR	1/3PR
ALL VENUES	22.1	38.9	26.8	44.4	27.7	41.5	18.9	28.2	25.6	43.1
N	17	24.9	21.7	31.1	22.5	29.5	20.1	27.2	20.2	32.1
E	18	36	20.9	40.6	18.6	33.8	11.5	21.6	32.3	55.4
I	40.8	77.2	48.8	82.2	54.6	82.6	27.8	41.4	27.8	49.8

TABLE II. Averages for the 1/4 and 1/3 PR results for the first author, over all the venues.

	USNEWS		NRCCS		NRCCE		ARWU		ARWU_CS	
	1/4PR	1/3PR	1/4PR	1/3PR	1/4PR	1/3PR	1/4PR	1/3PR	1/4PR	1/3PR
ALL VENUES	17.6	26.6	21.6	35.8	24	32.2	18	25.2	24.8	32.7
N	16.5	22.1	20.5	27.3	20.3	26.4	19.6	25.8	19.6	26.6
E	14.8	25.8	18.4	38	18.1	26.1	10.8	17.6	32.3	36.5
I	25	38.8	29.4	52.6	42.4	52.8	26	36	25.4	41.2

TABLE III. Averages for the mean 1/4 and 1/3 PR results for all authors, over all the venues.

References

[1] S. Nelakuditi, C. Gray, R. Roy Choudhury, "Snap Judgement of Publication Quality: How to Convince a Dean that you are a Good Researcher", ACM SIGMOBILE Mobile Computing and Communications Review, Volume 15, No. 2, 2011, pp. 20-23.

[2] http://grad-schools.usnews.rankingsandreviews.com/best-graduate-schools/top-science-schools/computer-science-rankings

[3] http://graduate-school.phds.org/rankings/

[4] http://www.arwu.org/FieldENG2010.jsp

[5]http://www.shanghairanking.com/SubjectCS2012.html

[6] Networking Conference Statistics, http://www.cs.ucsb.edu/~almeroth/conf/stats/

[7] http://vldb.org/archives/statistics.html

[8]http://students.engr.scu.edu/~sschaeck/acceptancerates/

[9]http://dl.acm.org/citation.cfm?id=1934902&picked=prox

[10]http://dl.acm.org/citation.cfm?id=1815961&picked=prox

Reviews and Authors' Response for paper: Not so Snap a Judgement: Discussing the Peer Reputation Metric

Reviews

This paper takes a closer look at Peers' Reputation (PR), a metric proposed earlier to indicate the quality of a research publication. PR metric, which is intended for making "snap judgements" on publication quality, conveys the selectivity of a conference with a tuple, say <1/3, 20>, indicating that 1/3 of the papers at that conference are from the top 20 universities in USA. The authors agree that PR is an improvement over the acceptance ratio (AR), a metric currently used to convey the selectivity of a conference publication. According to the authors, PR is an idea in the right direction, but it has four shortcomings.1) The evaluation of PR is based on networking conferences using Computer Science graduate school rankings, which may be different for the specific field of networking. 2) PR disregards the importance of the location of a conference. 3) PR is based on the affiliation of the first author, not all authors. 4) Only the ranking of US universities are considered by PR. Towards understanding the impact of these shortcomings on PR metric, they conduct a broader evaluation. They conclude that there is a need for a new finer-grain metric that will partially incorporate PR, but will alleviate its shortcomings.

A strength of the paper is its evaluation of PR, using 4 years of data for 11 conferences belonging to 3 different fields (networking, informatics, and electronics) using 3 US university and 2 world university rankings. Also, it brings attention to the location of a conference and its impact on PR which was not considered as a factor in PR. Moreover, this paper adds to the debate on publication quality, a hot topic of discussion of late. Therefore, I recommend accepting the paper for publication.

On the other hand, this paper has some weaknesses. The shortcomings of PR listed in this paper do not seem fundamental. A coarse-grain metric meant for snap judgement can be improved by a fine grain metric that takes many aspects into account. The problem is how do you assign relative weight to each of the aspects. It will be hard to design one that does not have shortcomings, particularly in rare cases. Therefore, it will be good if the authors can include at least a rough sketch of their proposed metric and a discussion of how it addresses the shortcomings of PR, without introducing new ones.

The argument about a new conference with IEEE fellows from lower ranked schools and postdocs from higher ranked schools is a reasonable argument. However, its very much possible that such a new conference is characterized as some aggregate function of the individuals PR metric. So the conference organized by IEEE Fellows could be the average of their PR values, which would make that conference look very strong. On the other hand, the postdocs' conference can be an average of their PR values, and if they have not been publishing in top venues in the past, that conference's PR metric might reflect that. In fact, the entire community can agree on a few top universities and top conferences, and every other university and conference can bootstrap from these rankings.

Authors mention that the PR value of a conference located outside the US decreased. How was the PR computed? Was the distribution derived only from the US-based authors or was it across all authors. Also, the correlation between location and PR is not consistent... in many cases, even if the conference happened outside the US, the PR value still decreased (MobiHoc Hong Kong vs. New Orleans).

In terms of writing, authors should summarize their findings better, at the end of each part of evaluation and perhaps again in the conclusion. In the current version, it is easier to get a list of shortcomings of the PR metric pointed out by the authors but it is not quite apparent whether their evaluation indeed confirms their points of view.

Authors' Response

The authors would like to sincerely thank the editor and the anonymous reviewers, for their very valuable comments which helped to improve and clarify the goals of this work.
Regarding the reviewers' suggestions:
We have included a rough sketch of our proposed metric, which we are working on, and a discussion of how it can address the shortcomings of PR, without introducing new ones, in our Conclusions section.

We agree with the reviewers' points regarding the possibilities of using some variants of PR for estimating the quality of specific conferences. We

also believe that it would be good (although difficult) for the community to agree on a few top universities and conferences so that other universities and conferences could bootstrap from those rankings. However, the paper in which PR was proposed did not offer these ideas, and that is why we propose methods to improve on the PR metric.

The PR value in the specific results that the reviewer is concerned with (on how it was computed) was derived only from US-based authors, following the original paper's methodology, in order to emphasize the validity of our points. In terms of the significance of location and its effect on PR, we mention in our paper that location is not always a critical factor, but it often seems to play a role in the PR results. Therefore, there are examples in our results where location seems to have small or no correlation to PR. Still, the results where location seems to be correlated to PR are many, and the correlation in those cases is clear. On the other hand, in the cases where the correlation is not apparent, there are often other factors that need to be taken into account when trying to explain the PR results. For example, in the case that the reviewer points out (Mobihoc 2008 in Hong Kong vs. Mobihoc 2009 in New Orleans) the Mobihoc 2008 conference received almost double the number of submissions that the Mobihoc 2009 conference received, and more than double than any other Mobihoc conference of the past 6 years. Therefore, Mobihoc 2008 was an outlier in this respect, and this fact makes any direct comparison between the 2008 version and any other version of the conference very difficult.

Finally, regarding the comment that has to do with the way our findings are presented, we hope that the final version of the paper clearly presents our findings and how they confirm our point of view.