Exploring and Understanding Scientific Metrics in Citation Netsworks

Mikalai Krapivin, Maurizio Marchese and Fabio Casati

Department of Information Engineering and Computer Science,
University of Trento, Italy
krapivin, marchese, casati@disi.unitn.it

Abstract. This paper explores scientific metrics in citation networks in scientific communities, how they differ in ranking papers and authors, and why. In particular we focus on network effects in scientific metrics and explore their meaning and impact. We initially take as example three main metrics that we believe significant; the standard citation count, the more and more popular h-index, and a variation we propose of PageRank applied to papers (called PaperRank) that is appealing as it mirrors proven and successful algorithms for ranking web pages and captures relevant information present in the whole citation network. As part of analyzing them, we develop generally applicable techniques and metrics for qualitatively and quantitatively analyzing such network-based indexes that evaluate content and people, as well as for understanding the causes of their different behaviors. We put the techniques at work on a dataset of over 260K ACM papers, and discovered that the difference in ranking results is indeed very significant (even when restricting to citation-based indexes), with half of the top-ranked papers differing in a typical 20-element long search result page for papers on a given topic, and with the top researcher being ranked differently over half of the times in an average job posting with 100 applicants.

Keywords: Scientific metrics, Scientometric, Page Rank Algorithm, Paper Rank, H-index, Divergence metric in ranking results.

1 Introduction

The area of scientific metrics in scholarly social networks (metrics that assess the quality and quantity of scientific productions) is a relevant area of research [1,2,3] aiming at the following two objectives: 1) measuring scientific papers, so that "good" papers can be identified by the scientific community and so that researchers can quickly find useful contributions when studying a given field, as opposed to browsing a sea of papers, and 2) measuring individual contributions and related reputation, to determine the impact of a scientist and to help screen and identify candidates for hiring and promotions in industry and academia.

Until only 30 years ago, the number of researchers and of conferences was relatively small, and it was relatively easy to assess papers and people by looking at papers published in international journals. With small numbers, the evaluation was essentially based on looking at the paper themselves. In terms of quantitative and measurable indexes, the number of publication was the key metric (if used at all). With the explosion of the number of researchers, journals, and conferences, the "number of publications" metric progressively lost meaning. On the other hand, this same explosion increased the need for quantitative metrics at least to "filter the noise". For example, a detailed, individual, qualitative analysis of hundreds of applications typically received today for any job postings becomes hard without quantitative measures for at least a significant preliminary filtering.

Recently, the availability of online databases and Web crawling made it possible to introduce and compute indexes based on the number of citations of papers (citation count, from hereafter denoted as CC) and its variations or aggregations, such as the impact factor and the h and g indexes [4]) to understand the impact of papers and scientists on the scientific community. More and more, Universities (including ours) are using these indexes as a way to filter or even decide how to fill positions by "plotting" candidates on charts based on several such indexes.

Besides "traditional" metrics, novel metrics for papers and authors are being proposed [5]: they are inspired at how the significance of Web pages is computed (essentially by considering papers as web pages, citations as links, and applying a variation of the PageRank algorithm [5,6]). PageRank-based metrics are emerging as important complement to citation counts as they incorporate important information present in the whole citation network, namely the "weight" (the reputation or authority) of the citing paper and its density of citations (how many other papers it references) in the metric. From a computational point of view, PageRank is a statistical algorithm: it uses a relatively simple model of "Random Surfer" [6] to determine the probability to visit a particular web page. Since random browsing through a graph is a stochastic Markov process, the model is fully described by Markov chain stochastic matrix. The most intriguing question about PageRank is how to compute one for very large networks such the web. The inventors of PageRank, Brin and Page, proposed a quite effective polynomial convergence method [6], similar to the Jacobi methods. Since then, a significant amount of research has been done in the exploration of the meaning of PageRank and proposals for different computation procedures [5,7,8,9]. When the attention is shifted from web pages to scientific citations, the properties of the citation graph – mainly its sparseness – has been used to simplify the computational problem [10]. Some recent work has also started to analyze and compare the effectiveness of the different ranking algorithms [11].

In the present work, we have based our computations on a variation of Page Rank (called PaperRank) for ranking scholarly documents. In particular, this paper performs an experimental study of scientific metrics based on citation networks with the goal of (1) assessing the extent of differences and variations on the evaluation results when choosing a certain metric over another, and (2) exploring and understanding the reasons behind these differences. We performed the analysis on a dataset consisting of over 260K ACM (Association for Computing Machinery) publications, a social scientific network of 244K authors and ca. one million citations. The results of the analysis are rather significant, in that even if we restrict to citation-based indexes, the choice of the specific index rather than another changes the result of filtering and selection of papers and evaluation of people about half of the times.

The structure of the paper is as follows. In Section 2 we describe the dataset and in Section 3 we focus on the presentation of the main indexes for papers and for authors. The in-depth exploration of the indexes is provided in Section 4 (for papers) and Section 5 (for authors), along with comments and discussions on the results and with the introduction of the appropriate meta-indexes. Finally, the major findings of the present work are summarized in Section 6.

2 Data set description and data preprocessing

The starting point for our analysis is a dataset of 266.788 papers published in ACM conferences or journals, and authored by 244.782 different authors. For any given paper in the set, we have all its references (outgoing citations), but we only have citations to it (incoming citations) from other papers in the dataset, and hence from ACM papers. To remove this bias (to the possible extent), we disregard references to non-ACM papers. In other words, we assume that the world, for our citation analysis, only consists of ACM papers. Including references to non-ACM papers would in fact unfairly lower the measure for PaperRank since, as we will show, PaperRank is based on both incoming and outgoing citations. Although we have no measurable evidence, given that we are comparing citation-based metrics we believe that the restriction to an “ACM world” does not change the qualitative results of the analysis.

The majority of papers in our dataset have been processed manually during the publishing process and all author’s names have been disambiguated by humans. This is important since Google Scholar or other autonomous digital libraries use machine learning-based unsupervised techniques to disambiguate the information and therefore they are prone to introduce mistakes. Thus, although incomplete, the ACM dataset has a high level of quality in particular in respect to authors and citations. A sample of the dataset format is available at this article’s companion web page [1].

3 Paper Rank and PR-Hirsch

3.1 Page Rank outline

The original Page Rank algorithm [6] ranks the nodes of a directed graph with N vertices. The rank of a node is determined by the following recursive formula, where *S(j)* is the quantity of outgoing links from a node *Pj*. $i, j \in \{1..n\}$ are just sequence numbers and *D* is the set of nodes such that there is a path in the graph from them to node *i*.

$$P_i = (1-d) \cdot \sum_{i \neq j}^{j \in D} P_j / S(j) + d / N \qquad \textbf{(1)}$$

[1] http://www.dit.unitn.it/~krapivin/exploring_citation%20_metrics_08.html

The formula can be seen in matrix form and the computation can be rewritten as an eigenvector problem: $\vec{r} = A \cdot \vec{r}$, where *A* is the transition matrix. This consideration exposes several potential problems in rank computation as discussed in [7, 12]. One of them is the presence of the nodes which link to other nodes but are not linked by other nodes, called *dangling nodes*. In this case, equation (1) may have no unique solution, or it may have no solution at all (it will lead to zero-rows occurrence in the transition matrix and uncertainty of the rank of the dangling nodes). Such problem may be resolved with the introduction of a dump-factor $0 < d < 1$.

The damp factor (proposed in the original PageRank article [6]) helps to achieve two goals at once: 1) faster convergence using iterative computational methods, 2) ability to solve the equation, since all the nodes must have al least d/N Page Rank even if they are not cited at all.

3.2 PaperRank

PageRank has been very successful in ranking web pages, essentially considering the reputation of the web page referring to a given page, and the outgoing link density (pages P linked by pages L where L has few outgoing links are considered more important than pages P cited by pages L where L has many outgoing links). PaperRank (PR) applies page rank to papers by considering papers as web pages and citations as links, and hence trying to consider not only citations when ranking papers, but also taking into account the rank of the citing paper and the density of outgoing citations from the citing paper.

From a computation perspective, PR is different from PageRank in that loops are very rare, almost inexistent. Situations with loop where a paper *A* cites a paper *B* and *B* cites *A* are possible when authors exchange their working versions and cite papers not yet published but accepted for publication. In our dataset, we have removed these few loops (around 200 loops in our set). Thus the damp factor is no longer needed to calculate PR. Furthermore, considering that a citation graph has $N>>1$ nodes (papers), each paper may potentially have from 1 to $N-1$ inbound links and the same quantity of outgoing ones. However, in practice citation graphs are extremely sparse, (articles normally have from 5 to 20 references) and this impact the speed of the computation of PR.

However, also in this case the problem may have no solution because of *initial papers* (papers who are cited but do not cite). To avoid this problem we slightly transform initial problem assigning a rank value equal to 1 to all initial nodes, and resetting it to zero at the end of the computation (as we want to emphasize that papers who are never cited have a null paper rank). Thus the problem become solvable and the Markov matrix may be easily brought to a diagonal form. We used fast and scalable recursive algorithm for calculating PaperRank, which corresponds to the modified equation: $\vec{r} = A \cdot \vec{r} + \vec{r}_0$.

3.3 PR-Hirsch

One of the most widely used indexes related to author is the H-index proposed by Jorge Hirsch in 2004 [4]. The H-index for an author is the maximum number *h* such that the author has at least *h* articles with *h* citations each. The H-index tries to value consistency in reputation: it is not important to have many papers, or many citations, but many papers with many citations.

We propose to apply a similar concept to measure authors based on PR, named PR-Hirsch (hereafter PRH). However, we cannot just say that PRH is the maximum number q such that an author has q papers with rank q or greater. In fact, while for H-index it is reasonable to compare the numbers of papers with the number of citations the papers have, for PRH the "meaning" is not so direct. The fact that a paper has a CC of 45 is telling us something we can easily understand (and correspondingly we can understand the H-index), while the fact that a paper has a PR of 6.34 or 0.55 has little "meaning". In order to define a PR-based Hirsch index, we therefore rescale PR to a value that can be meaningfully compared with the number of papers. Let's consider in some detail our set: we have a graph with N nodes (vertices) and n citations (edges). Each i-th node has PR equal to P_i, that expresses the probability for a random surfer to visit a node, as in the Page Rank algorithm. So let's assume that we run exactly n surfers (equal to the quantity of citations), and calculate the most probable quantity of surfers who visited node i. If the probability to visit the node i for one surfer is p_i, the expectation value Q_i for n surfers to visit the node i will be $p_i \cdot n$, *which is the most probable quantity of surfers, who visited node i*. We then sum probabilities since all surfers are independent. Finally, we normalize PR for each node according to the full probability condition. If the total sum of all PRs equals to M, the expected value for n surfers is:

$$Q_i = P_i \cdot n / M \quad (2)$$

Where P_i is a Paper Rank of the paper i, and n/M is a constant (≈5.9169 for our specific citation graph). So in other words we rescale PR to make it comparable with the quantity of citations. Indeed, Q_i is the most probable *quantity of surfers who visited a specific paper i*, whereas to compute Hirsch index we use the *quantity of citations for the paper i*.

Following the definition of H-index and the previous discussion, we define PR-Hirsch as the maximum integer number h such that an author has at least h papers with Q value (i.e. rescaled PR following equation (2)) equal or greater than h.

4 Exploring paper metrics

4.1 Plotting the difference

The obvious approach to exploring the effect of using PR vs CC in evaluating papers would consist in plotting these values for the different papers. Then, the density of points that have a high CC and low PR (or vice versa) would provide an indication of how often these measures can give different quality indication for a paper. This leads however to charts difficult to read in many ways: first, points overlap (many papers have the same CC, or the same PR, or both). Second, it is hard to get a qualitative indication of what is "high" and "low" CC or PR. Hence, we took the approach of dividing the CC and PR axis in bands.

Banding is also non-trivial. Ideally we would have split the axes into 10 (or 100) bands, e.g., putting in the first band the top 10% (top 1%) of the papers based on the metric, to give qualitative indications so that the presence of many papers in the corners of the chart would denote a high divergence. However the overlap problem would remain. Moreover, it would distort the charts in a significant way since the measures are discrete. For example the number of papers with 0 citations is well above 10%. If we neglect this issue and still divide in bands of equal size (number of papers), papers with the same measure would end up in different bands. This gives a very strong biasing in the chart (examples are provided in the companion web page[2]).

Finally, the approach we took (Figure 1) is to divide the X-axis in bands where each band corresponds to a different citation count (CC) measure. With this separation we built 290 different bands, since there are 290 different values for CC (even if there are papers with much higher CC, there are only 290 different CC values in the set). For the Y-axis we leverage mirrored banding, i.e., the Y-axis is divided into as many bands as the X-axis, also in growing values of PR. Each Y band contains the same number of papers as X (in other words, the vertical rectangle corresponding to band i in the X axis contains the same number of papers qi as the horizontal rectangle corresponding to band i of the Y-axis). We call a point in this chart as a square, and each square can contain zero, one, or many papers.

The reasoning behind the use of mirrored banding is that this chart emphasizes divergence as distance from the diagonal (at an extreme, plotting a metric against itself with mirrored banding would only put papers in the diagonal). Since the overlap in PR values is minimal (there are thousands of different values of PR and very few papers with the same PR values, most of which having very low CC and very low PR, and hence uninteresting), it does not affect in any qualitatively meaningful way the banding of the Y-axis.

Fig. 1 shows a very significant number of papers with a low CC but a very high PR. The color corresponds to the number of papers within a band (for actual values of PR and CC for each band see the companion web site) these are the white dots (a white color corresponds to one paper). Notice that while for some papers the divergence is extreme (top left) and immediately noticeable, there is a broad range of papers for which the difference is still very significant from a practical perspective. Indeed, the very dense area (bands 1-50) includes many excellent papers (CC numbers of around 40 are high, and even more considering that we only have citations from ACM papers). Even in that area, there are many papers for which the band numbers differ significantly if they are ranked by CC or PR.

In the subsequent discussion, we will qualitatively refer to papers with high PR and high CC as *popular gems*, to paper with high PR and low CC as *hidden gems*, to papers with low PR and high CC as *popular papers*, and to papers with low CC and PR as *dormant papers* (which is an optimistic term, on the assumption that they are going to be noticed sometime in the future).

[2] http://www.dit.unitn.it/~krapivin/exploring_citation%20_metrics_08.html

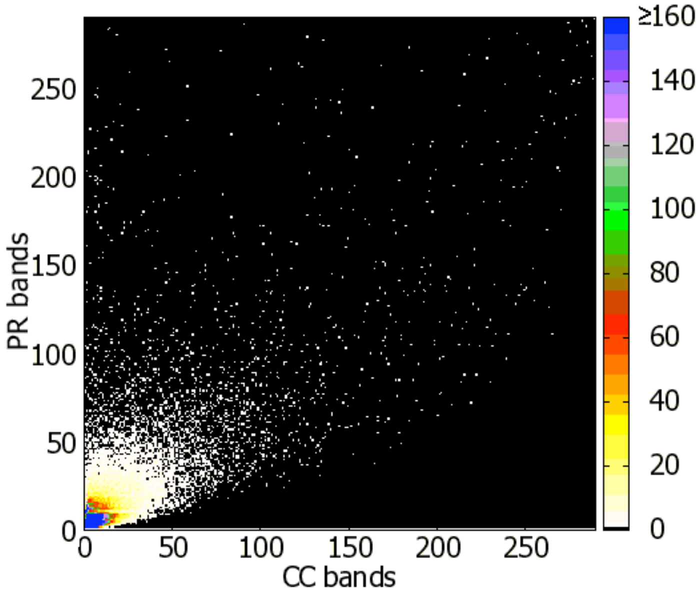

Fig. 1. CC vs PR. X axis plots CC bands, Y axis plots PR mirror-banded by CC.

4.2 Divergence

The plots and table above are an attempt to see the difference among metrics, but it is hard from them to understand what this practically means. We next try to quantitatively assess the difference in terms of *concrete effects* of using a metric over another for what metrics are effectively used, that is, ranking and selection. In the literature, the typical metric for measuring a difference between two rankings is the Kendall τ distance [13], measured as the number of steps needed to sort bi-ranked items so that any pair *A* and *B* in the two rankings will satisfy to the condition:

$$sign(R_1(A) - R_1(B)) = sign(R_2(A) - R_2(B))$$

where R_1 and R_2 are two different rankings. However, this measure does not give us an indication of the practical impact of using different rankings, both for searching papers and, as we will see later, for authors. What we really want to understand is to see the distance between two rankings based on the actual paper search patterns. Assume we are searching the Web for papers on a certain topic or containing certain words in the title or text. We need a way to sort results, and typically people will look at the top result, or at the top 10 or 20 results, disregarding the rest. Hence, the key metric to understand divergence of the two indexes is how often, on average, the top *t* results would contain different papers, with significant values for *t* = 1, 10, 20. For example, the fact that the papers ranked 16 and 17 are swapped in two different rankings is considered by the Kendall distance, but is in fact irrelevant from our perspective.

To capture this aspect, we propose a metric called *divergence*, which quantitatively measures the impact of using one scientometric index versus the other. Consider two metrics M_1 and M_2 and a set of elements (e.g., of papers) S. From this set S, we take a subset n of elements, randomly selected. For example, we take the papers related to a certain topic. These n papers are ranked, in two different rankings, according to two metrics M_1 and M_2, and we consider the top t elements. We call divergence of the two metrics, $Div_{M1,M2}$ *(t,n,S)*, the average number of elements that differ between the two sets (or, *t* minus the number of elements that are equal). For example, if *S* is our set of ACM papers, and *n* are 1000 randomly selected papers (say, the papers related to a certain topic or satisfying certain search criteria), $Div_{CC,PR}$ *(20,1000,S)* measures the average number of different papers that we would get in the typical 20-item long search results page. We measured the divergence experimentally for CC and PR, by sampling repetitively 1000 documents randomly from our dataset. We then stopped as soon as the measured divergence metric converged. Results are collected in the Table 1 below. As a particular case, $Div_{M1,M2}$ *(1,n,S)* measures how often does the top paper differs with the two indexes. The table is quite indicative of the difference, and much more explicit than the plots or other evaluation measures described above. In particular, the table shows that ca. 62% of the times, the top ranked paper differs with the two metrics.

Table 1. Experimentally measured divergence for the set of ACM papers.

t	$Div_{PR,CC}$ (t, 1000, S), in %	$Div_{PR,CC}$ (t, 1000, S),
1	62.40	0.62
10	49.94	4.99
20	46.42	9.28
40	43.29	17.31
60	42.51	25.5
80	41.75	33.39
100	40.52	40.52

Furthermore, and perhaps even more significantly, for the traditional 20-element search result page, nearly 50% of the papers would be different based on the metric used. This means that the choice of metric is very significant for any practical purposes, and that a complete search approach should use both metrics (provided that they are both considered meaningful ways to measure a paper). In general we believe that divergence is a very effective way to assess the difference of indexes, besides the specifics of CC and PR. We will also see the same index on authors, and the impact that index selection can therefore have on people's careers. Details on the experiments for producing these results and the number of measures executed are reported in the companion web page.

4.3 Understanding the differences.

We now try to understand *why* the two metrics differ. To this end, we separate the two factors that contribute to PR: (1) the PR measure of the citing papers and (2) the number of outgoing links of the citing papers.

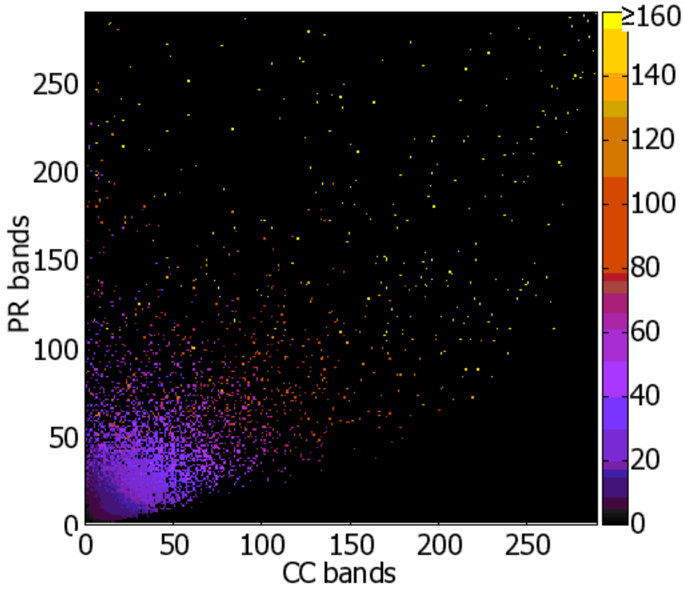

Fig. 2. Average potential weight

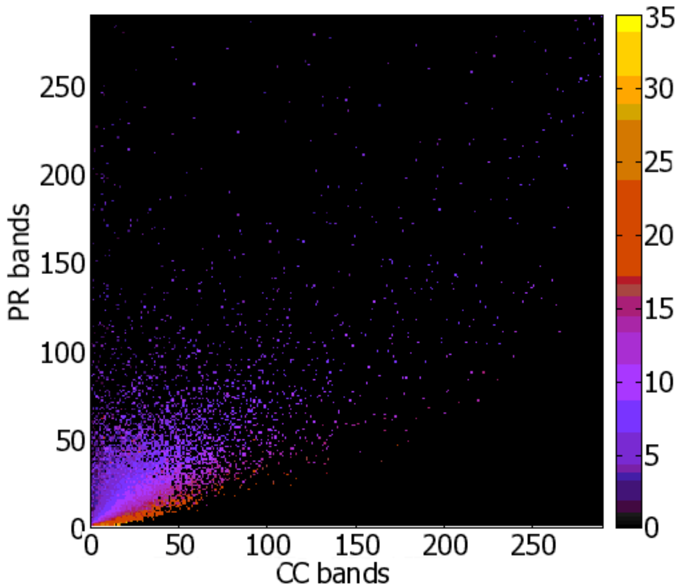

Fig. 3. Average dispersed weight

To understand the impact of the first factor, we consider for each paper P the weight of the papers citing it (we call this the *potential weight*, as it is the PR that the paper would have if all the citing papers P only cited P). We then plot (Figure 2) the average potential weight for the papers in a given square (intersection of a CC and a PR band) in the banded chart.

The estimation of the impact of outgoing links can be done in various ways. For example, we can take the same approach as for the computation of the previous potential weight and compute a double average over the outgoing links (for each paper P, compute the average number of outgoing links of the set C(P) of papers citing P, and then average them for all papers of a square in the CC vs PR chart). This is useful but suffers from the problem that if some papers (maybe "meaningless" paper with very low PR, possibly zero) have a very high number of outgoing links, they may lead us to believe that such high number of links may be the cause for a low PR value for a paper, but this is not the case (the paper is only loosing very few PR points, possibly even zero, due to these outgoing links). A high value of this measure therefore is not necessarily indicative of the number of outgoing links being a factor in low values of PR. Again, examples of these plots can be found in the companion web page.

What we really want to see when examining the effect of outgoing links from citing paper is the "*weight dispersion*", that is, how much weight of the incoming papers (i.e., how much potential weight) is dispersed through other papers as opposed to being transmitted to P. This is really the measure of the "damage" that outgoing links do to a PaperRank. We compute the dispersed weight index for a paper P (DW(P)) as the sum of the PR of the citing papers C(P) (that is, the potential weight of P) divided by the PR of P (the actual weight). Figure 3 plots the average dispersed weight for each square, again in a banded chart by CC and PR. The dark area in the bottom right corner is because there are no papers there.

Figure 2 and 3 tell us that outgoing links are the dominant effect for the divergence between CC and PR. Papers having a high CC and low PR have very high weight dispersion, while papers with high PR and low CC are very focused and able to capture nearly all potential weight. The potential weight chart (Fig. 2) also tends to give higher numbers for higher PR papers, but the distribution is much more uniform: i.e. there are papers in the diagonal or even below the diagonal and going from the top left to the bottom right the values do change but not in a significant way (especially when compared to the weight dispersion chart in Fig 3).

To see the difference concretely on a couple of examples, we take a "hidden gem" and a "popular paper" (see Fig. 4 and Fig. 5). The specific gem is the paper Computer system for inference execution and data retrieval, by R. E. Levien and M. E. Maron, 1967. This paper has 14 citations in our ACM-only dataset (Google Scholar shows 24 citations for the same paper). The PR of this "hidden gem" is 116.1, which is a very high result: only 9 papers have a greater rank. Figure 6 shows all the incoming citations for this paper up to two levels in the citation graph. The paper in the center is our "gem", because a heavyweight paper - that also has little dispersion (only two papers)- cites it. We point out here, that this also means that in some cases a pure PR may not be robust: the fact that our gem is cited by a heavyweight paper may be considered a matter of "luck" or a matter of great merit, as a highly respected "giant" is citing it.

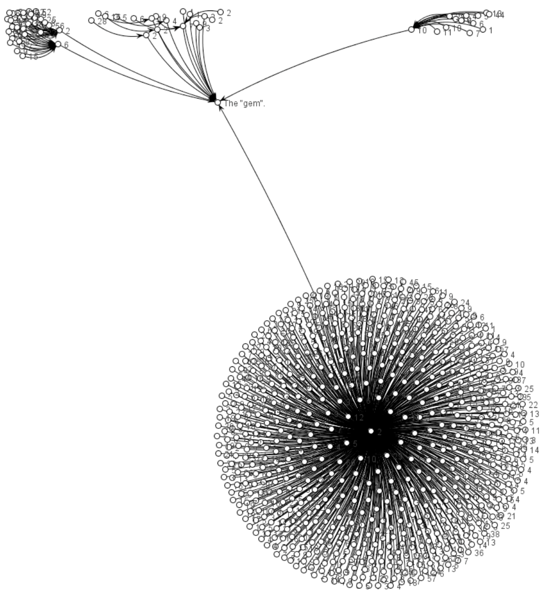

Fig. 4. “Hidden gem” in the dataset

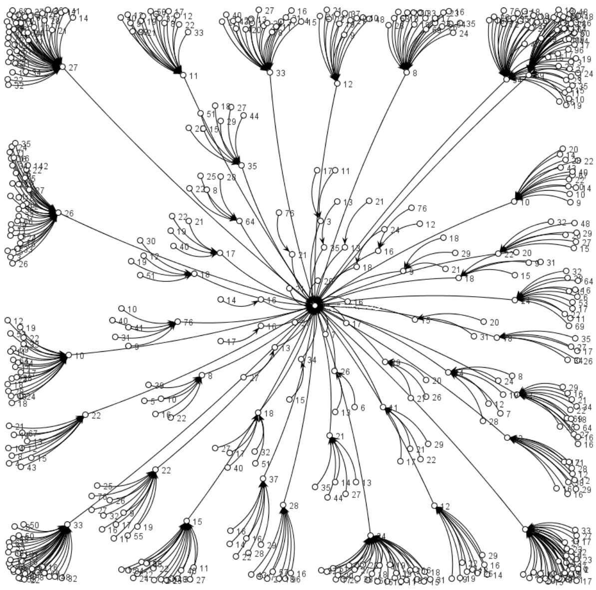

Fig. 5. "Popular paper" (in the center).

Again, discussing quality of indexes and which is “better” or “worse” is outside our analysis scope, as is the suggestion for the many variations of PR that could make it more robust.

We now consider a paper in the bottom of the CC vs PR plot, a paper with high number of citations but relatively low PR. The corresponding citation graph is shown in Fig. 5. This paper has 55 citations in our ACM dataset (158 citations in Google Scholar) and a relatively poor PR of 1.07. Comparing with papers in the same CC and PR band, this paper has a weight dispersion factor that is over twice that of papers in the same CC band and three times the one of papers in the same PR band, which explain why the increased popularity with respect to papers in the same PR band did not correspond to a higher PR.s

As a final comment, we observe that very interestingly there are papers with very low CC and very high PR, but much less papers - almost none - with very high CC and very low PR. If we follow the dispersion plot this is natural, as it would assume that the dispersed weight should be unrealistically high (many papers with hundreds of citations) which does not happen in practice, while it is possible to have "heavyweight" papers with very few citations that make the presence of paper gems (papers in the top left part) possible.

However, we believe that the absence of papers in the bottom right part and, more in general, the skew of the plot in Figure 1 towards the upper left is indicative of a "popularity bias". In the ideal case, an author A would read all work related to a certain paper P and then decide which papers to reference. In this case, citations are a very meaningful measure (especially if they are positive citations, as in the motto "standing on the shoulders of giants"). However this is impossible in practice, as nobody can read such a vast amount of papers. What happens instead is that author A can only select among the papers she "stumbles upon", either because they are cited by other papers or because they are returned first in search results (again often a result of high citation count) or because they are published in important venues. In any event, it is reasonable to assume that authors tend to stumble upon papers that are cited more often, and therefore these papers have a higher chance of being cited than the "hidden gems", even if maybe they do not necessarily have the same quality. We believe that it is for this reason that over time, once a paper increases with citation count, it necessarily increases with the weight, while gems may remain "hidden" over time. A detailed study of this aspect (and of the proper techniques for studying it) is part of our future work.

5 Exploring author metrics

5.1 Divergence

The same measure of divergence described for papers can be computed for authors. The only difference is that now the set S is a set of authors, and that the indexes are H-index and PRH instead of CC and PR. We also compute it for n=100, as the experiment we believe it is meaningful here is to consider replies to a typical job posting for academia or a research lab, generating, we assume, around 100 applications. Statistics for other values of n and more plots are reported in the companion web page (see previous footnote). Although nobody would only make a decision based on indexes, they are used more and more to filter applications and to make a decision in case of close calls or disagreements in the interview committees.

Table 2 tells us that almost two third of the times, the top candidate would differ. Furthermore, if we were to filter candidates (e.g., restrict to the top 20), nearly half of the candidates passing the cutoff would be different based on the index used

Table 2. Divergence between PRH and H, n=100.

t	$Div_{PRH,H}(t)$
1	59.30%
5	50.04%
10	46.13%
20	43.47%

This fact emphasizes once again that index selection, even in the case of both indexes based on citations, is key to determining the result obtained, be them searching for papers or hiring/promotion of employees. Notice also that we have been only looking at differences in the elements in the result set. Even more are the cases where the ranking of elements differ, even when the *t* elements are the same.

Another interesting aspect is that the divergence is very high. This is because most of the authors have a very low H and PRH. However, when we go to higher value of H and PRH, numbers are lower and the distribution is more uniform, in the sense that there are authors also relatively far away from the diagonal. Incidentally, we believe that this confirms the quality of divergence as a metric in terms of concretely emphasizing the fact that the choice of index, even among citation-based ones, has a decisive effect on the result.

6 Conclusions and future work

This paper has explored and tried to understand and explain the differences among citation-based indexes. In particular, we have focused on a variation of Page Rank algorithm specifically design for ranking papers – that we have named Paper Rank - and compared it to the standard citation count index. Moreover, we have analyzed related indexes for authors, in particular the Paper Rank Hirsch–index and the commonly-used H-index. We have explored in details the impact they can have in ranking and selecting both papers and authors. The main findings of this paper are:

1) PR and CC are quite different metrics for ranking papers. A typical search would return half of the times different results.

2) The main factor contributing to the difference is weight dispersion, that is, how much weight of incoming papers is dispersed through other papers as opposed to being transmitted to a particular paper.

3) For authors, the difference between PRH and H is again very significant, and index selection is likely to have a strong impact on how people are ranked based on the different indexes. Two thirds of the times the top candidate is different, in an average

application/selection process as estimated by the *divergence*. In addition to the findings, we believe that:
1) PR and PRH are complementary citation-based metrics capable of capturing the relevant information present in the whole citation network, namely the “weight” (the reputation or authority) of a citing paper and its “density of citations”.
2) Divergence can be a very useful and generally applicable metric, not only for comparing citation networks indexes, but also for comparing any two network-based ranking algorithms based on practical impact (results).
3) In our citation network, there are a significant number of “hidden gems” while there are very few “popular papers” (non gem). The working hypothesis for this fact (to be verified) is that this is due to citation bias driven by a “popularity bias” embedded in the author’s citation practices, i.e. authors tend to stumble upon papers that are cited more often, and therefore these papers have a higher chance of being cited.

The exploration of the citation bias hypothesis is our immediate future research, along with the extension of our dataset to a more complete coverage of the citation graph, to analyze the possible influence on the different indexes

Acknowledgments. This work has been supported by the LiquidPub project (FP7 FET-Open 213360.).

References

1. deSolla Price, D.: Little Science - Big Science. Columbia Univ. Press, New York, (1963)
2. Garfield, E.: Citation Indexing, ISI Press (1979)
3. Glänzel, W.: Bibliometrics as a research field, A course on theory and application of bibliometric indicators, Magyar Tudományos Akadémia, Course Handouts (2003)
4. Hirsch J. E.: An index to quantify an individual's scientific research output. In: Proceedings of the National Academy of Sciences, 102(46) pp.16569--16572, (2005)
5. Chen, P., Xie, H., Maslov S., Redner, S. “Finding Scientific Gems with Google”, Journal of Informetrics, Volume 1, Issue 1, pp. 8—15,Elsevier, (2007).
6. Brin S., Page L.: The anatomy of a large-scale hypertextual Web search engine. In: Computer Networks and ISDN Systems , Volume 30, Issues 1-7, pp. 107--117Elsevier, (1998)
7. Bianchini M., Gori M., Scarselli F.: Inside PageRank, ACM Transactions on Internet Technology, Vol. 5, No. 1, (2005)
8. Del Corso, G. M. Gull A., and Romani,F.: Fast PageRank Computation via a Sparse Linear System”, Internet Mathematics Vol. 2, No. 3, 251-273, (2005)s
9. Diligenti M., Marco Gori, M. Maggini, M.: Web Page Scoring Systems for Horizontal and Vertical Search, In: WWW Conference 2002, USA, May 7-11 (2002)ß
10. Sun Y. and Giles, L. C. : Popularity Weighted Ranking for Academic Digital Libraries. In: 29th European Conference on Information Retrieval 2007, pp. 605--612, Rome, Italy (2007)

11. Bernstam E.V., Herskovic J.R., Aphinyanaphongs Y.:Using Citation Data to Improve Retrieval from MEDLINE, Journal of the American Medical Informatics Association , Elsevier, Volume 13, Issue 1, Pages 96-105, (2006)
12. Langville A. N., Meyer C. D., "Deeper Inside PageRank". Internet Mathematics Vol. 1, No. 3, pp. 335--380, (2004)
13. Kendall M., Gibbons J. D.: Rank Correlation Methods. Edward Arnold, London (1990)ß