On a joint Research Publications and Authors Ranking

Dohy Hong, François Baccelli

On a joint Research Publications and Authors Ranking

Dohy Hong* François Baccelli†

Abstract

The paper introduces a new analysis technique for evaluating research activities which is based on a random walk on the bipartite graph of papers and authors. This technique is an extension of the PageRank family algorithm to this setting. It leads to a new ranking algorithm where the ranking of a paper/author depends on that of the papers/authors citing it/him or her. We compare the results against existing ranking methods through the analysis of simple scenarios.

Keywords: publication, citation, ranking, PageRank graph, random walk.

1 Introduction

There have been many publications (e.g. [12, 17, 14, 9, 16, 6]) presenting extensions of PageRank ideas to the context of the publication citation graph.

In this paper, we continue the existing work on the matter by considering a parameterized random exploration of the bipartite paper-author graph, which is defined below. The author graph alone has been considered in e.g. [12, 17, 8, 13, 4, 3]. The paper graph is also classical [2, 10, 5]. The joint graph was already considered in [9, 16, 14, 15]. We explain below how and why our approach differs from and continues these earlier approaches. In essence, all these have been limited to *local* properties of this bipartite graph or based on an exploration of a limited range on the relationships. For instance, the number of citations of a paper is its paper indegree. That of an author is the sum of the citations of its papers. Similarly, the H-index of an author and related indices are also local though non-additive metrics of this graph since they are also determined by the number of papers citing the paper of the considered author. The existing approaches on the author graph or the paper graph alone can be also interpreted as particular cases of our approach where the exploration possibility is arbitrarily limited.

Local characteristics such as the number of citations (the one that is used on the web site [1]) or the number of publications are of course relevant. They may give a good indication of the research activity. However, we believe that the existing metrics are not qualitative enough. By qualitative, we mean metrics where the ranking of a paper/author depends on that of the papers/authors citing it/him or her.

In §2, we present our model and in §3, we illustrate and compare different approaches through simple examples. Some simulation results are shown in §4.

2 Model

2.1 Graph structure We consider the bipartite graph $\mathcal{G}$ induced by the citation graph of papers and their authors: the nodes in this graph represent either an author ($\mathcal{A}$) or a paper ($\mathcal{R}$). The edges between nodes are naturally defined by the relationships:

- a paper $r \in \mathcal{R}$ is written by $a \in \mathcal{A}$;
- a paper r references a paper r'.

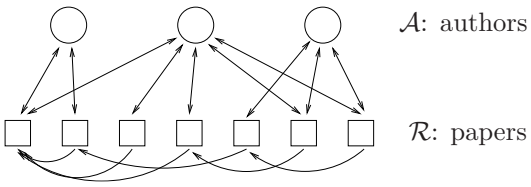

Figure 1: An example of paper-author graph.

Figure 1 shows an example of paper-author graph where papers are represented by squares and authors by circles. All author-paper edges (wrote/written by) are bidirectional, whereas all edges between papers (paper a citing paper b) are directional.

In the following, we call paper graph the graph obtained when only considering the nodes representing the papers. The author graph (see also definition in [12]) is the one obtained by directly linking authors when an author is cited by another one through a paper reference.

In the following, we call PR-G the PageRank extension (cf. §2.2) algorithm applied to the global paper-author graph and PR-A (resp. PR-P) the PageRank algorithm applied to the author (resp. paper) graph.

*Alcatel-Lucent Bell Labs Research, Route de Villejust, 91620 Nozay, France. E-mail: dohy.hong@alcatel-lucent.com

†INRIA, 23 avenue d'Italie, 75013 Paris, France. E-mail: francois.baccelli@inria.fr

2.2 Random walk The random walk algorithm we propose, PR-G, is defined as follows:

- Step0 (Global initialization): choose a node type;
- Step0.1 (Author initialization): choose an author a (by default uniformly); go to Step1;
- Step0.2 (Paper initialization): choose a paper r (by default uniformly); go to Step2;
- Step1 (from an author): choose a paper r written by a;
- Step2 (from a paper r): choose a paper r' cited by r; if none (or no information), return to Step0.2;
- Step2b: with probability θ return to Step2 with r'; otherwise, go to Step3;
- Step3: choose one of the authors a' of paper r'; if none (no information available), return to Step0.1;
- Step4: go to Step1 with a'.

At each step of the above random walk, we increment by one a counter $C(i)$ associated to the node i which is visited (a node can be a paper or an author).

There are several minor variants of the algorithm:

- above, the action "choose" is by default based on a uniform sampling.
  - in Step1, another option is to use weights; the weight of each paper may be inversely proportional to the number of co-authors;
  - In Step2, we may exclude the papers written by a co-author of a);
- in Step1, one can decide not to increment by one the counter associated to r; this depends on whether or not one wishes the evaluation of the co-authors to have a direct impact on the paper (which could be seen as a self appreciation or associated with the assumption that a good author is more likely to write a good paper);
- the test on θ in Step2b can be done at the beginning of Step2: this implies the possibility of a two hop jump from an author to a co-author through one of their common papers;
- we can also apply a global damping factor with probability d (as in the initial PageRank algorithm idea), namely at each step, one either executes the step as prescribed with probability $(1-d)$ and one reinitializes the random walk with probability d. An intuitive way to describe the role of the damping factor for PageRank on the web is to use a random surfer model where the damping factor would be the probability that the surfer gets bored after several clicks and switches to a random page of the web graph. The damping factor:
  - makes the Markov chain associated to the random walk irreducible (i.e. we have a single connected component);
  - prevents the random walk to stay too long in a *trap* position (a small group of nodes from which there are no outgoing links).

  The damping factor has also a direct influence on the convergence speed of the algorithm.

As for PageRank (on web pages), the aim of the above iterations is to evaluate the importance or the rank of each node, where the rank of a node is defined as the stationary probability that the random walk is located at this node.

Let us focus on the case with damping factor. In this case, this stationary distribution is the N dimensional vector Π solution of the following eigenvector problem:

$$A.\Pi = \Pi,$$

where $\Pi(i)$ is the stationary probability that the walk is in node i and where A is the $N \times N$ matrix defined as follows: if there are directed links from j, then

$$A(i,j) = \frac{1-d}{N} + d\frac{1}{N(j)}1_{j\to i},$$

where N is the total number of nodes, $N(j)$ is the number of outgoing links from j and $1_{j\to i}$ denotes the fact that there is a directed link form j to i. If there are no directed links from j, then $A(i,j) = 1/N$ for all i.

In our case, because nodes are of two types, we can reformulate $A(i,j)$ depending on the type of j as follows (here we assume that there are links from j and that the θ test is done at the beginning of Step2 to illustrate this reformulation):

$$\begin{aligned}
&\text{if } j \in \mathcal{A}, i \in \mathcal{R}: \\
&\qquad A(i,j) = \frac{1-d}{N} + \frac{d}{N_1(j)}1_{j \text{ authors } i}, \\
&\text{if } j \in \mathcal{A}, i \in \mathcal{A}: \\
&\qquad A(i,j) = \frac{1-d}{N},
\end{aligned}$$

and

$$\text{if } j \in \mathcal{R}, i \in \mathcal{R}:$$
$$A(i,j) = \frac{1-d}{N} + \frac{\theta d}{N_2(j)} 1_{j \text{ cites } i},$$
$$\text{if } j \in \mathcal{R}, i \in \mathcal{A}:$$
$$A(i,j) = \frac{1-d}{N} + \frac{(1-\theta)d}{N_3(j)} 1_{i \text{ co-authors } j},$$

where $N_1(j)$ is the number of papers authored by j, $N_2(j)$ is the number of references in paper j and $N_3(j)$ is the number of co-authors of paper j.

This eigenvector problem can also be seen as follows: it defines the importance of a paper or an author as a linear combination of the importances of the nodes (papers or authors) pointing to it in the paper-author graph. This is precisely the qualitative property stressed above. The ranking itself is then obtained by comparing/sorting the importances.

Thanks to the ergodic theorem for Markov chains, the ranking of node i, $\Pi(i)$ can be evaluated as the empirical frequency of the visits of the random walk at node i (see e.g. [11, 7]), that is through the counters used in connection with the above random walk algorithm.

Special cases Here are a few special cases:

- if $\theta = 0$, we obtain the author graph (cf. [12, 17]). The strict positiveness of θ plays an important role in PR-G. If we don't navigate the paper graph from a paper node (Step2b) more than once, we loose an important qualitative aspect of ranking resulting from the citation graph that we illustrate by a simple example below.

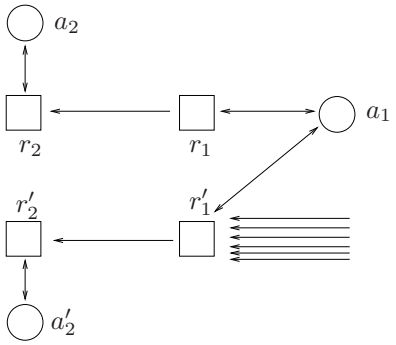

Figure 2: Illustration of the paper weight.

Assume a paper r_1 written by a_1 and citing a paper r_2 written by a_2. We assume that no one (else) cites r_1 and r_2. Now consider another paper r'_1 by a_1 and yet another paper r'_2 by a'_2. Assume that only r'_1 cites r'_2 and that r'_1 is cited by a large amount of other papers (cf. Figure 2). Assume that a_2 and a'_2 didn't write anything else. Then, in the author graph, a_2 and a'_2 have the same weight and r_2 and r'_2 have the same weight determined from that a_1. But qualitatively, the paper r'_2 should be better ranked than r_2 (also a'_2 better than a_2). Choosing $\theta > 0$, this is taken into account.

- Tahing $\theta = 1$ puts more emphasis on the paper graph; in [14], Step1 and Step2b are replaced by a uniform distribution of weights from papers to author (summing papers of an author) and from authors to papers, whereas our approach defines one global random walk where the parameter θ controls the emphasis of the paper graph within the global paper-author graph; the general drawback of only considering the paper graph is putting too much emphasis on the old papers (without control) in a configuration such as: p_n cites p_{n-1} who cites p_{n-2} etc. Then p_1 is inheriting weight from all future papers p_i, even if this effect can be bounded by playing with the damping factor (cf. §3.4).

3 Comparison and Analysis

3.1 First scenario S1: difference with H-index

We consider 2 groups of authors: group $A = \{a\}$ is reduced to one author, having written a single paper, and group B has a population size of N. We assume that each papers in group B has a single author and that each paper cites m papers selected as follows:

- with probability $p = s_b/(Nn)$ (s_b is a constant and n is the average number of publications of an author in group B, which is also assumed to be constant; we are interested in the asymptotic results for large N), the paper of a is referenced and the remaining $m-1$ references are to papers in group B;
- with probability $1-p$: all m papers are from group B.

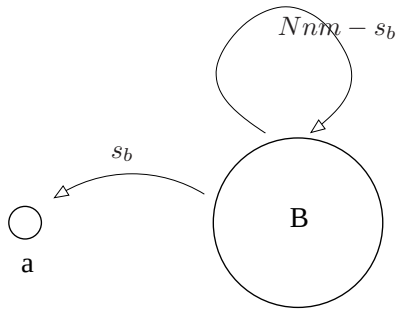

Figure 3: S1: simplified paper graph of the two groups.

In a random walk on the paper-author graph, the probability that, from a paper of B, the random walk jumps to the paper in group A by citation (Step2) is equal to p/m. The results of Table 1 summarize the results (up to a multiplicative constant and asymptotically for large N; for the H-index, we assume as a first approximation that each paper of B is cited m times, which is only true in average):

	a	$b \in B$
Nb of citations	s_b	mn
H-index	1	$m \wedge n$
PR-G	s_b	mn

Table 1: Scenario S1 (average per author).

For instance, if $s_b = 1000$, $m = 10$, $n = 10$, we get:

	a	$b \in B$
Nb of citations	1000	100
H-index	1	10
PR-G	1000	100

Table 2: Scenario S1: numerical example.

which means that a typical person b in group B has 100 citations with H-index 10, whereas a has 1000 citations with H-index 1. In this specific case, by construction, the random walk evaluations on the paper graph or the author graph or the bipartite graph all give the same results.

3.2 Second scenario S2: difference with the number of citations We introduce this second scenario to show how the PR-A/P/G variants can differ from the number of citations. We consider 3 groups of authors: group $A = \{a\}$ is reduced to one author having written a single paper. Groups B and C have a population size of N_B and N_C, respectively. We set $N = 1 + N_B + N_C$.

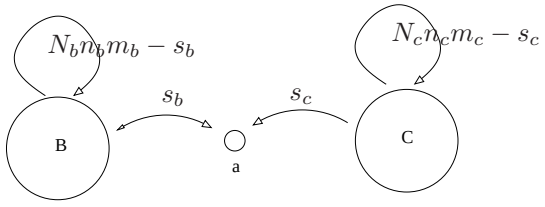

Figure 4: S2: simplified paper graph of the three groups.

We assume that an author in group B (resp. C) wrote each of his/her papers alone that each paper references m_b (resp. m_c) papers:

- with probability $p_b = s_b/(N_B \times n_b)$ (resp. $p_c = s_c/(N_C \times n_c)$), one of the m_b (resp. m_c) is the paper of a, with n_b (resp. n_c) the average number of publications per author in group B (resp. C); the $m_b - 1$ (resp. $m_c - 1$) being of group B (resp. C);
- with probability $1 - p_b$ (resp. $1 - p_c$), all m_b (resp. m_c) are papers from group B (resp. C).

Therefore, in a random walk on this graph, the probability that from a paper of B (resp. C) we jump to a paper of A by citation is equal to: p_b/m_b (resp. p_c/m_c).

Then, we have the following results (up to a multiplicative constant for PR-G) when N_B and $N - C$ are large enough:

	a	$b \in B$	$c \in C$
Nb of citations	$s_b + s_c$	$m_b n_b$	$m_c n_c$
H-index	1	$m_b \wedge n_b$	$m_c \wedge n_c$
PR-G	1	$\frac{m_b n_b}{s_b}\frac{N_B}{N}$	$\frac{m_c n_c}{s_c}\frac{N_C}{N}$

Table 3: Scenario S2.

When N_B goes to infinity (with N_C fixed), we have the limit (PR-G is renormalized):

	a	$b \in B$	$c \in C$
Nb of citations	$s_b + s_c$	$m_b n_b$	$m_c n_c$
H-index	1	$m_b \wedge n_b$	$m_c \wedge n_c$
PR-G	s_b	$m_b n_b$	0

Table 4: Scenario S2: asymptotic values.

The parameters $s_b, s_c, m_b, m_c, n_b, n_c$ are free parameters (assuming N_B and N_C large enough). For instance, if $s_b = s_c = 100$, $m_b = m_c = 20$, $n_b = n_c = 20$, we have:

	a	$b \in B$	$c \in C$
Nb of citations	200	400	400
H-index	1	20	20
PR-A	100	400	0

Table 5: Scenario S2: numerical values.

Here, let us highlight the differences observed for the authors of group C: when the size of population B increases, their scores are constant for the H-index and

the number of citations, whereas their score goes to zero for PR-A/P/G.

The interpretation is simple: the authors of the sets A and B are better ranked than C because they are "alive", that is, they have a probability to be cited by the large population in B, whereas the activity of group C is "dead" in the sense that the activity of this group is separated from the global (B) activity.

3.3 Scenario S3: comparing PR-A/P/G In the scenario S3, we consider 6 papers $p_1, p_2, p_3, q_1, q_2, q_3$ authored respectively by a, b, c, a, b, d connected as indicated in Figure 5.

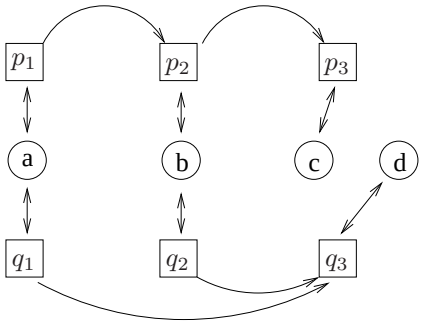

Figure 5: S3: paper-author graph (without the reinitialization edges).

In this scenario, if the damping factor is 0, then the eigenvector (Π) associated with the paper graph (cf. Figure 6) is:

$$(\pi_{p_1}, \pi_{p_2}, \pi_{p_3}; \pi_{q_1}, \pi_{q_2}, \pi_{q_3}) = (1, 2, 3; 1, 1, 3)\,\frac{1}{11}. \tag{3.1}$$

Figure 6: S3: paper graph.

In particular, in the paper graph, p_3 and q_3 are equally ranked. By summing the weights of all papers for each author, we get:

$$(\pi_a, \pi_b, \pi_c, \pi_d) = (2, 3, 3, 3)\,\frac{1}{11}, \qquad \text{(PR-P)}$$

We see that the authors b, c, d are not differentiated.

Now, the eigenvector corresponding to the author graph (cf. Figure 7) is:

$$(\pi_a, \pi_b, \pi_c, \pi_d) = (4, 6, 7, 9)\,\frac{1}{26}, \qquad \text{(PR-A)}$$

Here we see the differentiations between the authors b, c, d.

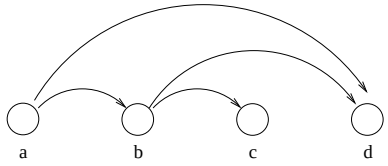

Figure 7: S3: author graph.

Now applying PR-G, we get (equalities up to a normalization constant):

$$\begin{aligned}
(\pi_a, \pi_b, \pi_c, \pi_d) &= (4, 6, 5+\theta, 7-\theta)\,, \qquad \text{(PR-G)}\\
(\pi_{p_1}, \pi_{p_2}, \pi_{p_3}) &= (2, 3+\theta, 5+\theta)\\
(\pi_{q_1}, \pi_{q_2}, \pi_{q_3}) &= (2, 3-\theta, 7-\theta)\,.
\end{aligned}$$

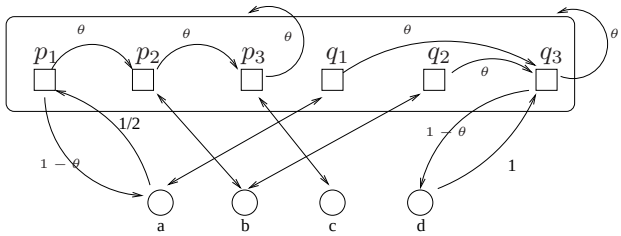

Figure 8: S3: bipartite graph with reinitialization. From a paper, we use the citation link with probability θ and go to the author space with $1-\theta$; when we reach the terminal papers p_3 and q_3 with probability θ we choose a random paper position. From an author, we choose one of the paper written by this author with equiprobability.

If we take $\theta = 1$, we find back the paper graph eigenvector (3.1). If $\theta = 0$, we have the following weights for authors: $(4, 6, 5, 7) \times 1/22$; we hence find a slightly different result than with PR-A. While the relative ranking of a and d is intuitive, the comparison between b and c depends on the way a paper with no citation is weighted. If one wishes to give less weight to papers without citation, one could jump to a random author position from the terminal papers p_3 and q_3.

To illustrate the difference of approaches, we slightly modify the scenario S3 by suppressing the node q_2 and adding a citation from p_2 to q_3 (Scenario S3b, cf. Figure 9). By such a transformation, the author graph is not modified.

However, for PR-G, we get (equalities up to a normalization constant):

$$\begin{aligned}
(\pi_a, \pi_b, \pi_c, \pi_d) &= (2, 2, 2, 3)\\
(\pi_{p_1}, \pi_{p_2}, \pi_{p_3}) &= (1, 2, 2)\\
(\pi_{q_1}, -, \pi_{q_3}) &= (1, -, 3)\,.
\end{aligned}$$

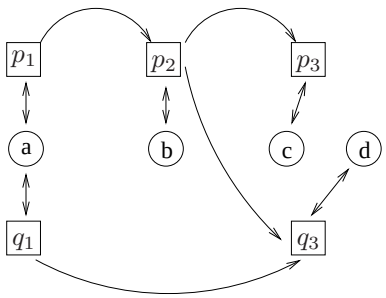

Figure 9: Scenario S3b.

In this result, we have the same ranking for a, b, c: b is cited once and not a, but a produced 2 papers, which in this particular example case is equivalent to b; knowing the equality between a and b, the equality of b and c is obvious: b and c are both cited once by the equally weighted authors a and b. In this case, the relative weights of papers with PR-G are the same as the ones with PR-P because there is a direct bijection (and this explains why we have scores not depending on θ) between papers and authors for b, c, d. For a the lack of dependence on θ comes from the equal probability to visit p_1 and q_1.

To further illustrate the difference of approaches, we modify again the scenario S3 by suppressing the node p_1 and adding a citation from q_1 to p_2 (Scenario S3c; cf. Figure 10). By such a transformation, the author graph is still not modified.

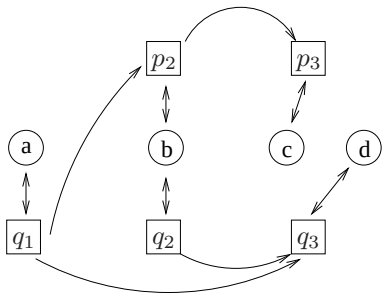

Figure 10: Scenario S3c.

However, for PR-G, we get (equalities up to a normalization constant):

$$\begin{aligned}(\pi_a, \pi_b, \pi_c, \pi_d) &= (4, 10, 9+\theta, 11-\theta)\\ (-, \pi_{p_2}, \pi_{p_3}) &= (-, 5+\theta, 9+\theta)\\ (\pi_{q_1}, \pi_{q_2}, \pi_{q_3}) &= (4, 5-\theta, 11-\theta)\,.\end{aligned}$$

From PR-P, we get in this case: $(-, 3, 5; 2, 2, 5)$: for PR-P, q_1 and q_2 are equality ranked (no citation); but for PR-G, we see that q_2 is better ranked, because q_2 is written by b who was cited for another paper. Also, q_3 is better ranked than p_3: q_3 was cited twice. For PR-P, p_3 receives the same score because it is cited by p_2 itself cited by q_1.

The examples above illustrate the qualitative gain of considering the author-paper graph. We see that the results of PR-G correspond better to our qualitative expectations.

3.4 About PR-G and PR-P The difference between PR-G and PR-P can be illustrated at least through two situations: the first one is when two papers are cited in the same way, but only differ by the importance of the authors citing directly or indirectly those two papers. PR-P won't differentiate them, whereas PR-G will.

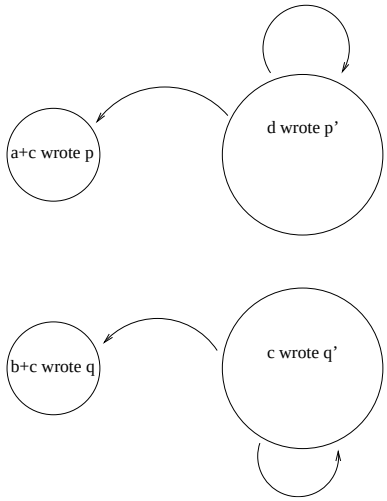

Figure 11: Comparison PR-P and PR-G: simplified graph.

An example is shown in Figure 11: authors a and c wrote p, b and c wrote q, d wrote p' (p' could be a set of papers) and c wrote q' (q' could be a set of papers).

With PR-P, a and b will be ranked equally, whereas PR-G (which explores the paper-author graph in Figure 12) will differentiate between them because c will be better ranked than d.

The second situation is shown in Figure 13: we illustrate a case where PR-P may give too high a weight to a paper at the end of a chain of citations (impact of indirect citations): with PR-P, a and b have differences depending only on the damping factor, whereas with PR-G, b will receive much higher a score because of Step1.

However, if the papers citing a were all written also by a, we would end up with both PR-G and PR-P

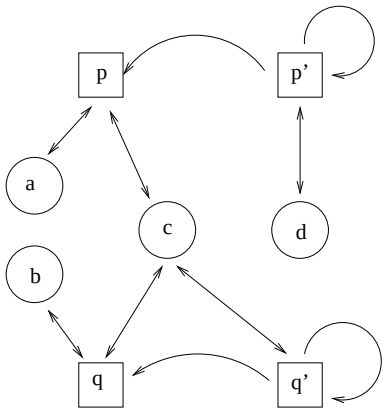

Figure 12: PR-G: paper-author graph.

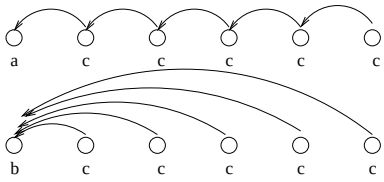

Figure 13: Comparison PR-P and PR-G: simplified graph where each circle is a paper written by a, b or c.

having a similar score.

4 Simulation Setting

In order to further illustrate our approach, we consider a paper-author graph generated by a simple model and evaluate it through simulations. We first generate the publications every time step, then associate the authors to each paper and build the citation graph. We use below a simple citation model where the probability to cite a paper has a dependence on the difference d of publication dates of the form: $1/d^{\alpha}$.

4.1 Simulation scenario S4 The simulation scenario S4 is shown on Figure 14: it is composed of four groups of authors:

- group G is the the *genius* group (whose work is cited by everybody);
- group A is the unique group that is cited by G and references only papers within A and B (and G);
- group B who references papers within A and B (and G);
- and group C who references papers only within its group C (and G).

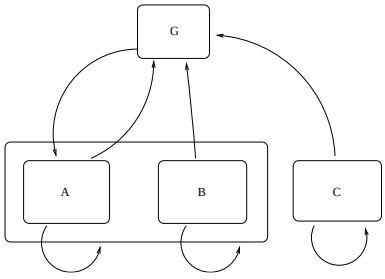

Figure 14: Scenario S4: simplified graph representation of author groups.

The groups A and B are meant to represent the majority of the research community (say within an activity domain). Group C represents an isolated group.

4.2 Simulation parameters Here are the simulation parameters:

- The number of time stepsis N_{time} (number of publication dates);
- The numbers of authors in each group are N_A, N_B, N_C, N_G, respectively;
- The number of published papers per time step is N_p; for the sake of simplicity we set $N_p = N = N_A + N_B + N_C + N_G$ and at each time step exactly 1 paper is published by each author (no co-author), so that the total number of publications per author is exactly N_{time};
- The number of references per paper is n_r for all authors, except for authors from C, for which it is n'_r;
- the number of references going to G is controlled by a probability q_g (applied on each reference).

5 Simulation Results

5.1 Comparison of all approaches We set $N_{time} = 120$, $N_G = 1, N_A = N_B = 100$, $N_C = 10$, $n_r = 20$, $n'_r = 40$, $q_g = 0.05$, $\alpha = 1.5$. For PR-G, we set $\theta = 0.7$.

In Table 6, we see that G has the largest number of citations (by a factor 5 to 12), and is the best ranked whatever the approaches. Also results for A and B are always close by. The only difference is here between A, B and C: all local evaluations (H-index, number of citations) give a higher score to C , whereas PR-A/P/G gives a lower score to C.

5.2 Comparison of PR-A/P/G In order to better understand the properties of the proposed ranking, we

	G	A	B	C
Nb. cit. per author	26077	2283	2265	4514
Avg citations (%)	4.958	0.434	0.431	0.858
H-index	117	23.34	23.23	40.6
PR-A (%)	4.675	0.495	0.451	0.061
PR-P (%)	4.423	0.482	0.443	0.294
PR-G (%)	4.647	0.491	0.447	0.149

Table 6: Comparison (S4) per author. For an easier comparison, the average values are per author (and per group) and in percentage (so that, multiplying the average values by the size of population in the group per column and summing those values we get 100).

now assume that (scenario S4b) the author g of group G writes two types of papers: the first one gathers papers that are cited by A, B, C and the second one gathers papers that are never cited. In A, the population is separated in 50+50 (A_1, A_2): the authors of the first subgroup are cited by the first type of papers of G and the second one by the second type. The results are shown in Table 7.

	G	A_1	A_2	B	C
Nb. cit. (%)	4.958	0.434	0.433	0.431	0.859
H-index	60	23.36	23.4	23.23	40.6
PR-A (%)	4.721	0.495	0.496	0.450	0.075
PR-P (%)	4.444	0.521	0.439	0.443	0.318
PR-G (%)	4.695	0.522	0.459	0.447	0.156

Table 7: Comparison (S4b) per author.

We see that only PR-P and PR-G can differentiate between A_1 and A_2. PR-G generally gives ranking results between PR-A and PR-P.

5.3 Analysis of the ranking dynamics Below we define a scenario S4c where at time step 120, we replace all new references to group G by references to group C. We can see on Figure 15 that PR-G quickly adapts the ranking to this change. Its scores are between those of PR-A and PR-P.

Figure 16 also illustrates the evolution of the ranking obtained by the different approaches. We focus here on the time interval $[10, 24]$. The citation rule, at each time step, is that defined above. We compare the evolution of the following ratio: the ranking scores of the papers of G published at time 10, divided by the average score of all papers published at time 10. We see that, for the number of citations, this ratio is close to 10 and is stable. For PR-P, this ratio starts a bit above the average value and takes much more time to converge; this is because the papers of G published at time 10 need time to be cited by others to gain rank. For PR-A, this ratio converges almost as quickly as the number of citations: this is because the scores of the papers in question are mainly based on the score of its author which is stable. Finally, for PR-G, this ratio is between PR-A and PR-P, taking both effects into account.

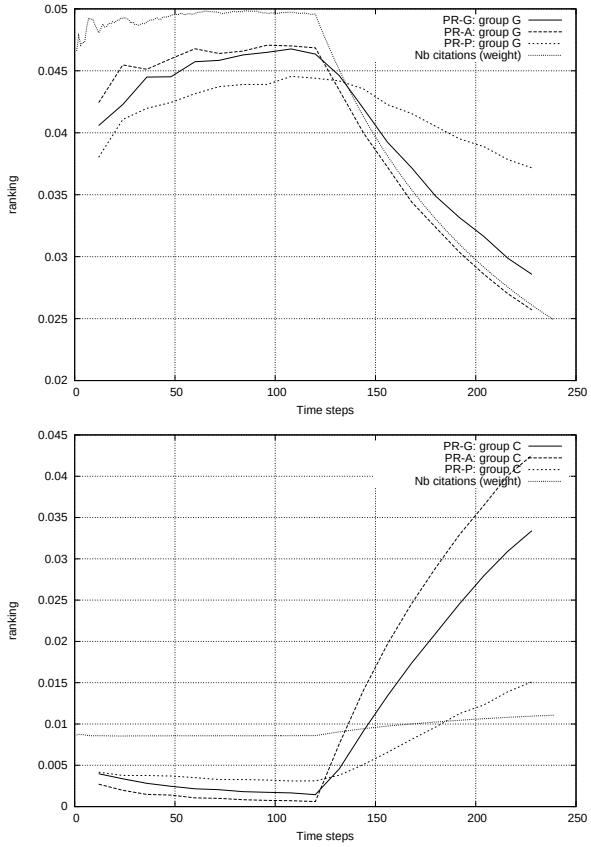

Figure 15: Scenario S4c. Top: ranking evolution of group G. Bottom: ranking evolution of group C.

6 Conclusion

In this paper, we proposed a global bipartite graph ranking algorithm for jointly ranking papers and authors. We compared this ranking mechanism to existing metrics through simple cases to illustrate the improvements brought by this type of ranking compared to the local metrics used today. This was done both in the static and the dynamic cases. The approach we proposed has the advantage of being global and of matching intuition.

We can also extend the approach to a general framework (for instance in the context of heterogeneous crowd generated contents) which will be the object of a future paper.

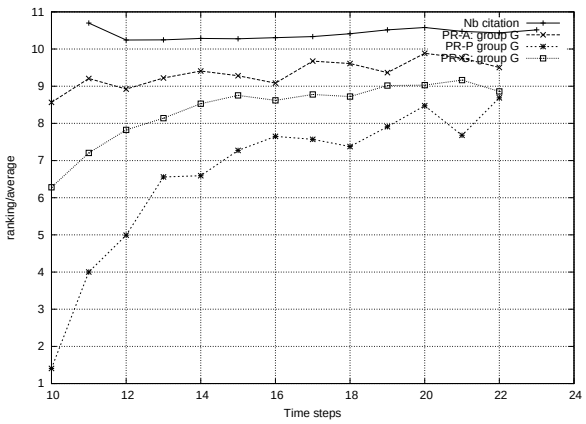

Figure 16: Scenario S4c: ranking dynamics of group G: score of the papers of G published at $t = 10$.

In a future work, we also hope to validate our approach through evaluations on large publication data bases.

References

[1] http://academic.research.microsoft.com.

[2] P. Chen, H. Xie, S. Maslov, and S. Redner. Finding scientific gems with google's pagerank algorithm. *Informetrics*, 1(1):8–15, January 2007.

[3] Y. Ding, E. Yan, A. R. Frazho, and J. Caverlee. Pagerank for ranking authors in co-citation networks. *CoRR*, 2010.

[4] D. Fiala, F. Rousselot, and K. Jeek. Pagerank for bibliographic networks. *Scientometrics*, 76(1):135–158, 2008.

[5] M. Gori and A. Pucci. Research paper recommender systems: A random-walk based approach. 2006.

[6] M. Krapivin, M. Marchese, and F. Casati. Exploring and understanding citation-based scientific metrics. *Advances in Complex Systems*, 13(1):59–81, 2010.

[7] A. N. Langville and C. D. Meyer. Deeper inside pagerank. *Internet Mathematics*, 1(3):335–400, 2004.

[8] X. Liu, J. Bollen, M. L. Nelsom, and H. V. de Sompel. Co-authorship networks in the digital library research community. *Inf. Process. Manage.*, 41(6):1462–1480, 2005.

[9] N. Ma, J. Guan, and Y. Zhao. Bringing pagerank to the citation analysis. *Information Processing and Management*, 44:800–810, 2008.

[10] S. Maslov and S. Redner. Promise and pitfalls of extending google's pagerank algorithm to citation networks. *The Journal of Neuroscience*, 28(44):11103–11105, October 2008.

[11] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. *Technical Report Stanford University*, 1998.

[12] F. Radicchi, S. Fortunato, B. Markines, and A. Vespignani. Diffusion of scientific credits and the ranking of scientists. *Phys. Rev.*, E 80, 2009.

[13] E. Yan and Y. Ding. Discovering author impact: A pagerank perspective. *Inf. Process. Manage.*, 47(1):125–134, 2011.

[14] E. Yan, Y. Ding, and C. R. Sugimoto. P-rank: An indicator measuring prestige in heterogeneous scholarly networks. *Journal of the American Society for Information Science and Technology*, 62(3):467–477, March 2011.

[15] J. Zhang, C. Ma, C. Zhao, J. Zhang, L. Yi, and X. Mao. A novel ranking framework for linked data from relational databases. *Tsinghua Science & Technology*, 15(6):642–649, December 2010.

[16] D. Zhou, S. Orshanskiy, H. Zha, and C. Giles. Co-ranking authors and documents in a heterogeneous network. *ICDM*, pages 739–744, 2007.

[17] K. Zyczkowski. Citation graph, weighted impact factors and performance indices. *Scientometrics*, 85(1):301–315, 2010.