Generalized Comparison of Graph-based Ranking Algorithms for Publications and Authors*

Antonis Sidiropoulos† Yannis Manolopoulos‡
asidirop@csd.auth.gr manolopo@csd.auth.gr

Data Engineering Lab,§ Department of Informatics,
Aristotle University, Thessaloniki, 54124 Greece

August, 2005

Abstract

Citation analysis helps in evaluating the impact of scientific collections (journals and conferences), publications and scholar authors. In this paper we examine known algorithms that are currently used for Link Analysis Ranking, and present their weaknesses over specific examples. We also introduce new alternative methods specifically designed for citation graphs. We use the SCEAS system as a base platform to introduce these new methods and perform a generalized comparison of all methods. We also introduce an aggregate function for the generation of author ranking based on publication ranking. Finally, we try to evaluate the rank results based on the prizes of '*VLDB 10 Year Award*', '*SIGMOD Test of Time Award*' and '*SIGMOD E.F.Codd Innovations Award*'.

1 Introduction

The ranking algorithms that are used in bibliometrics, in general can be separated into two classes. We refer to the first class as *collection-based ranking algorithms*. In this class a weighted citation graph is used. The graph nodes represent collections, whereas the weighted edges represent the total number of citations made from one collection to another. The ISI Impact Factor belongs to this ranking class [5, 6, 7] and uses the journals as collections. The collections could also be proceedings of conferences. In [15] we have presented a method alternative to Impact Factor, in which the *Collection Importance* is computed by a clustering algorithm. The latter work belongs in this class

*A shorter version of this paper is under submission to the Journal of Systems and Software
†http://delab.csd.auth.gr/~asidirop
‡http://delab.csd.auth.gr/~manolopo
§http://delab.csd.auth.gr/

of algorithms. Alternatively, the collections could be technical reports of universities and institutes or any other set of publications. In the this respect, the publications of an author could also be considered as a collection. Thus, we may rank authors by using this class of ranking.

On the other hand, there exists a class of *publication-based ranking* methods. According to this approach, the graph nodes represent publications, whereas an edge from node x to node y represents a citation from paper x to paper y. Computing the ranking at the publication level has the advantage that only a single process is performed to evaluate more than one entity: the paper itself, the collection it belongs to and, finally, the authors as well. The last two computations can be made by an aggregate average of the first one or by a more sophisticated function. All the ranking algorithms that where initially used to rank web pages belong in this class. Two of the most well known algorithms of this category are `PageRank` by Brin and Page [2] and Kleinberg's `HITS`, which ranks the elements as `Hubs` and `Authorities` [10, 12, 11]. Another widely accepted algorithm is `SALSA`, which is a variation of `HITS` [13].

In this paper we focus in the second class of ranking algorithms. In particular, the structure of this paper is as follows. In the next section we present known algorithms used for publication ranking. We also review their weak characteristics in the case of bibliometrics. In Section 3 we present a set of ranking approaches discussing their weaknesses and strengths. In particular, we examine variations of the SCEASRank algorithm [1], that has been introduced in [16], as well as variations of HITS and SALSA algorithms. In Section 4 we present the performed experiments. Specifically, we compare the computation speed of all algorithms and assess the ranking results, which are compared by using several methods. The used dataset is the contents of our SCEAS library. Finally, in Section 5 we evaluate the results hypothetically assuming: if we were to decide for the prizes of '*VLDB 10 Year Award*' and '*SIGMOD Test of Time Award*' by using one of the previously mentioned algorithms, could we be able to prize the correct (same) publications and authors? In the second part of the above section an aggregate function for ranking authors is presented. The author ranking is computed by using the publication ranking as input. The evaluation is made by comparing to '*SIGMOD E.F.Codd Innovations Award*'. The last section concludes the paper.

2 Ranking Methods

In this section we present known algorithms used for ranking web graphs. These algorithms could also be used in bibliometrics for citation graph-based ranking. Throughout this paper we use the symbols of Table 1 to present all the algorithms in a unifying way.

2.1 Citation Count

Ranking publications by counting the incoming citations is the simplest and fastest way. We refer to this algorithm as the Citation Count (`CC`). Thus, the score of a publication x is the in-degree of

[1]We have built a web-based library called SCEAS (standing for *Scientific Collection Evaluator by Advanced Scoring*) with data extracted from DBLP, and accessible through the url: *http://delab.csd.auth.gr/sceas*.

Table 1: Notations

I_x :	The set of Publications that cite x
$\|I_x\|$:	The number of Publications that cite x
O_x :	The set of Publications that are cited by x
$\|O_x\|$:	The number of Publications cited by x
d :	Damping factor (set to 0.85 for PageRank)
b :	Citation importance (usually set to 1)
a :	Exponential Factor (> 1, usually set to ϵ)

the graph node x:

$$CC_x \;=\; |I_x| \tag{1}$$

This unweighted ranking is the method that has been used for several years. However, this approach is questionable as all citations should not count the same. For instance, when a paper gets citations from good papers, then it should have a better ranking. This is the reason why several ranking algorithms have been introduced.

2.2 Balanced Citation Count

Another simple ranking method is the Balanced Citation Count (`BCC`). In this model, citations do not count equally but their importance is a function of the out-degree of the citing node:

$$BCC_x \;=\; \sum_{\forall y \in I_x} \frac{1}{|O_y|} \tag{2}$$

Here, the weight of a citation from node y is equal to $\frac{1}{|O_y|}$ rather than 1 as in the `CC` method (Equation 1). This means that the enforcement, which a node y gives to nodes that points to, sums up to 1 rather than $|O_y|$ as in `CC`. Thus, all the scores BCC_x sum up to $|V|$, which is the number of nodes in the graph. However, this method has the same disadvantage as `CC` does; there are no weights to represent the importance of the citing papers.

2.3 PageRank

`PageRank` takes into account the importance of the citing papers. Originally, the `PageRank` score, PR, has been defined as [2]:

$$PR(A) \;=\; (1-d) \;+\; d\,(\frac{PR(t1)}{C(t1)} + \ldots + \frac{PR(tn)}{C(tn)})$$

where $t1, \ldots, tn$ are pages linking to page A, C is the number of outgoing links from a page (out-degree) and d is a damping factor, usually set to 0.85. Using the symbols of Table 1, the last equation is equivalent to:

$$PR_x \ = \ (1-d) \ + \ d \sum_{\forall y \in I_x} \frac{PR_y}{|O_y|} \tag{3}$$

In simple words, `PageRank` assigns high score to a node if it is pointed by highly ranked nodes.

By definition `PageRank` gives high score to a node x, if there is a big connected component C where some of its nodes point to x. The more and larger cycles contained in C, the greater score x will get. This happens because there is feedback from nodes that belong into cycles to themselves. In bibliometrics, in this unusual case of existing cycles, they mainly represent self-citations. Consequently, it is not reasonable to let self-citations influencing the score. Removing the cycles will change the results. For example, Table 2 shows the rank results of Figure 1 graph (which consists of three connected components). We remark that node 0 gets 4 citations, whereas nodes 10 and 6 get 3 citations each. However, the `PageRank` score of nodes 10 and 6 is about 3 times higher than the score of node 0. This happens because nodes 10 and 6 are parts of citation cycles.

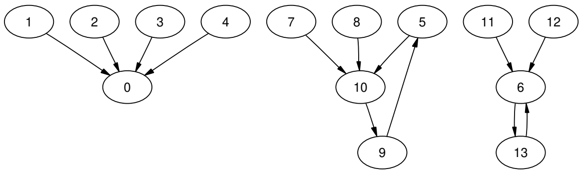


Figure 1: An example of 3 graphs.

Table 2: Rank results for the graphs of Figure 1.

CC	PR	HA	BHA	SA	BSA	P	PS	BPS	EPS	BEPS	SCEAS
0 4.000	6 1.919	0 1.000	0 1.000	0 0.658	13 0.803	5 0.333	6 2.757	13 39.000	0 1.150	0 1.472	0 1.472
6 3.000	13 1.781	6 0.000	6 0.000	6 0.493	6 0.596	13 0.333	10 2.506	6 37.000	6 1.030	6 1.433	6 1.433
10 3.000	10 1.661	10 0.000	13 0.000	10 0.493	10 0.000	6 0.111	0 2.183	5 33.000	10 0.993	10 1.356	10 1.356
9 1.000	9 1.562	5 0.000	10 0.000	5 0.164	9 0.000	9 0.111	13 2.051	9 31.000	13 0.584	13 0.895	13 0.895
5 1.000	5 1.477	9 0.000	9 0.000	9 0.164	5 0.000	10 0.111	9 1.913	10 31.000	9 0.573	9 0.867	9 0.867
13 1.000	0 0.660	13 0.000	5 0.000	13 0.164	0 0.000	0 0.000	5 1.590	0 4.000	5 0.452	5 0.687	5 0.687
8 0.000	1 0.150	1 0.000	1 0.000	1 0.000	1 0.000	1 0.000	1 0.000	1 0.000	1 0.000	1 0.000	1 0.000
7 0.000	2 0.150	2 0.000	2 0.000	2 0.000	2 0.000	2 0.000	2 0.000	2 0.000	2 0.000	2 0.000	2 0.000
4 0.000	3 0.150	3 0.000	3 0.000	3 0.000	3 0.000	3 0.000	3 0.000	3 0.000	3 0.000	3 0.000	3 0.000
3 0.000	4 0.150	4 0.000	4 0.000	4 0.000	4 0.000	4 0.000	4 0.000	4 0.000	4 0.000	4 0.000	4 0.000
2 0.000	7 0.150	7 0.000	7 0.000	7 0.000	7 0.000	7 0.000	7 0.000	7 0.000	7 0.000	7 0.000	7 0.000
12 0.000	8 0.150	8 0.000	8 0.000	8 0.000	8 0.000	8 0.000	8 0.000	8 0.000	8 0.000	8 0.000	8 0.000
11 0.000	11 0.150	11 0.000	11 0.000	11 0.000	11 0.000	11 0.000	11 0.000	11 0.000	11 0.000	11 0.000	11 0.000
1 0.000	12 0.150	12 0.000	12 0.000	12 0.000	12 0.000	12 0.000	12 0.000	12 0.000	12 0.000	12 0.000	12 0.000

A second characteristic of `PageRank` is that it is designed (specifically for the web) so that a page score is mostly affected by the scores of the pages that point to it and less by the number of the incoming links (i.e. in our case the number of citations). For example, in Table 3 (Figure 2), node 0 gets higher score than node 1, although node 1 gets 6 citations. This is another weak point for the case of bibliometrics.

Table 3: Rank results for the graph of Figure 2.

CC	PR	HA	BHA	SA	BSA	P	PS	BPS	EPS	BEPS	SCEAS
1 6.000	0 0.928	1 1.000	1 0.973	1 0.986	0 1.000	0 0.000	1 3.865	0 7.000	1 1.763	1 2.207	1 2.207
0 1.000	1 0.915	0 0.000	0 0.233	0 0.164	1 0.002	1 0.000	0 3.135	1 6.000	0 0.812	0 1.180	0 1.180
7 0.000	2 0.150	2 0.000	2 0.000	2 0.000	2 0.000	2 0.000	2 0.000	2 0.000	2 0.000	2 0.000	2 0.000
6 0.000	3 0.150	3 0.000	3 0.000	3 0.000	3 0.000	3 0.000	3 0.000	3 0.000	3 0.000	3 0.000	3 0.000
5 0.000	4 0.150	4 0.000	4 0.000	4 0.000	4 0.000	4 0.000	4 0.000	4 0.000	4 0.000	4 0.000	4 0.000
4 0.000	5 0.150	5 0.000	5 0.000	5 0.000	5 0.000	5 0.000	5 0.000	5 0.000	5 0.000	5 0.000	5 0.000
3 0.000	6 0.150	6 0.000	6 0.000	6 0.000	6 0.000	6 0.000	6 0.000	6 0.000	6 0.000	6 0.000	6 0.000
2 0.000	7 0.150	7 0.000	7 0.000	7 0.000	7 0.000	7 0.000	7 0.000	7 0.000	7 0.000	7 0.000	7 0.000

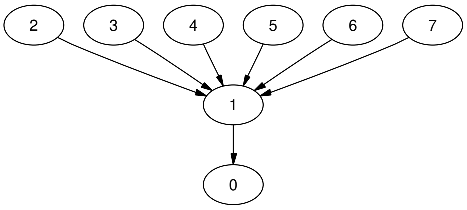
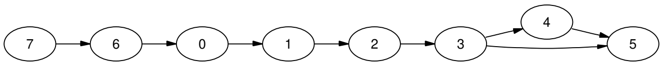
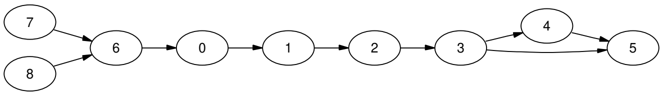


Figure 2: A second example of a graph.

A third characteristic is that a change in the score of a node j affects the score of node i even if their connecting path is very long. This case should also not arise in bibliometrics. In other words, adding a new node (publication) to our graph should impact its neighbors (the publications that the new paper cites). On the other hand, the scores of the rest nodes (the neighbors of neighbors etc.) should be much less impacted by the addition. For example, consider the graph of Figure 3(a), where a new node is added with an extra link to node 6, resulting in Figure 3(b). In this case, the score of node 5 increases by 7% and the score of node 4 increases by 6.8%, although they are 5 nodes away from each other. This is another case where `PageRank` has been proven not to behave smoothly in a bibliometrics environment.



(a)



(b)

Figure 3: A third example of graph.

Table 4: Rank results for the graph of Figure 3(a).

CC	PR	HA	BHA	SA	BSA	P	PS	BPS	EPS	BEPS	SCEAS
5 2.000	5 0.767	5 0.851	5 0.909	5 0.496	5 0.935	0 0.000	5 2.302	5 7.000	5 0.773	5 0.765	5 0.765
6 1.000	3 0.623	4 0.526	4 0.416	0 0.372	4 0.355	1 0.000	4 1.144	3 5.000	4 0.386	3 0.578	3 0.578
4 1.000	2 0.556	0 0.000	3 0.000	1 0.372	3 0.000	2 0.000	3 1.120	2 4.000	3 0.386	2 0.571	2 0.571
3 1.000	1 0.478	1 0.000	2 0.000	2 0.372	2 0.000	3 0.000	2 1.074	1 3.000	2 0.384	1 0.553	1 0.553
2 1.000	4 0.415	2 0.000	1 0.000	3 0.372	1 0.000	4 0.000	1 0.989	4 3.000	1 0.378	0 0.503	0 0.503
1 1.000	0 0.386	3 0.000	0 0.000	6 0.372	0 0.000	5 0.000	0 0.831	0 2.000	0 0.357	6 0.368	6 0.368
0 1.000	6 0.278	6 0.000	6 0.000	4 0.248	6 0.000	6 0.000	6 0.540	6 1.000	6 0.279	4 0.290	4 0.290
7 0.000	7 0.150	7 0.000	7 0.000	7 0.000	7 0.000	7 0.000	7 0.000	7 0.000	7 0.000	7 0.000	7 0.000

In Table 6 (Figure 4) `PageRank` ranks node 0 first and node 3 second. In a web graph this is a reasonable result; however, in a citation graph node 3 should be the first in the rank table.

Table 5: Rank results for the graph of Figure 3(b).

CC	PR	HA	BHA	SA	BSA	P	PS	BPS	EPS	BEPS	SCEAS
6 2.000	5 0.820	5 0.851	5 0.909	6 0.626	5 0.935	0 0.000	5 2.287	5 8.000	5 0.769	5 0.767	5 0.767
5 2.000	3 0.689	4 0.526	4 0.416	5 0.417	4 0.355	1 0.000	4 1.143	3 6.000	6 0.555	6 0.736	6 0.736
4 1.000	2 0.635	6 0.000	6 0.000	0 0.313	3 0.000	2 0.000	3 1.140	2 5.000	0 0.432	0 0.639	0 0.639
3 1.000	1 0.570	0 0.000	3 0.000	1 0.313	2 0.000	3 0.000	2 1.134	1 4.000	1 0.397	1 0.603	1 0.603
2 1.000	0 0.494	1 0.000	0 0.000	2 0.313	1 0.000	4 0.000	1 1.124	4 3.500	2 0.388	2 0.590	2 0.590
1 1.000	4 0.443	2 0.000	1 0.000	3 0.313	0 0.000	5 0.000	0 1.104	0 3.000	3 0.385	3 0.585	3 0.585
0 1.000	6 0.405	3 0.000	2 0.000	4 0.209	6 0.000	6 0.000	6 1.068	6 2.000	4 0.385	4 0.292	4 0.292
8 0.000	7 0.150	7 0.000	7 0.000	7 0.000	7 0.000	7 0.000	7 0.000	7 0.000	7 0.000	7 0.000	7 0.000
7 0.000	8 0.150	8 0.000	8 0.000	8 0.000	8 0.000	8 0.000	8 0.000	8 0.000	8 0.000	8 0.000	8 0.000

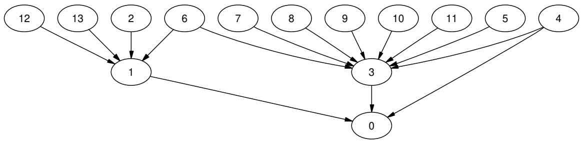


Figure 4: Final example of a graph.

2.4 HITS

`HITS` has been proposed to rank web pages retrieved when searching through a browser. The notion behind `HITS` is the discrimination between hubs and authorities. Hubs are pages with good links, whereas authorities are pages with good content. Any node can be a hub or authority. Thus, `HITS` computes 2 vectors of scores. Originally the scores for hubs and authorities have been defined as [10]:

$$\begin{aligned} \vec{a'} &= A^T * \vec{h} \\ \vec{h'} &= A * \vec{a} \end{aligned}$$

where A is the adjacency matrix of the citation graph, with $A_{i,j} = 1$ if publication i cites publication j, and zero otherwise. $\vec{a}$ is a vector where its i-th element stands for the authority score of publication i, whereas vector $\vec{h}$ contains the scores of hub nodes. By using the terminology of Table 1, `HITS Authority` (`HA`) and `HITS Hub` (`HH`) scores can be computed as:

$$\begin{aligned} HA_x &= \sum_{\forall y \in I_x} HH_y \\ HH_x &= \sum_{\forall y \in O_x} HA_y \end{aligned} \tag{4}$$

Table 6: Rank results for the graph of figure 4.

CC	PR	HA	BHA	SA	BSA	P	PS	BPS	EPS	BEPS	SCEAS
3 8.000	0 1.607	3 0.960	3 0.886	3 0.848	0 0.797	0 0.000	0 7.130	0 13.000	3 2.378	3 2.575	3 2.575
1 4.000	3 1.043	1 0.218	0 0.424	1 0.424	3 0.587	1 0.000	3 5.247	3 7.000	0 1.952	0 2.341	0 2.341
0 3.000	1 0.596	0 0.177	1 0.189	0 0.318	1 0.143	2 0.000	1 2.623	1 3.500	1 1.189	1 1.288	1 1.288
9 0.000	2 0.150	2 0.000	2 0.000	2 0.000	2 0.000	3 0.000	2 0.000	2 0.000	2 0.000	2 0.000	2 0.000
8 0.000	4 0.150	4 0.000	4 0.000	4 0.000	4 0.000	4 0.000	4 0.000	4 0.000	4 0.000	4 0.000	4 0.000
7 0.000	5 0.150	5 0.000	5 0.000	5 0.000	5 0.000	5 0.000	5 0.000	5 0.000	5 0.000	5 0.000	5 0.000
6 0.000	6 0.150	6 0.000	6 0.000	6 0.000	6 0.000	6 0.000	6 0.000	6 0.000	6 0.000	6 0.000	6 0.000
5 0.000	7 0.150	7 0.000	7 0.000	7 0.000	7 0.000	7 0.000	7 0.000	7 0.000	7 0.000	7 0.000	7 0.000
4 0.000	8 0.150	8 0.000	8 0.000	8 0.000	8 0.000	8 0.000	8 0.000	8 0.000	8 0.000	8 0.000	8 0.000
2 0.000	9 0.150	9 0.000	9 0.000	9 0.000	9 0.000	9 0.000	9 0.000	9 0.000	9 0.000	9 0.000	9 0.000
13 0.000	10 0.150	10 0.000	10 0.000	10 0.000	10 0.000	10 0.000	10 0.000	10 0.000	10 0.000	10 0.000	10 0.000
12 0.000	11 0.150	11 0.000	11 0.000	11 0.000	11 0.000	11 0.000	11 0.000	11 0.000	11 0.000	11 0.000	11 0.000
11 0.000	12 0.150	12 0.000	12 0.000	12 0.000	12 0.000	12 0.000	12 0.000	12 0.000	12 0.000	12 0.000	12 0.000
10 0.000	13 0.150	13 0.000	13 0.000	13 0.000	13 0.000	13 0.000	13 0.000	13 0.000	13 0.000	13 0.000	13 0.000

Formally, HITS is computed over a graph subset, which is the result set of a user search. In the case of bibliometrics, we could use the entire citation graph as a set for applying HITS and rank all the publications. In practice, HITS ranking is not appropriate for the bibliometrics field. The reason is that publications get high authority score only if there are hubs pointing to it. However, this should not be the main concern in bibliometrics. In addition, as Borodin et al. proved in [1], a Hub is penalized when it points to 'poor' authorities. This is also shown in Figure 1, where the authority score of node 0 converges to 1, whereas all other authorities get a score of 0. The same behavior of HITS appears in Table 3 (Figure 2) where the score of node 1 also converges to 1.

2.5 Prestige

Specifically for bibliometrics, Kleinberg in [10] proposes a model where authorities directly endorse other authorities. This is what Chakrabarti calls Prestige [3]. The rank score of this method is:

$$\vec{a'} = A^T * \vec{a}$$

or according to our notation, the previous expression is equivalent to:

$$P_x = \sum_{\forall y \in I_x} P_y \tag{5}$$

From Equation 5 it is obvious that in the ideal case that there are no cycles in our citation graph:

- The values of $\vec{P}$ converge to $\vec{0}$.
- Even if we find a way that $\vec{P}$ does not converge to zero, a node x will never get a greater score than node y, if x points to y. The proof of this sentence is simple, since there are no negative values in the score vector. Thus, in the best case it will be $P_x = P_y$, but in general and most common case it will hold that $P_x \leq P_y$.

From all the examples presented so far, we see that Prestige gives a rank score only to nodes that belong to cycles or to nodes pointed by members of cycles. All the remaining nodes converge to a score of 0.

2.6 SALSA

SALSA, proposed by Lempel and Moran [13], is a variation of HITS as it uses a balance between the in- and out-degrees. The formulation of SALSA score by using our notation is:

$$\begin{aligned} SA_x &= \sum_{\forall y \in I_x} \frac{SH_y}{|O_y|} \\ SH_x &= \sum_{\forall y \in O_x} \frac{SA_y}{|I_y|} \end{aligned} \tag{6}$$

SALSA behaves much better than HITS in the cases of Figures 1, 2 and 4. On the other hand, as we can see in the example of Figure 3, node 6 comes first in the rank table after the addition of one citation.

3 Our Ranking Methods

Summarizing the weaknesses of the previous algorithms we observe that:

- CC and BCC do not take into account the importance of the citing publications
- Prestige does not have good behavior when citation circles exist; the score for non-members of circles converges to zero.
- PageRank has the same problem with Prestige. The members of circles get higher scores.
- HITS and SALSA are based on the notion of hubs and authorities, which is not appropriate for evaluating publications.

Having in mind the above remarks, in the sequel we will define some new ranking methods, appropriate for evaluation of publications.

3.1 B-HITS

To bypass the weakness of HITS, we propose the (Balanced HITS) B-HITS algorithm. According to this variant, a node is not only enforced by links from hubs, but also by links from other authorities. This way actually we change the original notion of '*authority*' to '*noticeable elements*' or '*valuable elements*' and, thus, the formula for computing the scores becomes:

$$\begin{aligned} \vec{a'} &= (1-p) * A^T\vec{h} + p * A^T\vec{a} \\ \vec{h'} &= A * \vec{a} \end{aligned}$$

where p is the percentage of authorities endorsement to other authorities. By using our terminology, the above equations are equivalent to:

$$\begin{aligned} BHA_x &= (1-p) * \sum_{\forall y \in I_x} BHH_y + p * \sum_{\forall y \in I_x} BHA_y \\ BHH_x &= \sum_{\forall y \in O_x} BHA_y \end{aligned} \qquad (7)$$

At this stage, for brevity we avoid finding an optimized value for p, although further investigation is needed (probably based on the graph characteristics). Assuming a fair balancing, in our experiments we have used $p = 0.5$.

3.2 B-SALSA

The endorsement balancing of the authorities and hubs to authorities in B-HITS can be used in SALSA as well. The resulting equation is:

$$\begin{aligned} BSA_x &= (1-p) * \sum_{\forall y \in I_x} \frac{BSH_y}{|O_y|} + p * \sum_{\forall y \in I_x} \frac{BSA_y}{|O_y|} \\ BSH_x &= \sum_{\forall y \in O_x} \frac{BSA_y}{|I_y|} \end{aligned} \tag{8}$$

Actually, this will cause SALSA to practically resemble to PageRank as we will see in the following section.

3.3 SCEAS PS: Publication Score

In our SCEAS system [15], the Publication Score (PS) of a node x equals the sum of the $PS + b$ of all nodes pointing directly to x. Thus, each citation to x from a publication y gives a constant factor b plus the score PS_y to depict its importance. Thus the resulting equation is:

$$PS_x = \sum_{\forall y \in I_x} (b + PS_y) \tag{9}$$

This approach called PS, is a hybrid between CC and Prestige, where CC is multiplied by the factor b (i.e. $\vec{PS} = b * \vec{CC} + \vec{P}$). Prestige has the disadvantage that it converges to zero in the absence of cycles. Using PS we overcome the problem of Prestige.

If there are no circles in the graph, then PS could be computed recursively. In such a case, if node x points to y, then node x will never get a score higher than y. For example, for $b = 1$, in Figure 2 the resulting vector should be $\vec{PS} = (7, 6, 0, 0, 0, 0, 0, 0)$ since $PS_1 = 6 * b = 6$ and $PS_0 = PS_1 + b = 7$.

Usually, circles do exist. In such a case, PS, like HITS and Prestige, needs a normalization step to converge. Without normalization all these algorithms should lead the score vector to converge to ∞. The normalization keeps the 'energy' stable during the iterations. Usually, the $||.||_1$ is used for normalization and in most cases we normalize so that $||.||_1 = 1$.

For the PS case, factor b complicates the task of normalization. If we normalize so that $||\vec{PS}||_1 = 1$, then for large graphs we will have $\vec{PS} \approx \vec{CC}$. The starting vector for $\vec{PS}$ is $\vec{0}$, which means that after the first iteration $||\vec{PS}||_1$ should be equal to $|E| * b$. Thus, we can keep $||\vec{PS}||_1$ stable by normalizing it with:

$$||\vec{PS}||_1 = |E| * b$$

In the example of Figure 2, the resulting vector is $\vec{PS} = (3.13, 3.86, 0, 0, 0, 0, 0, 0)$, which means that finally $PS_1 > PS_0$. Thus, with normalization we overcome the fact that node x could never get a higher score than node y, if x points to y.

3.4 SCEAS BPS: Balanced Publication Score

We may elaborate Equation 9 by adopting a balancing factor as a function of the number of outgoing citations. This results in a Balanced Publication Score (BPS) as follows:

$$BPS_x = \sum_{\forall y \in I_x} \frac{BPS_y + b}{|O_y|} \tag{10}$$

BPS embeds the reasoning of PageRank and SALSA into PS. From the tables of the given examples, we remark that BPS concludes to the same ranking as PageRank. This will also be shown in the following section.

3.5 SCEAS EPS: Exponentially Weighted Publication Score

Another elaborated version of PS (Equation 9) is the Exponentially Weighted Publication Score (EPS). The EPS score of node x is the exponentially weighted sum of the EPSs of all nodes pointing to x directly:

$$EPS_x = \sum_{\forall y \in I_x} (EPS_y + b) * a^{-1} \tag{11}$$

This metric actually takes into account the size of the tree formed by the citations pointing directly or indirectly to x. If there is an indirect link from node x to node y, then the score of y is a function of the score of x multiplied by a^{-d}, where d is the distance between the two nodes.

Here, like in the PS case, normalization is not necessary if cycles do not exist. For $a = e$, the computation of ranking for Figure 2 (without normalization) would result in $E\vec{P}S = (1.179, 2.207, 0, 0, 0, 0, 0, 0)$, since $EPS_1 = 6*b*a^{-1} = 6*e^{-1} = 2.207$ and $EPS_0 = (EPS_1+b)*a^{-1} = (2.207+1)*e^{-1} = 1.179$.

In general, we need a normalization step when cycles exist. For the same reasons as in the case of PS, the $||E\vec{P}S||_1$ should not be normalized to 1. In addition, we should not use the same normalization value as in PS because this could lead EPS to be identical to PS. For example, assume that we derive a normalization factor n such that $||n * \vec{PS}||_1 = |E| * b$. If, during the same iteration, we were seeking for a normalization factor k for EPS such that $||k * E\vec{P}S||_1 = |E| * b$, then $k = a * n$ should hold. This would lead the two algorithms to be identical since it would produce $E\vec{P}S = \vec{PS} * (k/n) * a^{-1} = \vec{PS} * (a * n/n) * a^{-1} = \vec{PS}$, which means that the ranking would be exactly the same. To avoid this, we should include factor a in the normalization process, so that: $||E\vec{P}S||_1 = |E| * b * a^{-1}$. Thus, the normalization factor is equally important in the EPS case.

3.6 SCEAS BEPS: Balanced Exponentially Weighted Publication Score

A hybrid method based on Equations 10 and 11 is the Balanced Exponentially Weighted Publication Score (BEPS). The $BEPS$ score of node x is the exponentially weighted sum of the $BEPS_y$ divided by the number of citations made from publication y, $\forall y \in I_x$:

$$BEPS_x \;=\; \sum_{\forall y \in I_x} \frac{BEPS_y + b}{|O_y|} * a^{-1} \tag{12}$$

3.7 SCEAS General

A dumping factor d may be added into Equation 12 which would lead to the following equation, as a generalized formula of `BEPS` and `PageRank`:

$$S_x \;=\; (1-d) + d * \sum_{\forall y \in I_x} \frac{S_y + b}{|O_y|} * a^{-1} \tag{13}$$

From now on we will refer to this method with the name `SCEASRank`. For $d = 1$, `SCEASRank` is equivalent to `BEPS`. For $b = 0$ and $a = 1$ `SCEAS` is equivalent to `PageRank`. `PageRank` uses a value of $d = 0.85$ to balance the precision and the convergence speed. A value of d closer to 1 results in better precision for the scores. Also, the value $d = 1$ should lead `PageRank` to converge to zero. Therefore, a value $d < 1$ is necessary for `PageRank`. For `SCEAS` it is safe to use any value for d (where $0 < d \leq 1$) if $b > 0$. Also, the convergence speed is mainly affected by the factor a rather than d. In our case, it is safe to use any greater factor d than 0.85, such as 0.99.

4 Experiments

4.1 The Dataset

Our SCEAS system uses the DBLP data having the time-stamp of 2005-05-19. Table 7 describes in detail the qualitative characteristics of the DBLP citation graph. We observe that only 1.31% of the publications have their citations stored (V_O), whereas only 2.92% of them have in-degree (V_I). Therefore, these publications are actually being ranked. The distribution of these citations per year and source is described in [16]. Here, we notice that the available citations (e.g. the V_O set) relate to publications that are published in a selected set of the most important proceedings and journals according to the DBLP administrators. In other words, the set of citing papers has been already filtered with the criterion of conference/journal importance. Practically, in our dataset all citing papers have a minimum quality standard, and, therefore, it is expected that the differences in the rank orders produced in the course of our experiments by the various methods will be rather smooth.

4.2 Computation Speed

According to the definition of `SCEASRank`, it is obvious that we come up with a very fast convergence for $a = e$. In Figures 5-6, the x-axis represents the number of iterations needed by each algorithm

Table 7: Database and citation graphs properties.

Symbol	Definition	Properties
$G_\infty = (V_\infty, E_\infty)$	The universal citation graph	
$G_{DBLP} = (V_{DBLP}, E_{DBLP})$	Our DBLP mirror database	$G_{DBLP} \subset G_\infty$
V_{DBLP}	Nodes in the citation graph	$\lvert V_{DBLP} \rvert = 624893$
V_{INP}	The set of in proceedings	$\lvert V_{INP} \rvert = 391543$
V_{ART}	The set of articles in journals	$\lvert V_{ART} \rvert = 233350$
E_{DBLP}	Citations in the graph	$\lvert E_{DBLP} \rvert = 100210$
$E_{DBLP-} = \{(i \to j) : i \in V_{DBLP} \wedge j \notin V_{DBLP}\}$	Citations to non DBLP entities	$\lvert E_{DBLP-} \rvert = 67971$
$V_I = \{x \in V_{DBLP} : \lvert I_x \rvert > 0\}$		$\lvert V_I \rvert = 18273$
$V_O = \{x \in V_{DBLP} : \lvert O_x \rvert > 0\}$		$\lvert V_O \rvert = 8183$
		$\lvert V_O \cup V_I \rvert = 20831$

to compute the ranks for the DBLP dataset. Axis y shows in a logarithmic scale the value of:

$$\delta \;=\; ||\vec{x}_l - \vec{x}_{l-1}||_1 \tag{14}$$

where $\vec{x}_l$ is the vector with scores $\{S_1, S_2, \ldots, S_V\}$ after l iterations. In the cases of `HITS` and `SALSA`, δ is computed as the sum of the Norm-1 of the authorities vector plus the Norm-1 of the hubs vector:

$$\delta \;=\; ||\vec{a}_l - \vec{a}_{l-1}||_1 + ||\vec{h}_l - \vec{h}_{l-1}||_1 \tag{15}$$

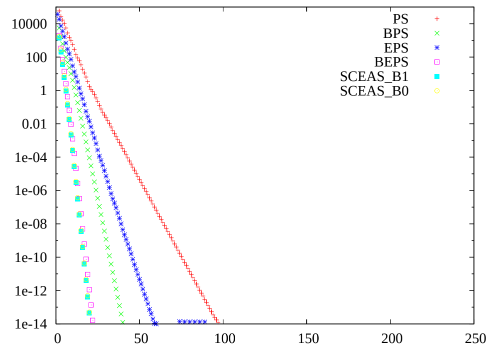


Figure 5: Computation Speed of `SCEAS` variations on DBLP collection.

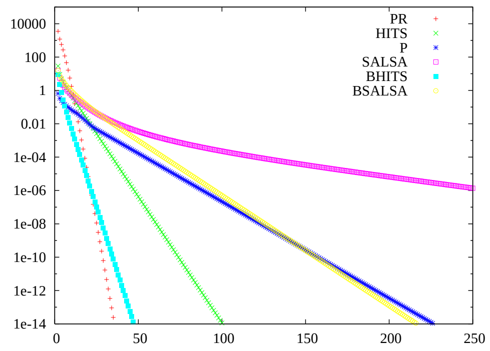


Figure 6: Computation Speed of `PageRank`, `HITS`, `SALSA`, `Prestige` on DBLP collection.

Table 8: The tangent of the lines of Figures 5 and 6.

Algorithm	`SCEAS_B0`	`SCEAS_B1`	`BEPS`	`PR`	`BPS`	`EPS`	`BHITS`	`PS`	`HITS`	`BSALSA`	`P`	`SALSA`
Angle	1.029	1.030	1.106	1.804	2.057	3.176	3.193	5.428	6.605	15.184	17.270	71.621

The termination condition for each algorithm is $\delta < \epsilon$, where ϵ is a very small number. Actually, as described in [8], this number could not be predefined for `PageRank`, since it depends on the citation graph. It is obvious that each algorithm needs a different value of ϵ as a termination condition. Regardless of this, the plot shows that the SCEAS curves are much steeper than the curves of the other algorithms. This means that SCEAS converges faster than the other methods no matter what the actual values of δ and ϵ are. For better understanding, Table 8 shows the tangent of the angle $((x2 - x1)/(y1 - y2))$ of each curve. From the latter table it is clear that SCEAS is the fastest algorithm, BEPS follows with small difference, PageRank and BPS come after at almost half the speed of convergence compared to SCEAS, whereas the remaining algorithms have a very slow convergence speed.

At this point, it is worth noticing that for all algorithms except HITS, SALSA and their variations, the computation could be done in only one step iff:

1. The graph had no cycles, and
2. The computation was made recursively by starting from the dangling nodes.

However, the graph contains a few cycles, and, thus, one step computing is infeasible.

4.3 Results Comparison

In this section we compare the results of all the algorithms mentioned above. This task is not trivial, since they are widely accepted in their own application field. In particular, we perform a many-fold statistical comparison based on the results for DBLP collection:

- By counting the number of common elements out of the top x rank table elements for each pair of algorithms.
- By plotting the function $Top(x)$, where x is the number of top elements in the rank tables.
- By computing the distance between all pairs of rank tables based on the Kendall's tau [9] that is also used in [1].
- By computing the simple distance between all pairs of rank tables.
- By computing the weighted distance between all pairs of rank tables.
- By using q-q plots.

4.3.1 Comparison based on the number of common elements

Consider that all ranking methods are executed on the DBLP dataset and for each method the top 20 elements are extracted. Table 9 depicts the number of common elements among the top 20 elements for each pair of ranking methods. As shown in this table, SCEAS_B1 (for $b = 1, d = 0.85, a = e$), SCEAS_B0 (for $b = 0, d = 0.85, a = e$) and BEPS (for $b = 1, d = 1, a = e$) are very close to each other. PageRank is very close to BPS, since BPS is equivalent to PageRank without the damping factor. `SALSA` authority gives the top 20 elements based on the citation count (`CC`). The variants `PS` and `EPS` are also very close to each other. In Table 9, as well as in the following Tables which present distances of the ranking algorithms, the cells that are framed denote that the two algorithms compared are dissimilar (few common elements or great distance). The cells that are marked with a gray background denote that the two algorithms are similar. The light gray denotes high similarity while the dark gray simply similarity.

Table 9: Common elements between the top 20 DBLP publications for each pair of ranking methods.

	CC	BCC	PR	HA	P	SA	BHA	BSA	PS	BPS	EPS	BEPS	SCEAS_B1	SCEAS_B0
CC	-	18	12	13	5	20	7	13	10	11	11	15	16	16
BCC	18	-	12	11	5	18	7	13	10	11	11	14	15	15
PR	12	12	-	11	12	12	8	16	17	19	18	17	16	16
HA	13	11	11	-	6	13	10	12	10	11	11	12	11	11
P	5	5	12	6	-	5	8	10	13	13	12	9	9	9
SA	20	18	12	13	5	-	7	13	10	11	11	15	16	16
BHA	7	7	8	10	8	7	-	10	8	8	9	8	7	7
BSA	13	13	16	12	10	13	10	-	14	16	15	17	16	16
PS	10	10	17	10	13	10	8	14	-	18	19	14	13	13
BPS	11	11	19	11	13	11	8	16	18	-	18	16	15	15
EPS	11	11	18	11	12	11	9	15	19	18	-	15	14	14
BEPS	15	14	17	12	9	15	8	17	14	16	15	-	19	19
SCEAS_B1	16	15	16	11	9	16	7	16	13	15	14	19	-	20
SCEAS_B0	16	15	16	11	9	16	7	16	13	15	14	19	20	-

4.3.2 Comparison based on the $Top(x)$ Function

One could argue that counting the number of common elements among the top 20 elements of two rank lists is not indicative of the similarity of these lists. For this reason, we define the function $Top(a_1, a_2, x)$ such that for two ranking algorithms a_1 and a_2, $Top(a_1, a_2, x)$ gives the number of common elements in the top x rank lists divided by the number of nodes:

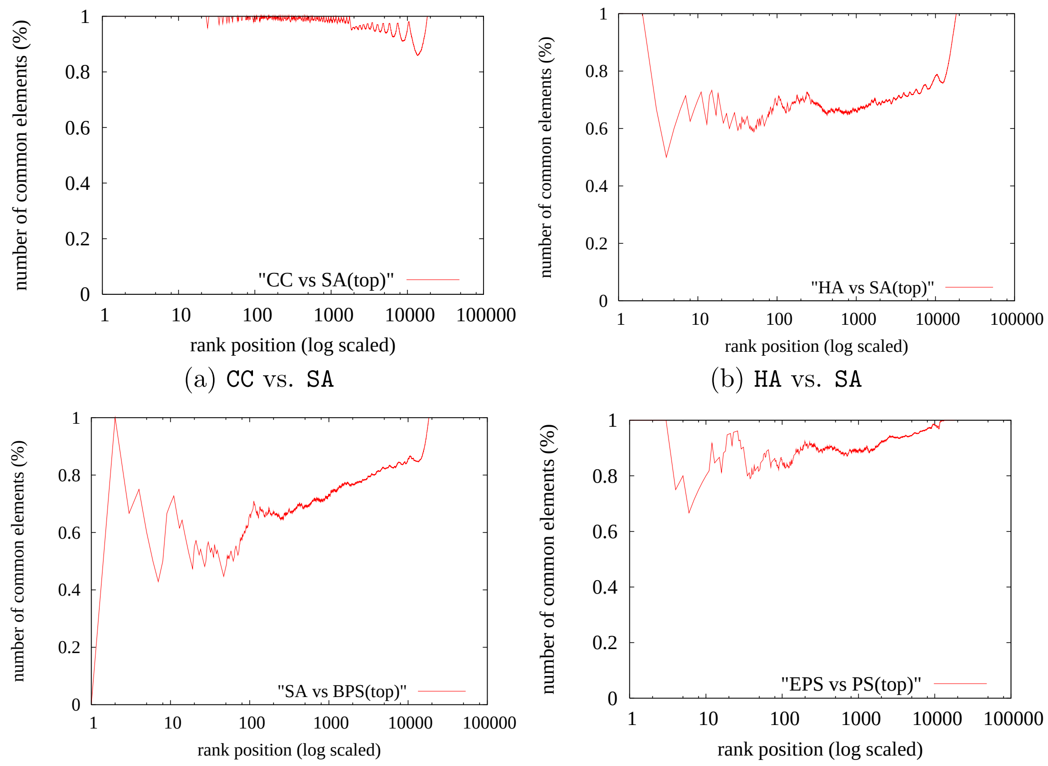
$$R(a, x) = \{i \in V : P_a(i) \leq x\}$$

$$Top(a_1, a_2, x) = |R(a_1, x) \cap R(a_2, x)|/|V| \quad (16)$$

where $P_a(i)$ is the position of node i in the rank table produced by algorithm a and $R(a, x)$ is the set of top x elements of the rank table. It is obvious that $Top(a_1, a_2, x) \underset{x \to |V|}{\longrightarrow} 1$. Some plots of the above function are shown in Figure 7. For example, Figure 7a confirms that `SALSA` is almost equal to `CC` as also shown in Table 9. Also `EPS` is very close to `PS` (see Figure 7d). On the other hand, `SALSA` is far from `BPS` (see Figure 7c) and finally `HITS Authorities` is stably on 50% related to `SALSA Authorities`.

4.3.3 Comparison based on Kendall's tau

The distance between any pair of rank tables can be computed by Kendall's tau with penalty p [9, 1]. To perform this computation, the *violating set* $\mathcal{V}(a_1, a_2)$ and the *weakly violating set* $\mathcal{W}(a_1, a_2)$ must



(a) CC vs. SA (b) HA vs. SA

(c) SA vs. BPS (d) EPS vs. PS

Figure 7: Comparison over DBLP rank results with plot of function $C(a_1, a_2, x)$.

be defined:

$$\mathcal{V}(a_1, a_2) = \{(i, j) \in V : (a_1(i) < a_2(j) \wedge a_2(i) > a_1(j)) \vee (a_1(i) > a_2(j) \wedge a_2(i) < a_1(j))\}$$
$$\mathcal{W}(a_1, a_2) = \{(i, j) \in V : (a_1(i) = a_1(j) \wedge a_2(i) \neq a_2(j)) \vee (a_1(i) \neq a_1(j) \wedge a_2(i) = a_2(j))\}$$

where $a_1(i)$ is the computed score for element i by the rank method a_1. The violating set is the set of all node pairs for which their ranking is different for the two rank tables (a_1 and a_2). The weakly violating set is the set of all node pairs which are ranked at the same position by one algorithm, but in different position by another. Kendall's tau penalizes by 1 for each member of the violating set, and by p each member of the weakly violating set:

$$K^{(p)}(a_1, a_2) = \sum_{\forall i, j \in V} \mathcal{I}^{(p)}_{a_1 a_2}(i, j)$$

$$\mathcal{I}^{(p)}_{a_1 a_2}(i, j) = \begin{cases} 1 & \text{if } (i, j) \in \mathcal{V}(a_1, a_2) \\ p & \text{if } (i, j) \in \mathcal{W}(a_1, a_2) \\ 0 & \text{otherwise} \end{cases}$$

Actually, Kendall's tau with penalty $p = 0$ gives the number of swaps that will be performed by bubble sort to produce the second rank table starting from the first. Finally, the weak rank distance, as defined in [1] is:

$$d^{(0)}(a_1, a_2) \;=\; \frac{1}{|V|(|V|-1)/2} K^{(0)}(a_1, a_2) \tag{17}$$

whereas the results are shown in Table 10. The strict rank distance (see Table 11) is:

$$d^{(1)}(a_1, a_2) \;=\; \frac{1}{|V|(|V|-1)/2} K^{(1)}(a_1, a_2) \tag{18}$$

where $|V|$ is the number of graph vertexes. Thus, $d^{(0)}(a_1, a_2)$ and $d^{(1)}(a_1, a_2)$ are normalized in the scale of [0..1].

Table 10: Differences in ranking of DBLP citation graph computed by $d^{(0)}(a_1, a_2)$.

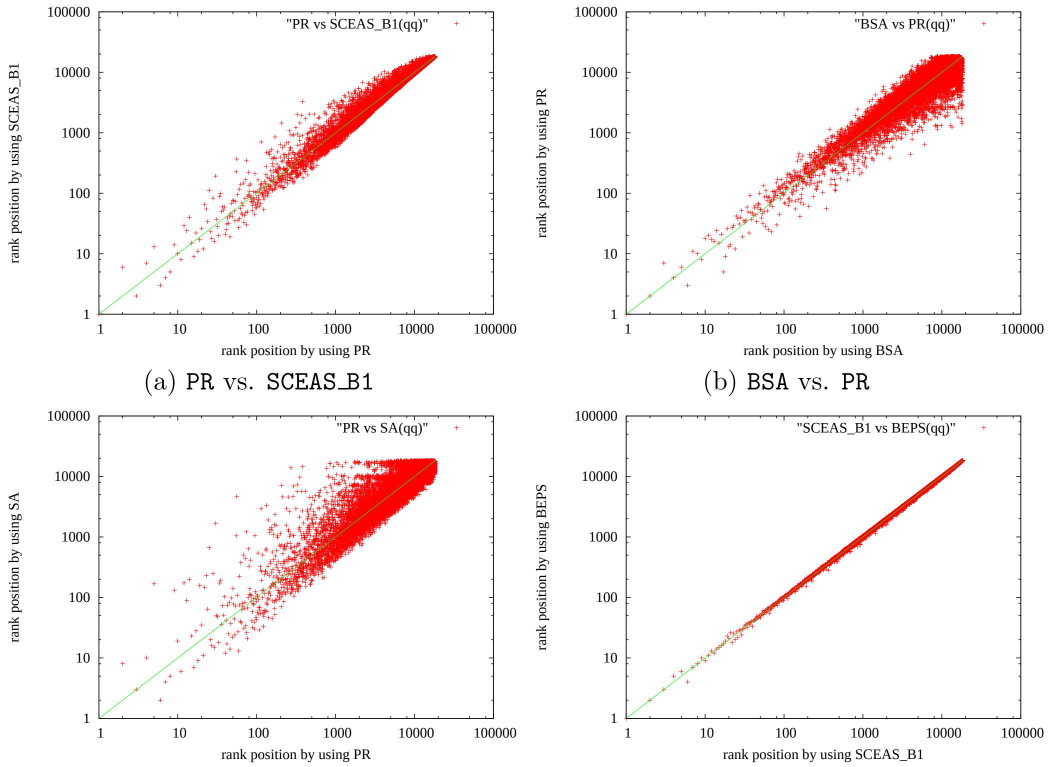
	CC	BCC	PR	HA	P	SA	BHA	BSA	PS	BPS	EPS	BEPS	SCEAS_B1	SCEAS_B0
CC	-	7.6%	7.6%	12.8%	15.0%	0.8%	15.5%	5.0%	7.6%	7.8%	5.8%	7.3%	7.3%	7.3%
BCC	7.6%	-	6.7%	34.3%	25.2%	11.1%	34.5%	23.0%	24.1%	7.8%	22.7%	3.0%	2.6%	2.6%
PR	7.6%	6.7%	-	31.5%	20.2%	11.7%	30.6%	19.1%	18.5%	1.0%	17.3%	3.6%	4.1%	4.1%
HA	12.8%	34.3%	31.5%	-	15.9%	21.3%	12.1%	16.0%	17.6%	31.1%	17.3%	33.1%	33.3%	33.3%
P	15.0%	25.2%	20.2%	15.9%	-	18.9%	12.1%	13.4%	9.2%	19.4%	10.9%	22.9%	23.2%	23.2%
SA	0.8%	11.1%	11.7%	21.3%	18.9%	-	24.0%	13.8%	13.5%	12.0%	11.8%	11.1%	11.1%	11.1%
BHA	15.5%	34.5%	30.6%	12.1%	12.1%	24.0%	-	15.3%	16.0%	30.0%	16.6%	32.8%	33.0%	33.0%
BSA	5.0%	23.0%	19.1%	16.0%	13.4%	13.8%	15.3%	-	10.2%	18.7%	9.2%	21.0%	21.3%	21.3%
PS	7.6%	24.1%	18.5%	17.6%	9.2%	13.5%	16.0%	10.2%	-	17.6%	1.9%	21.5%	21.9%	21.9%
BPS	7.8%	7.8%	1.0%	31.1%	19.4%	12.0%	30.0%	18.7%	17.6%	-	16.6%	4.7%	5.2%	5.2%
EPS	5.8%	22.7%	17.3%	17.3%	10.9%	11.8%	16.6%	9.2%	1.9%	16.6%	-	20.2%	20.6%	20.6%
BEPS	7.3%	3.0%	3.6%	33.1%	22.9%	11.1%	32.8%	21.0%	21.5%	4.7%	20.2%	-	0.4%	0.4%
SCEAS_B1	7.3%	2.6%	4.1%	33.3%	23.2%	11.1%	33.0%	21.3%	21.9%	5.2%	20.6%	0.4%	-	0.0%
SCEAS_B0	7.3%	2.6%	4.1%	33.3%	23.2%	11.1%	33.0%	21.3%	21.9%	5.2%	20.6%	0.4%	0.0%	-

Table 11: Differences in ranking of DBLP citation graph computed by $d^{(1)}(a_1, a_2)$.

	CC	BCC	PR	HA	P	SA	BHA	BSA	PS	BPS	EPS	BEPS	SCEAS_B1	SCEAS_B0
CC	-	30.2%	30.5%	35.8%	37.9%	11.5%	38.6%	28.1%	25.6%	30.7%	23.9%	30.2%	30.2%	30.2%
BCC	30.2%	-	7.1%	35.0%	46.4%	23.6%	35.2%	23.6%	29.4%	8.2%	28.0%	3.5%	3.0%	3.0%
PR	30.5%	7.1%	-	31.7%	41.3%	24.2%	30.8%	19.3%	23.4%	1.0%	22.2%	3.6%	4.1%	4.1%
HA	35.8%	35.0%	31.7%	-	37.2%	33.7%	12.1%	16.1%	22.6%	31.3%	22.3%	33.3%	33.5%	33.5%
P	37.9%	46.4%	41.3%	37.2%	-	43.0%	33.4%	34.7%	25.5%	40.5%	27.1%	44.0%	44.3%	44.3%
SA	11.5%	23.6%	24.2%	33.7%	43.0%	-	36.4%	26.2%	26.8%	24.5%	25.1%	23.6%	23.5%	23.5%
BHA	38.6%	35.2%	30.8%	12.1%	33.4%	36.4%	-	15.3%	21.1%	30.1%	21.7%	33.0%	33.2%	33.2%
BSA	28.1%	23.6%	19.3%	16.1%	34.7%	26.2%	15.3%	-	15.2%	18.8%	14.3%	21.2%	21.4%	21.4%
PS	25.6%	29.4%	23.4%	22.6%	25.5%	26.8%	21.1%	15.2%	-	22.5%	1.9%	26.4%	26.8%	26.8%
BPS	30.7%	8.2%	1.0%	31.3%	40.5%	24.5%	30.1%	18.8%	22.5%	-	21.4%	4.7%	5.2%	5.2%
EPS	23.9%	28.0%	22.2%	22.3%	27.1%	25.1%	21.7%	14.3%	1.9%	21.4%	-	25.1%	25.5%	25.5%
BEPS	30.2%	3.5%	3.6%	33.3%	44.0%	23.6%	33.0%	21.2%	26.4%	4.7%	25.1%	-	0.4%	0.4%
SCEAS_B1	30.2%	3.0%	4.1%	33.5%	44.3%	23.5%	33.2%	21.4%	26.8%	5.2%	25.5%	0.4%	-	0.0%
SCEAS_B0	30.2%	3.0%	4.1%	33.5%	44.3%	23.5%	33.2%	21.4%	26.8%	5.2%	25.5%	0.4%	0.0%	-

4.3.4 Comparison based on q-q plots

Another tool to compare the rankings is by means of q-q plots. Each point $x(i,j)$ of a q-q plot corresponds to some node x. The coordinates of the point (i,j) mean that the node is ranked at the i-th position by the first algorithm (x-axis) and at the j-th position by the second algorithm (y-axis). In Figure 8a `PageRank` is compared against `SCEAS`. We see that all the points are quite close to the line $y = x$. Figures 8b-8c show the comparison of `PageRank` to `SALSA` and `B-SALSA`. It is clear that `B-SALSA` is closer to `PageRank` than `SALSA`. Finally, Figure 8d shows the similarity of `SCEAS` (with $b = 1$) to `BEPS`. We can see that all points are pretty much distributed on the line $y = x$.



(a) `PR` vs. `SCEAS_B1` (b) `BSA` vs. `PR`

(c) `PR` vs. `SA` (d) `SCEAS_B1` vs. `BEPS`

Figure 8: Comparison over DBLP rank results with q-q plots.

4.3.5 Comparison based on Simple Distance

For brevity we avoid presenting further comparisons with q-q plots. To summarize a q-q plot into one number, we can compute for all points (i,j) their distances from the line $y = x$ and then sum them up. The distance of point (i,j) from the line $y = x$ is equal to $|j - i|/\sqrt{2}$. Thus, an alternative

Simple Distance measure of algorithms a_1 and a_2 is:

$$D(a_1, a_2) \;=\; \frac{1}{|V|} \sum_{\forall i} \frac{|P_{a_1}(i) - P_{a_2}(i)|}{|V|} \tag{19}$$

where $|V|$ is the number of nodes and $P_{a_1}(i)$ is the position of node i in the rank table of algorithm a_1. The results are shown in Table 12. Function $D(a_1, a_2)$ is also normalized in the scale of [0..0.5], since 0.5 is the value of distance D for a reverse ordering. This comparison method is known as the Spearman's footrule [4].

Table 12: Differences in ranking of DBLP citation graph by $D(a_1, a_2)$.

	CC	BCC	PR	HA	P	SA	BHA	BSA	PS	BPS	EPS	BEPS	SCEAS_B1	SCEAS_B0
CC	-	13.3%	13.3%	17.1%	23.0%	7.9%	18.8%	11.5%	13.9%	13.4%	12.6%	13.1%	13.1%	13.1%
BCC	13.3%	-	5.3%	23.0%	24.3%	12.3%	23.3%	15.7%	17.8%	6.1%	16.9%	2.5%	2.2%	2.2%
PR	13.3%	5.3%	-	21.2%	21.2%	12.8%	20.7%	13.1%	14.1%	0.9%	13.3%	2.8%	3.2%	3.2%
HA	17.1%	23.0%	21.2%	-	18.3%	18.4%	8.6%	11.2%	14.0%	20.9%	13.8%	22.1%	22.2%	22.2%
P	23.0%	24.3%	21.2%	18.3%	-	22.7%	15.9%	17.0%	13.1%	20.7%	14.3%	22.9%	23.1%	23.1%
SA	7.9%	12.3%	12.8%	18.4%	22.7%	-	20.5%	13.2%	14.7%	13.1%	13.3%	12.3%	12.3%	12.3%
BHA	18.8%	23.3%	20.7%	8.6%	15.9%	20.5%	-	10.9%	13.1%	20.3%	13.6%	22.1%	22.2%	22.2%
BSA	11.5%	15.7%	13.1%	11.2%	17.0%	13.2%	10.9%	-	8.9%	12.8%	8.2%	14.3%	14.5%	14.5%
PS	13.9%	17.8%	14.1%	14.0%	13.1%	14.7%	13.1%	8.9%	-	13.6%	1.5%	16.1%	16.3%	16.3%
BPS	13.4%	6.1%	0.9%	20.9%	20.7%	13.1%	20.3%	12.8%	13.6%	-	12.8%	3.6%	4.0%	4.0%
EPS	12.6%	16.9%	13.3%	13.8%	14.3%	13.3%	13.6%	8.2%	1.5%	12.8%	-	15.2%	15.4%	15.4%
BEPS	13.1%	2.5%	2.8%	22.1%	22.9%	12.3%	22.1%	14.3%	16.1%	3.6%	15.2%	-	0.4%	0.4%
SCEAS_B1	13.1%	2.2%	3.2%	22.2%	23.1%	12.3%	22.2%	14.5%	16.3%	4.0%	15.4%	0.4%	-	0.0%
SCEAS_B0	13.1%	2.2%	3.2%	22.2%	23.1%	12.3%	22.2%	14.5%	16.3%	4.0%	15.4%	0.4%	0.0%	-

4.3.6 Comparison based on Weighted Distance

During comparing rankings any difference in a high position is practically more important than a difference in a low position. Neither Kendall's tau nor Spearman's footrule do make such discrimination. For this reason, we define an alternative rank distance measure. We set the *Weighted Distance* D_w of two rank methods a_1 and a_2 as:

$$\begin{aligned} w(a_1, a_2, i) &= \tfrac{1}{min(P_{a_1}(i), P_{a_2}(i))} \\ d_w(a_1, a_2, i) &= |P_{a_1}(i) - P_{a_2}(i)| * w(a_1, a_2, i) \\ D_w(a_1, a_2) &= \tfrac{1}{|V| * \sum_{\forall i \in V} w(a_1, a_2, i)} \sum_{\forall i \in V} d_w(a_1, a_2, i) \end{aligned} \tag{20}$$

The weight function $w(a_1, a_2, i)$ is linear and inversely proportional to the minimum rank position of node i. This means that if an element is ranked 1st by a_1 and 2nd by a_2, then the weight will be 1 and the Weighted Distance $d_w = 1$. More sample cases are shown in Table 13.

In Table 14 we can see which algorithms are close to each other. `PageRank` is very close to `SCEAS`, `BEPS`, `BPS` and finally `BCC`. `PageRank` computes a balanced score based on out-degrees, thus it is rather closer to `BCC` than to `CC`. On the other hand, `HITS Authorities (HA)` is rather closer to `CC`

Table 13: Example of weighted distance measure cases.

$P_{a_1}(i)$	$P_{a_2}(i)$	$w(a_1, a_2, i)$	$d_w(a_1, a_2, i)$
1	2	1	1
10	11	0.1	0.1
50	52	0.02	0.04
1	11	1	10
100	110	0.01	0.1
1010	1000	0.001	0.01

than to BCC. Prestige (P) is quite close to PS but far from the remaining algorithms. SALSA (SA) is quite close to CC and to other 'authorities only' algorithms (closer than HA). B-HITS (BHA) seems to deviate from HA. B-SALSA (BSA) lies at a distance with respect to all the remaining algorithms.

Table 14: Differences in ranking of DBLP citation graph by $D_w(a_1, a_2)$

	CC	BCC	PR	HA	P	SA	BHA	BSA	PS	BPS	EPS	BEPS	SCEAS_B1	SCEAS_B0
CC	-	3.6%	3.6%	5.1%	10.4%	1.6%	7.1%	2.7%	5.0%	3.8%	3.9%	3.3%	3.3%	3.3%
BCC	3.6%	-	1.8%	8.0%	12.0%	3.2%	9.3%	4.4%	6.8%	2.2%	5.8%	0.7%	0.6%	0.6%
PR	3.6%	1.8%	-	6.5%	8.6%	3.5%	6.9%	3.3%	4.4%	0.2%	3.9%	0.9%	1.0%	1.0%
HA	5.1%	8.0%	6.5%	-	6.6%	5.6%	4.3%	3.4%	4.2%	6.4%	3.8%	7.0%	7.1%	7.1%
P	10.4%	12.0%	8.6%	6.6%	-	10.8%	5.2%	6.2%	3.7%	8.1%	4.5%	10.4%	10.6%	10.6%
SA	1.6%	3.2%	3.5%	5.6%	10.8%	-	7.9%	3.3%	5.5%	3.7%	4.4%	3.0%	3.0%	3.0%
BHA	7.1%	9.3%	6.9%	4.3%	5.2%	7.9%	-	3.9%	3.9%	6.6%	4.0%	8.0%	8.2%	8.2%
BSA	2.7%	4.4%	3.3%	3.4%	6.2%	3.3%	3.9%	-	2.7%	3.2%	2.2%	3.9%	3.9%	3.9%
PS	5.0%	6.8%	4.4%	4.2%	3.7%	5.5%	3.9%	2.7%	-	4.1%	0.6%	5.5%	5.7%	5.7%
BPS	3.8%	2.2%	0.2%	6.4%	8.1%	3.7%	6.6%	3.2%	4.1%	-	3.6%	1.2%	1.3%	1.3%
EPS	3.9%	5.8%	3.9%	3.8%	4.5%	4.4%	4.0%	2.2%	0.6%	3.6%	-	4.8%	4.9%	4.9%
BEPS	3.3%	0.7%	0.9%	7.0%	10.4%	3.0%	8.0%	3.9%	5.5%	1.2%	4.8%	-	0.1%	0.1%
SCEAS_B1	3.3%	0.6%	1.0%	7.1%	10.6%	3.0%	8.2%	3.9%	5.7%	1.3%	4.9%	0.1%	-	0.0%
SCEAS_B0	3.3%	0.6%	1.0%	7.1%	10.6%	3.0%	8.2%	3.9%	5.7%	1.3%	4.9%	0.1%	0.0%	-

4.3.7 Discussion about Comparisons

Starting with the first comparison method in Table 9, we can detect 3 groups of algorithms that gave similar results:

1. BEPS, SCEAS_B1, SCEAS_B0
2. CC, BCC, SA
3. BPS, PR, EPS, PS

On the other hand, Prestige (P) seems to be the most dissimilar to most of the algorithms and mainly to CC, BCC, SA and HA. It is obvious that Table 9, does not give inside information since it is based on a very simple metric.

Referring to the *Top(x)* function we can have a better understanding of the (dis-)similarity of the algorithms. Unfortunately, it is very difficult to study 91 plots (which is the total number of combinations). From the plots of Figure 7, we see that in the first 100 elements, the percentage of common elements is almost stable and very close to the value depicted in Table 9.

The next metric, Kendall's tau, may give more information. In the weak distance table (see Table 10), we observe that the group of algorithms `BEPS - SCEAS_B0 - SCEAS_B1` is very strong and remains strong in the strict rank distance table, too. The next group we found with the *Top20* metric (`CC-BCC-SA`) seems to be violated in the weak distance table, as only `CC` and `SA` remain in the group. From Table 11 we remark that even `CC` and `SA` deviate further from each other, and, thus, the group disappears. In both Tables (10 and 11) it is shown that the last group of (`BPS-PR-EPS-PS`) is split into two groups: `BPS-PR` and `EPS-PS`. In addition, the group `BPS-PR` is very close to the group `BEPS-SCEAS_B1-SCEAS_B0`, whereas `BCC` is close to `BEPS-SCEAS_B1-SCEAS_B0` but not that close to `BPS-PR`.

Table 12 (Spearman's footrule) shows the same groups as tables 10-11. In addition, in this table it is more clear that `HA`, `P` and `BHA` are distant from almost all the remaining algorithms.

5 Evaluation of Results

This section is separated in two parts. First, we consider the papers of VLDB and SIGMOD conferences and we evaluate our ranking by comparing our results to the awarded papers by the *'VLDB 10 Year Award'* and the *'SIGMOD Test of Time Award'*. Secondly, we use the publications rank results to produce a rank table for authors. We evaluate our results by comparing them to the *'Edgar F. Codd Innovations Award'*.

5.1 Ranking Papers

It is a rather tough task to evaluate ranking algorithms for scientific publications, since it is subjective to decide which is better. As a criterion to verify that our ranking results are appropriate, we rely on two well known awards for publications: the *'VLDB 10 Year Award'* and the *'SIGMOD Test of Time Award'*. We accept that if an algorithm gives high rank positions to the awarded publications, then this algorithm can be used more safely for evaluating publications.

We compare all ranking methods: `PageRank`, `SCEAS` and variants, `HITS` and `B-HITS (Authorities)`, `SALSA` and `B-SALSA (Authorities)`, `Prestige`, `CC` and `BCC`. `SCEAS` with $b = 0$ is not presented here, since it produces similar results with `SCEAS` with $b = 1$ for the DBLP dataset. The testing scenario is the following:

1. Execute all ranking algorithms on the DBLP citation graph.
2. Extract the papers that have been published in the proceedings of VLDB and SIGMOD conferences (projection of V_{DBLP} to V_{vldb} and V_{sigmod}, respectively).

3. Organize and sort the rank tables grouped by conference and year. Some of these rank tables are presented in Appendix A.

4. Check the position of the awarded papers in the above rank tables.

Table 15: SIGMOD awared papers.

Year	Title	CC	BCC	PR	HA	P	SA	BHA	BSA	PS	BPS	EPS	BEPS	SCEAS_B1
1988	A Case for Redundant Arrays of Inexpensive Disks (RAID). D.A. Patterson, G.A. Gibson, R.H. Katz	2	1	1	8	7	2	10	3	4	1	4	1	1
1989	F-Logic: a Higher-Order language for Reasoning about Objects, Inheritance, and Scheme. M. Kifer, G. Lausen	5	4	6	5	4	5	5	5	8	6	7	4	4
1990	Encapsulation of Parallelism in the Volcano Query Processing System. G. Graefe	10	9	10	5	9	10	7	10	7	10	7	9	9
1990	Set-Oriented Production Rules in Relational Database Systems. J. Widom, S.J. Finkelstein	3	3	3	4	15	3	4	3	3	3	3	3	3
1992	Extensible/Rule Based Query Rewrite Optimization in Starburst. H. Pirahesh, J.M. Hellerstein, W. Hasan	3	2	1	2	5	3	5	2	3	1	3	1	1
1992	Querying Object-Oriented Databases. M. Kifer, W. Kim, Y. Sagiv	1	5	2	1	8	1	1	3	2	2	1	3	3
1993	Mining Association Rules between Sets of Items in Large Databases. R. Agrawal, T. Imielinski, A.N. Swami	1	1	1	6	26	1	5	1	1	1	1	1	1
1994	From Structured Documents to Novel Query Facilities. V. Christophides, S. Abiteboul, S. Cluet, M. Scholl	2	2	2	7	8	2	3	2	1	2	2	2	2
1994	Shoring Up Persistent Applications. M.J. Carey, D.J. DeWitt, M.J. Franklin, N.E. Hall, M.L. McAuliffe, J.F. Naughton, D.T. Schuh, M.H. Solomon, C.K. Tan, O.G. Tsatalos, S.J. White, M.J. Zwilling	1	1	1	1	1	1	2	1	2	1	1	1	1
		28	28	27	39	83	28	42	30	31	27	29	25	25

Table 16: VLDB awared papers.

Year	Title	CC	BCC	PR	HA	P	SA	BHA	BSA	PS	BPS	EPS	BEPS	SCEAS_B1
1986	Object and File Management in the EXODUS Extensible Database System. M.J. Carey, D.J. DeWitt, J.E. Richardson, E.J. Shekita	2	3	2	1	2	2	2	2	1	2	2	3	3
1987	The R^+-Tree: a Dynamic Index for Multi-Dimensional Objects. T.K. Sellis, N. Roussopoulos, C. Faloutsos	1	1	1	1	19	1	1	1	1	1	1	1	1
1988	Disk Shadowing. D. Bitton, J. Gray	2	1	1	5	8	2	11	3	2	1	2	1	1
1989	ARIES/NT: a Recovery Method Based on Write-Ahead Logging for Nested Transactions. K. Rothermel, C. Mohan	6	9	6	14	1	6	13	7	1	6	2	7	7
1990	Deriving Production Rules for Constraint Maintainance. S. Ceri, J. Widom	1	1	1	3	17	1	6	1	1	1	1	1	1
1991	A Transactional Model for Long-Running Activities. U. Dayal, M. Hsu, R. Ladin	4	2	2	24	22	4	17	7	12	2	9	2	2
1992	Querying in Highly Mobile Distributed Environments. T. Imielinski, B.R. Badrinath	12	7	3	43	10	12	30	15	11	3	12	4	4
1993	Universality of Serial Histograms. Y.E. Ioannidis	5	9	6	8	11	5	8	4	4	6	4	7	7
1994	Fast Algorithms for Mining Association Rules in Large Databases. R. Agrawal, R. Srikant	1	1	1	4	16	1	8	1	1	1	1	1	1
1995	W3QS: a Query System for the World-Wide Web. D. Konopnicki, O. Shmueli	3	2	4	7	3	3	1	4	2	4	2	3	3
		37	36	27	110	109	37	97	45	36	27	36	30	30

In Tables 15-16 we show the awarded publications and their rank position for all ranking methods. For example, the paper entitled ‘Fast Algorithms for Mining Association Rules in Large DBs, 1994’ (see Table 16) is ranked 1st by `PageRank`, 1st by `SCEAS`, 4th by `HITS` (as an authority), 16th by `Prestige` and 1st by `CC, BCC, PR, SA, BSA`. Notice again that this way we do not evaluate the awarding committees but the ranking methods. In these detailed tables we observe that the awarded papers are generally highly ranked. Apparently some deviations and exceptions do exist. These exceptions may exist for several reasons:

- Our citations sample may be not large enough: e.g. an awarded publication may get a lot of citations from scientific domains that are not included in the DBLP dataset.

- By definition, the awards are subjective: e.g. an awards committee decision may be based on objective factors (such as the number of received citations) but also may combine other measures and indicators of impact.
- It is rather unusual for an author to be awarded twice by the same organization.

To get an overall view of the performance of the ranking methods, we sum the positions of the awarded publications in the last row. The smaller this sum, the better performance of the ranking method. In both tables we remark that the `HITS Ranking (Authorities)` and `Prestige` are by far the worst methods. This verifies our remarks explained in Section 2. `BHITS` is slightly better than `HITS` for the case of Table 16. Among the other ten methods, `CC, BCC, SA, BSA, PS` and `EPS` seem to be in the same medium category, but they still remain quite good ranking methods. Finally, we see that `PageRank, BPS, BEPS` and `SCEAS` are very close to each other and they alternate at the winning position in the two Tables 15 and 16. In the sequel, we will ignore the `HITS Rank (Authorities)` and `Prestige`, since they are not suitable for our ranking purposes.

Since all algorithms (except `HITS` and `Prestige`) ended up with a similar behavior, we will try to produce a single rank table by averaging their results along the reasoning of [14]. In simple words, we will compute all the ten rankings and assign to each paper a number of points (5 to 1) depending on its position in the specific rank table. For example, in each table the first paper gets 5 points, the second one gets 4 points, and so on. Thus, if a paper is ranked first in all 10 rankings, then it will get 50 points. Therefore, we repeat steps 3 and 4 of the previous scenario to produce the new rank tables.

This latter computation is reported in Tables 17-18, where the column 'Pos' represents the position of the corresponding paper. It can easily be seen that the majority of the awarded publications are ranked in the top 3 positions of these new rank tables. After exhaustive experiments we also concluded that the sum of the positions is smaller by using this averaged approach in comparison to any stand alone ranking method. In particular, we remark that in the SIGMOD case (see Table 17) the sum of positions (i.e. 26) is better than the average sum of Table 15 and slightly larger than the best algorithms of this case (i.e. 25 for both `BEPS` and `SCEAS General`) mainly due to the 1990/2000 outlier paper. In the VLDB case (Table 18) the sum (i.e. 28) again is smaller than the average sum of Table 16 and slightly larger than `PageRank` (27) and `BPS` (27).

In the Appendix A we present a detailed ranking of SIGMOD'95 and VLDB'95 publications for all of the methods. Thus, we may suggest and with high probability predict the publications that will be awarded during the next few years. Apparently, the possibility of an error increases as we consider more distant years in the future. In Table 20 we present our predictions for the next 5 years of the VLDB 10 Year Award case, and in Table 19 for the next 5 years of the SIGMOD Test of Time Award. For each year, we pop the top three papers with the highest average score.

Table 17: Sum of rank positions of pubs awarded with the SIGMOD Test of Time over DBLP.

Year	Title	Pos	Points
1988	A Case for Redundant Arrays of Inexpensive Disks (RAID). D.A. Patterson, G.A. Gibson, R.H. Katz	1	37
1989	F-Logic: a Higher-Order language for Reasoning about Objects, Inheritance, and Scheme. M. Kifer, G. Lausen	5	9
1990	Encapsulation of Parallelism in the Volcano Query Processing System. G. Graefe	8	0
1990	Set-Oriented Production Rules in Relational Database Systems. J. Widom, S. J. Finkelstein	3	27
1992	Extensible/Rule Based Query Rewrite Optimization in Starburst. H. Pirahesh, J.M. Hellerstein, W. Hasan	2	37
1992	Querying Object-Oriented Databases. M. Kifer, W. Kim, Y. Sagiv	3	31
1993	Mining Association Rules between Sets of Items in Large Databases. R. Agrawal, T. Imielinski, A.N. Swami	1	45
1994	From Structured Documents to Novel Query Facilities. V. Christophides, S. Abiteboul, S. Cluet, M. Scholl	2	37
1994	Shoring Up Persistent Applications. M.J. Carey, D.J. DeWitt, M.J. Franklin, N.E. Hall, M.L. McAuliffe, J.F. Naughton, D.T. Schuh, M.H. Solomon, C.K. Tan, O.G. Tsatalos, S.J. White, M.J. Zwilling	1	44
	Sum of positions	26	

Table 18: Sum of rank positions of pubs awarded with the VLDB 10 Year Award over DBLP.

Year	Title	Pos	Points
1986	Object and File Management in the EXODUS Extensible Database System. M.J. Carey, D.J. DeWitt, J.E. Richardson, E.J. Shekita	2	34
1987	The R^+-Tree: a Dynamic Index for Multi-Dimensional Objects. T.K. Sellis, N. Roussopoulos, C. Faloutsos	1	42
1988	Disk Shadowing. D. Bitton, J. Gray	1	38
1989	ARIES/NT: a Recovery Method Based on Write-Ahead Logging for Nested Transactions. K. Rothermel, C. Mohan	6	5
1990	Deriving Production Rules for Constraint Maintenance. S. Ceri, J. Widom	1	43
1991	A Transactional Model for Long-Running Activities. U. Dayal, M. Hsu, R. Ladin	3	24
1992	Querying in Highly Mobile Distributed Environments. T. Imielinski, B.R. Badrinath	4	10
1993	Universality of Serial Histograms. Y.E. Ioannidis	6	6
1994	Fast Algorithms for Mining Association Rules in Large Databases. R. Agrawal, R. Srikant	1	45
1995	W3QS: A Query System for the World-Wide Web. D. Konopnicki, O. Shmueli	3	30
	Sum of positions	28	

5.2 Ranking Authors

We may rely on our method of computing scores for publications and compute scores for authors as well. One approach could be to compute the average score of all their publications. This is again not a trivial task. For instance, author A has 200 publications with only 40 ones being 'first class'. Assume that these high quality publications have a score of 10 points each, whereas the remaining ones have a score of 1 point. Author B has in total 20 publications, with 10 publications of them being 'first class'. It is reasonable to consider that author A should be ranked higher than author B for his scientific contribution, because A has 4 times the number of first class publications than author B. However, if we just compute the average of all publication scores, then authors A and B would have 2.8 and 5.5 points respectively. Therefore, it is not fair to take into account all the publications a person has authored. In addition, it is not fair to take into account different number of publications for each author (e.g. 40 publications for A and 10 for B).

In our approach, we take into account the same number of publications for all authors so that the score results could be comparable. Therefore, our problem now is to choose the number of publications of each author that should be considered in the ranking. We performed the following experiment to determine this number. We computed the average score for each author by using his best x publications, $\forall x \in \{1, 3, 5, 8, 10, 15, 20, 25, 30, 40\}$. Thus, we produced 10 rankings for every ranking method. As a testbed we used the authors that were awarded the '*SIGMOD Edgar F. Codd Innovations Award*'. The higher these authors were ranked, the better the evaluation was

Table 19: *SIGMOD Test of Time Award* prediction for years 2005-2008 (1995-1998).

Year	Title	Pos	Points
2005 (1995)	View Maintenance in a Warehousing Environment. Y. Zhuge, H. Garcia-Molina, J. Hammer, J. Widom	1	44
	An Effective Hash Based Algorithm for Mining Association Rules. J.S. Park, M. Chen, P.S. Yu	2	40
	Nearest Neighbor Queries. N. Roussopoulos, S. Kelley, F. Vincent	3	36
	Incremental Maintenance of Views with Duplicates. T. Griffin, L. Libkin	4	19
	Balancing Histogram Optimality and Practicality for Query Result Size Estimation. Y.E. Ioannidis, V. Poosala	5	4
2006 (1996)	Implementing Data Cubes Efficiently. V. Harinarayan, A. Rajaraman, J.D. Ullman	1	50
	A Query Language and Optimization Techniques for Unstructured Data. P. Buneman, S.B. Davidson, G.G. Hillebrand, D. Suciu	2	40
	BIRCH: an Efficient Data Clustering Method for Very Large Databases. T. Zhang, R. Ramakrishnan, M. Livny	3	27
	Improved Histograms for Selectivity Estimation of Range Predicates. V. Poosala, Y.E. Ioannidis, P.J. Haas, E.J. Shekita	4	23
	Data Access for the Masses through OLE DB. J.A. Blakeley	5	5
2007 (1997)	Online Aggregation. J.M. Hellerstein, P.J. Haas, H.J. Wang	1	50
	An Array-Based Algorithm for Simultaneous Multidimensional Aggregates. Y. Zhao, P. Deshpande, J.F. Naughton	2	40
	Improved Query Performance with Variant Indexes. P.E. O'Neil, D. Quass	3	22
	Dynamic Itemset Counting and Implication Rules for Market Basket Data. S. Brin, R. Motwani, J.D. Ullman, S. Tsur	4	11
	The SR-tree: An Index Structure for High-Dimensional Nearest Neighbor Queries. N. Katayama, S. Satoh	5	10
2008 (1998)	Catching the Boat with Strudel: Experiences with a Web-Site Management System. M.F. Fernandez, D. Florescu, J. Kang, A.Y. Levy, D. Suciu	1	45
	Your Mediators Need Data Conversion! S. Cluet, C. Delobel, J. Siméon, K. Smaga	2	35
	New Sampling-Based Summary Statistics for Improving Approximate Query Answers. P. B. Gibbons, Y. Matias	3	24
	Integrating Mining with Relational Database Systems: Alternatives and Implications. S. Sarawagi, S. Thomas, R. Agrawal	4	23
	Exploratory Mining and Pruning Optimizations of Constrained Association Rules. R.T. Ng, L.V.S. Lakshmanan, J. Han, A. Pang	5	8

considered. In Appendix 5.2.1 we present the results that were produced by this experiment. Based on this experiment, we concluded that the average of 25 best publications is the most appropriate measure (and alternatively the average of the best 30 publications of each author).

In Tables 21-22 we compare the various ranking methods. In these tables we present the rank positions of the awarded authors for each ranking method by taking into account the average score of the best 25 and 30 publications of each author, respectively. It is obvious that column `SCEAS_B1` of Table 21 is identical to column 'best25' of Table 25, and column `SCEAS_B1` of Table 22 is identical to column 'best30' of Table 25.

Table 20: *VLDB 10 Year Award* prediction for years 2005-2008 (1995-1998).

Year	Title	Pos	Points
2005 (1995)	Aggregate-Query Processing in Data Warehousing Environments. A. Gupta, V. Harinarayan, D. Quass	1	46
	Discovery of Multiple-Level Association Rules from Large Databases. J. Han, Y. Fu	2	35
	W3QS: a Query System for the World-Wide Web. D. Konopnicki, O. Shmueli	3	30
	Sampling-Based Estimation of the Number of Distinct Values of an Attribute. P.J. Haas, J.F. Naughton, S. Seshadri, L. Stokes	4	23
	Mining Generalized Association Rules. R. Srikant, R. Agrawal	5	8
2006 (1996)	Querying Heterogeneous Information Sources Using Source Descriptions A.Y. Levy, A. Rajaraman, J.J. Ordille	1	47
	On the Computation of Multidimensional Aggregates. S. Agarwal, R. Agrawal, P. Deshpande, A. Gupta, J.F. Naughton, R. Ramakrishnan, S. Sarawagi	2	39
	The X-tree : an Index Structure for High-Dimensional Data S. Berchtold, D.A. Keim, H. Kriegel	3	28
	Querying Multiple Features of Groups in Relational Databases. D. Chatziantoniou, K.A. Ross	4	16
	Answering Queries with Aggregation Using Views. D. Srivastava, S. Dar, H.V. Jagadish, A.Y. Levy	5	8
2007 (1997)	Optimizing Queries across Diverse Data Sources. L.M. Haas, D. Kossmann, E.L. Wimmers, J. Yang	1	47
	Don't Scrap It, Wrap It! A Wrapper Architecture for Legacy Data Sources. M.T. Roth, P.M. Schwarz	2	37
	DataGuides: Enabling Query Formulation and Optimization in Semistructured Databases. R. Goldman, J. Widom	3	35
	Selectivity Estimation Without the Attribute Value Independence Assumption. V. Poosala, Y.E. Ioannidis	4	20
	To Weave the Web. P. Atzeni, G. Mecca, P. Merialdo	5	4
2008 (1998)	Optimal Histograms with Quality Guarantees. H.V. Jagadish, N. Koudas, S. Muthukrishnan, V. Poosala, K.C. Sevcik, T. Suel	1	31
	Hash Joins and Hash Teams in Microsoft SQL Server. G. Graefe, R. Bunker, S. Cooper	2	28
	Clustering Categorical Data: An Approach Based on Dynamical Systems. D. Gibson, J.M. Kleinberg, P. Raghavan	3	20
	A Quantitative Analysis and Performance Study for Similarity-Search Methods in High-Dimensional Spaces. R. Weber, H. Schek, S. Blott	4	16
	Using Schema Matching to Simplify Heterogeneous Data Translation. T. Milo, S. Zohar	5	12

The last two rows of Tables 21-22 show the rank position of the awarded author that ranked last ('lowest ranking point') and the 'sum of ranking positions' of all the awarded authors. These two numbers serve as metrics for comparing the rankings. The lower these numbers, the better ranking is performed. In this table we can see that `HITS authorities` and `Prestige` are by far the worst methods again, since the 'sum of ranking points' and the 'lowest ranking point' are about 2 and 3 times greater than the respective numbers computed for the other methods. Plain Citation Count (`CC`) and Balanced Citation Count (`BCC`) give acceptable results. By far the best methods are

Table 21: Positions of awarded authors of DBLP by average of best 25 publications.

Author Name	Rank Position by:												
	CC	BCC	PR	HA	P	SA	BHA	BSA	PS	BPS	EPS	BEPS	SCEAS_B1
Michael Stonebraker	2	3	4	2	4	2	3	2	2	5	2	4	4
Jim Gray	3	1	2	9	9	3	10	4	6	2	5	1	1
Philip Bernstein	6	7	6	6	2	6	1	5	5	6	6	8	8
David DeWitt	1	4	8	4	55	1	21	8	11	8	8	5	5
C. Mohan	29	32	40	62	176	29	95	40	41	43	34	34	33
David Maier	9	10	13	8	21	9	23	12	14	14	13	11	11
Serge Abiteboul	13	19	23	16	15	13	62	21	26	24	22	22	21
Hector Garcia-Molina	21	16	20	115	291	21	201	36	69	21	50	19	19
Rakesh Agrawal	15	12	16	82	407	15	182	28	70	17	45	12	12
Rudolf Bayer	73	51	24	117	32	74	19	30	17	20	24	36	40
Patricia Selinger	38	39	28	24	73	38	41	25	19	26	18	32	34
Donald Chamberlin	14	9	5	11	3	14	9	7	3	4	4	7	7
Ronald Fagin	17	18	10	17	5	17	7	9	9	10	11	14	14
Lowest Ranking Point	73	51	40	117	407	74	201	40	70	43	50	36	40
Sum of Ranking Points	241	221	199	473	1093	242	674	227	292	200	242	205	209

PageRank, B-SALSA, BPS, BEPS and SCEAS. Also, B-SALSA improves SALSA by more than 50%. In Appendix B we present the full rank results for authors.

5.2.1 Ranking Authors Experiment

In this Section we present the results produced by the experiment mentioned in Section 5.2 aiming in finding the number of publications we should take into account for each author. We present the rank results of CC (Table 23), BEPS (Table 24) and SCEAS (Table 25) methods. For brevity, we present only the awarded authors and their position for each selected number of 'top' publications. For example, Hector-Garcia Molina is ranked 81st in the ranking by SCEAS, if the score is produced by the average of top-1 publication of each author. He is ranked 41st in the ranking, if the score is produced by the average of top-3 publications, and so on.

The last two rows of Tables 23- 25 show the rank position of the awarded author that ranked last ('lowest ranking point') and the 'sum of ranking positions' of all the awarded authors. These two numbers are our metrics for comparing the rankings. The lower these numbers, the better ranking is performed. The lowest ranking point is significantly higher when computing the average for the top 1-3 publications of each author. This is due to the fact that the co-authors of a top publication take advantage and 'climb up' the ranking results. Therefore, by increasing the number of the selected top publications, the awarded authors move towards the top of the rank table. This trend holds until the number of the selected publications becomes 25. The same remark holds when we consider the notion of 'sum of ranking positions'. Also note that in column 40 we miss 2 of the awarded authors, since they have less than 40 publications in the DBLP digital library. For brevity, we have not included the respective tables for other ranking methods, since the results are similar and the smallest ranking point is the same. Thus, after this experiment we decided to rank the authors by averaging their 25 best publications. As a second choice we also selected the averaging of 30-best publications.

Table 22: Positions of awarded authors of DBLP by average of best 30 publications

Author Name	Rank Position by:												
	CC	BCC	PR	HA	P	SA	BHA	BSA	PS	BPS	EPS	BEPS	SCEAS_B1
Michael Stonebraker	1	2	4	1	4	1	3	2	2	4	1	3	3
Jim Gray	3	1	2	10	9	3	10	4	6	2	5	1	1
Philip A. Bernstein	6	7	6	6	2	6	1	5	5	6	6	7	7
David DeWitt	2	3	8	3	50	2	21	8	11	8	8	5	5
C. Mohan	29	30	39	57	162	29	90	41	39	42	34	32	32
David Maier	8	12	14	9	20	8	23	13	14	14	13	11	12
Serge Abiteboul	12	19	22	15	14	12	58	20	23	23	21	21	21
Hector Garcia-Molina	21	14	20	101	270	20	184	35	64	20	46	18	17
Rakesh Agrawal	14	11	16	68	371	14	168	26	61	17	40	12	11
Rudolf Bayer	72	53	24	111	31	73	19	32	17	22	25	41	41
Patricia Selinger	42	43	31	25	68	42	39	28	19	27	19	34	35
Donald Chamberlin	16	9	5	11	3	16	9	7	3	5	4	8	8
Ronald Fagin	19	18	10	17	5	19	7	9	9	10	11	15	16
Lowest Ranking Point	72	53	39	111	371	73	184	41	64	42	46	41	41
Sum of Ranking Points	245	222	201	434	1009	245	632	230	273	200	233	208	209

Table 23: Positions of awarded authors by CC.

Author Name	Rank Position by:									
	best1	best3	best5	best8	best10	best15	best20	best25	best30	best40
C. Mohan	102	92	79	54	48	36	34	29	29	26
David J. DeWitt	34	16	12	7	5	4	1	1	2	2
David Maier	44	19	13	11	9	9	8	9	8	7
Donald D. Chamberlin	5	4	4	6	7	10	11	14	16	
Hector Garcia-Molina	146	58	42	35	31	27	26	21	21	15
Jim Gray	8	3	2	2	2	2	2	3	3	4
Michael Stonebraker	23	9	5	4	4	3	3	2	1	1
Patricia G. Selinger	3	13	17	21	23	31	36	38	42	
Philip A. Bernstein	52	21	15	13	10	7	6	6	6	5
Rakesh Agrawal	101	44	35	27	25	20	15	15	14	12
Ronald Fagin	86	33	23	20	20	19	16	17	19	18
Rudolf Bayer	45	42	48	62	65	73	73	73	72	69
Serge Abiteboul	83	31	25	19	19	16	14	13	12	8
Lowest Ranking Point	146	92	79	62	65	73	73	73	72	69
Sum of Rank Points	732	385	320	281	268	257	245	241	245	167

A final remark with respect to this table is the following. Quite interesting and easily explained is the fact that there are several authors whose ranking position gets higher with increasing number of top publications (with S. Abiteboul advancing the most), whereas the opposite holds for a few other authors (e.g. P. Selinger due to her specific work on System R and R. Bayer due to his work on B-trees and related structures). Jim Gray steadily holds top positions.

6 Conclusion

In this report we proposed and experimentally examined a family of new alternative methods (in our SCEAS system) for scientific publications evaluation, besides the known algorithms of PageRank and HITS. We also evaluated the above methods by using the DBLP digital library as a training

Table 24: Positions of awarded authors by `BEPS`.

Author Name	Rank Position by:									
	best1	best3	best5	best8	best10	best15	best20	best25	best30	best40
C. Mohan	104	105	84	69	61	48	42	34	32	27
David J. DeWitt	30	20	15	13	12	7	7	5	5	3
David Maier	39	28	21	17	16	13	12	11	11	8
Donald D. Chamberlin	9	4	4	4	4	4	5	7	8	
Hector Garcia-Molina	81	44	38	30	27	22	22	19	18	13
Jim Gray	3	3	2	2	2	2	2	1	1	1
Michael Stonebraker	21	10	9	7	5	5	4	4	3	2
Patricia G. Selinger	7	22	24	26	26	32	33	32	34	
Philip A. Bernstein	57	27	18	14	11	8	8	8	7	5
Rakesh Agrawal	48	33	27	22	20	17	15	12	12	7
Ronald Fagin	41	29	22	18	18	15	17	14	15	14
Rudolf Bayer	22	25	29	33	38	35	37	36	41	35
Serge Abiteboul	148	64	48	41	37	26	25	22	21	16
Lowest Ranking Point	148	105	84	69	61	48	42	36	41	35
Sum of Rank Points	610	414	341	296	277	234	229	205	208	131

Table 25: Positions of awarded authors by `SCEAS_B1`.

Author Name	Rank Position by:									
	best1	best3	best5	best8	best10	best15	best20	best25	best30	best40
C. Mohan	104	103	84	69	61	48	41	33	32	26
David J. DeWitt	30	18	14	12	11	7	5	5	5	3
David Maier	41	26	20	17	16	12	12	11	12	8
Donald D. Chamberlin	9	5	4	4	4	5	7	7	8	
Hector Garcia-Molina	81	41	36	28	26	22	22	19	17	12
Jim Gray	3	3	2	2	2	2	1	1	1	1
Michael Stonebraker	21	9	9	7	5	4	4	4	3	2
Patricia G. Selinger	7	21	23	25	27	32	33	34	35	
Philip A. Bernstein	67	28	17	13	12	9	8	8	7	5
Rakesh Agrawal	47	33	25	20	20	15	14	12	11	7
Ronald Fagin	49	29	22	18	17	16	17	14	16	14
Rudolf Bayer	23	24	29	33	40	35	40	40	41	39
Serge Abiteboul	146	64	46	41	35	26	25	21	21	16
Lowest Ranking Point	146	103	84	69	61	48	41	40	41	39
Sum of Rank Points	628	404	331	289	276	233	229	209	209	133

set and the awarded publications of '*VLDB 10 Year Award*' and '*SIGMOD Test of Time Award*' as an evaluation set for the publications ranking method. Additionally, we presented author ranking based on the publication rank results and we used the '*SIGMOD Edgar F. Codd Innovations Award*' as an evaluation set. Evidently, it is not our intention to suggest which of them should be awarded in the coming years. However, we believe that our objective approach is of help to the respective committees. In all the above cases the performance of our `SCEAS` method was in general better than the other methods.

References

[1] A. Borodin, J. S. Rosenthal, G. O. Roberts, and P. Tsaparas. Link Analysis Ranking: Algorithms, Theory and Experiments. *ACM Transactions on Internet Technologies*, 5(1):231–297, 2005.

[2] S. Brin and L. Page. The Anatomy of a Large-scale Hypertextual Web Search Engine. In *Proceedings 7th WWW Conference*, pages 107–117, 1998.

[3] S. Chakrabarti. *Mining the Web: Discovering Knowledge from Hypertext Data*, chapter 7.1, pages 205–206. Morgan Kaufmann Publishers, 2003.

[4] P. Diaconis and RL Graham. Spearman's Footrule as a Measure of Disarray. *Journal of the Royal Society of Statistics, Series B*, 39:262–268., 1977.

[5] E. Garfield. Science Citation Index. http://www.isinet.com/isi/.

[6] E. Garfield. Citation Analysis as a Tool in Journal Evaluation. *Essays of an Information Scientist*, 1:527–544, 1972.

[7] E. Garfield. The Impact Factor, 1994. http://www.isinet.com/isi/hot/essays/ journalcitation-reports/7.html.

[8] Sepandar D. Kamvar and Taher H. Haveliwala. The Condition Number of the PageRank Problem. Technical report, Stanford University, 2003.

[9] M.G. Kendall. *Rank Correlation Methods*. Charles Griffin and Co., London, 1970.

[10] J.M. Kleinberg. Authoritative Sources in a Hyperlinked Environment. *Journal of the ACM*, 46(5):604–632, 1999.

[11] J.M. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins. The Web as a Graph: Measurements, Models, and Methods. In *Proceedings 5th COCOON Conference*, pages 1–17, 1999.

[12] J. Kleinjnen and W. Van Groenendaal. Measuring the Quality of Publications: New Methodology and Case Study. *Information Processing and Management*, 36(4):551–570, 2000.

[13] Ronny Lempel and Shlomo Moran. SALSA: the Stochastic Approach for Link-structure Analysis. *ACM Transactions on Information Systems*, 19(2):131–160, 2001.

[14] R.K. Rainer and M.D. Miller. Examining Differences across Journal Rankings. *Communications of the ACM*, 48(2):91–94, 2005.

[15] A. Sidiropoulos and Y. Manolopoulos. A New Perspective to Automatically Rank Scientific Conferences Using Digital Libraries. *Information Processing and Management*, 41(2):289–312, 2005.

[16] Antonis Sidiropoulos and Yannis Manolopoulos. A Citation-based System to Assist Prize Awarding. *SIGMOD Record*, 34(4):54–60, Dec 2005.

Appendix A: Ranking Publications of SIGMOD 1995 and VLDB 1995

In this section we present the rank results by some methods for the SIGMOD 1995 Conference in proceedings as well as the VLDB 1995 Conf. Note here that our database includes citation only until 2000. Thus, 1995 publications could get citations during 5 years only. As we can see by all methods for the *SIGMOD* case (Tables 26-36) three publications alterna te for the winning place. The same is noticable for the *VDBL* case in Tables 37-47.

Table 26: Rank of DBLP components belonging to sigmod 1995 by CC

Score	Title
49	View Maintenance in a Warehousing Environment. Y. Zhuge, H. Garcia-Molina, J. Hammer, J. Widom
46	Nearest Neighbor Queries. N. Roussopoulos, S. Kelley, F. Vincent
44	An Effective Hash Based Algorithm for Mining Association Rules. J. S. Park, M. Chen, P. S. Yu
37	Incremental Maintenance of Views with Duplicates. T. Griffin, L. Libkin
28	Balancing Histogram Optimality and Practicality for Query Result Size Estimation. Y. E. Ioannidis, V. Poosala
22	Broadcast Disks: Data Management for Asymmetric Communications Environments. S. Acharya, R. Alonso, M. J. Franklin, S. B. Zdonik
21	FastMap: A Fast Algorithm for Indexing, Data-Mining and Visualization of Traditional and Multimedia Datasets. C. Faloutsos, K. Lin
14	Adapting Materialized Views after Redefinitions. A. Gupta, I. S. Mumick, K. A. Ross
14	Efficient Maintenance of Materialized Mediated Views. J. J. Lu, G. Moerkotte, J. Schü, V. S. Subrahmanian
10	A Critique of ANSI SQL Isolation Levels. H. Berenson, P. A. Bernstein, J. Gray, J. Melton, E. J. O'Neil, P. E. O'Neil

Table 27: Rank of DBLP components belonging to sigmod 1995 by BCC

Score	Title
4.612	An Effective Hash Based Algorithm for Mining Association Rules. J. S. Park, M. Chen, P. S. Yu
4.594	View Maintenance in a Warehousing Environment. Y. Zhuge, H. Garcia-Molina, J. Hammer, J. Widom
3.4	Nearest Neighbor Queries. N. Roussopoulos, S. Kelley, F. Vincent
3.369	Broadcast Disks: Data Management for Asymmetric Communications Environments. S. Acharya, R. Alonso, M. J. Franklin, S. B. Zdonik
3.237	Incremental Maintenance of Views with Duplicates. T. Griffin, L. Libkin
2.626	A Critique of ANSI SQL Isolation Levels. H. Berenson, P. A. Bernstein, J. Gray, J. Melton, E. J. O'Neil, P. E. O'Neil
2.551	Balancing Histogram Optimality and Practicality for Query Result Size Estimation. Y. E. Ioannidis, V. Poosala
2.236	Fault Tolerant Design of Multimedia Servers. S. Berson, L. Golubchik, R. R. Muntz
1.726	FastMap: A Fast Algorithm for Indexing, Data-Mining and Visualization of Traditional and Multimedia Datasets. C. Faloutsos, K. Lin
1.598	Adapting Materialized Views after Redefinitions. A. Gupta, I. S. Mumick, K. A. Ross

Table 28: Rank of DBLP components belonging to sigmod 1995 by PR

Score	Title
2.941	An Effective Hash Based Algorithm for Mining Association Rules. J. S. Park, M. Chen, P. S. Yu
2.775	View Maintenance in a Warehousing Environment. Y. Zhuge, H. Garcia-Molina, J. Hammer, J. Widom
2.004	Nearest Neighbor Queries. N. Roussopoulos, S. Kelley, F. Vincent
1.884	Incremental Maintenance of Views with Duplicates. T. Griffin, L. Libkin
1.763	A Critique of ANSI SQL Isolation Levels. H. Berenson, P. A. Bernstein, J. Gray, J. Melton, E. J. O'Neil, P. E. O'Neil
1.712	Balancing Histogram Optimality and Practicality for Query Result Size Estimation. Y. E. Ioannidis, V. Poosala
1.664	Broadcast Disks: Data Management for Asymmetric Communications Environments. S. Acharya, R. Alonso, M. J. Franklin, S. B. Zdonik
1.325	Adapting Materialized Views after Redefinitions. A. Gupta, I. S. Mumick, K. A. Ross
1.204	FastMap: A Fast Algorithm for Indexing, Data-Mining and Visualization of Traditional and Multimedia Datasets. C. Faloutsos, K. Lin
1.182	Fault Tolerant Design of Multimedia Servers. S. Berson, L. Golubchik, R. R. Muntz

Table 29: Rank of DBLP components belonging to sigmod 1995 by HA

Score	Title
1.908	Nearest Neighbor Queries. N. Roussopoulos, S. Kelley, F. Vincent
1.186	View Maintenance in a Warehousing Environment. Y. Zhuge, H. Garcia-Molina, J. Hammer, J. Widom
1.075	Incremental Maintenance of Views with Duplicates. T. Griffin, L. Libkin
0.9303	Balancing Histogram Optimality and Practicality for Query Result Size Estimation. Y. E. Ioannidis, V. Poosala
0.6591	Efficient Maintenance of Materialized Mediated Views. J. J. Lu, G. Moerkotte, J. Schü, V. S. Subrahmanian
0.5741	FastMap: A Fast Algorithm for Indexing, Data-Mining and Visualization of Traditional and Multimedia Datasets. C. Faloutsos, K. Lin
0.5189	Topological Relations in the World of Minimum Bounding Rectangles: A Study with R-trees. D. Papadias, Y. Theodoridis, T. K. Sellis, M. J. Egenhofer
0.4233	Towards an Effective Calculus for Object Query Languages. L. Fegaras, D. Maier
0.3906	A General Solution of the n-dimensional B-tree Problem. M. Freeston
0.3874	The LyriC Language: Querying Constraint Objects. A. Brodsky, Y. Kornatzky

Table 30: Rank of DBLP components belonging to sigmod 1995 by P

Score	Title
2.887e-103	View Maintenance in a Warehousing Environment. Y. Zhuge, H. Garcia-Molina, J. Hammer, J. Widom
2.26e-103	Incremental Maintenance of Views with Duplicates. T. Griffin, L. Libkin
1.615e-103	Adapting Materialized Views after Redefinitions. A. Gupta, I. S. Mumick, K. A. Ross
1.422e-103	Efficient Maintenance of Materialized Mediated Views. J. J. Lu, G. Moerkotte, J. Schü, V. S. Subrahmanian
4.387e-104	Applying Update Streams in a Soft Real-Time Database System. B. Adelberg, H. Garcia-Molina, B. Kao
5.339e-105	Parallel Database Systems 101. J. Gray
8.276e-144	Hypergraph Based Reorderings of Outer Join Queries with Complex Predicates. G. Bhargava, P. Goel, B. R. Iyer
7.289e-145	A Database Interface for File Updates. S. Abiteboul, S. Cluet, T. Milo
4.505e-145	Data Extraction and Transformation for the Data Warehouse. C. Squire
4.505e-145	Copy Detection Mechanisms for Digital Documents. S. Brin, J. Davis, H. Garcia-Molina

Table 31: Rank of DBLP components belonging to sigmod 1995 by SA

Score	Title
8.113	View Maintenance in a Warehousing Environment. Y. Zhuge, H. Garcia-Molina, J. Hammer, J. Widom
7.616	Nearest Neighbor Queries. N. Roussopoulos, S. Kelley, F. Vincent
7.285	An Effective Hash Based Algorithm for Mining Association Rules. J. S. Park, M. Chen, P. S. Yu
6.126	Incremental Maintenance of Views with Duplicates. T. Griffin, L. Libkin
4.636	Balancing Histogram Optimality and Practicality for Query Result Size Estimation. Y. E. Ioannidis, V. Poosala
3.642	Broadcast Disks: Data Management for Asymmetric Communications Environments. S. Acharya, R. Alonso, M. J. Franklin, S. B. Zdonik
3.477	FastMap: A Fast Algorithm for Indexing, Data-Mining and Visualization of Traditional and Multimedia Datasets. C. Faloutsos, K. Lin
2.318	Efficient Maintenance of Materialized Mediated Views. J. J. Lu, G. Moerkotte, J. Schü, V. S. Subrahmanian
2.318	Adapting Materialized Views after Redefinitions. A. Gupta, I. S. Mumick, K. A. Ross
2.117	The Lotus Notes Storage System. K. Moore

Table 32: Rank of DBLP components belonging to sigmod 1995 by BHA

Score	Title
0.2548	Nearest Neighbor Queries. N. Roussopoulos, S. Kelley, F. Vincent
0.1271	Adapting Materialized Views after Redefinitions. A. Gupta, I. S. Mumick, K. A. Ross
0.1218	View Maintenance in a Warehousing Environment. Y. Zhuge, H. Garcia-Molina, J. Hammer, J. Widom
0.08248	Incremental Maintenance of Views with Duplicates. T. Griffin, L. Libkin
0.07995	FastMap: A Fast Algorithm for Indexing, Data-Mining and Visualization of Traditional and Multimedia Datasets. C. Faloutsos, K. Lin
0.07879	Balancing Histogram Optimality and Practicality for Query Result Size Estimation. Y. E. Ioannidis, V. Poosala
0.07783	Efficient Maintenance of Materialized Mediated Views. J. J. Lu, G. Moerkotte, J. Schü, V. S. Subrahmanian
0.06473	Topological Relations in the World of Minimum Bounding Rectangles: A Study with R-trees. D. Papadias, Y. Theodoridis, T. K. Sellis, M. J. Egenhofer
0.0357	Towards an Effective Calculus for Object Query Languages. L. Fegaras, D. Maier
0.03546	The LyriC Language: Querying Constraint Objects. A. Brodsky, Y. Kornatzky

Table 33: Rank of DBLP components belonging to sigmod 1995 by BSA

Score	Title
2.342	Nearest Neighbor Queries. N. Roussopoulos, S. Kelley, F. Vincent
2.334	View Maintenance in a Warehousing Environment. Y. Zhuge, H. Garcia-Molina, J. Hammer, J. Widom
1.743	An Effective Hash Based Algorithm for Mining Association Rules. J. S. Park, M. Chen, P. S. Yu
1.677	Incremental Maintenance of Views with Duplicates. T. Griffin, L. Libkin
1.519	Balancing Histogram Optimality and Practicality for Query Result Size Estimation. Y. E. Ioannidis, V. Poosala
0.881	FastMap: A Fast Algorithm for Indexing, Data-Mining and Visualization of Traditional and Multimedia Datasets. C. Faloutsos, K. Lin
0.7384	Broadcast Disks: Data Management for Asymmetric Communications Environments. S. Acharya, R. Alonso, M. J. Franklin, S. B. Zdonik
0.7267	Adapting Materialized Views after Redefinitions. A. Gupta, I. S. Mumick, K. A. Ross
0.723	Efficient Maintenance of Materialized Mediated Views. J. J. Lu, G. Moerkotte, J. Schü, V. S. Subrahmanian
0.491	Topological Relations in the World of Minimum Bounding Rectangles: A Study with R-trees. D. Papadias, Y. Theodoridis, T. K. Sellis, M. J. Egenhofer

Table 34: Rank of DBLP components belonging to sigmod 1995 by EPS

Score	Title
1.485	View Maintenance in a Warehousing Environment. Y. Zhuge, H. Garcia-Molina, J. Hammer, J. Widom
1.377	Nearest Neighbor Queries. N. Roussopoulos, S. Kelley, F. Vincent
1.264	An Effective Hash Based Algorithm for Mining Association Rules. J. S. Park, M. Chen, P. S. Yu
1.009	Incremental Maintenance of Views with Duplicates. T. Griffin, L. Libkin
0.7907	Balancing Histogram Optimality and Practicality for Query Result Size Estimation. Y. E. Ioannidis, V. Poosala
0.7243	Adapting Materialized Views after Redefinitions. A. Gupta, I. S. Mumick, K. A. Ross
0.5426	FastMap: A Fast Algorithm for Indexing, Data-Mining and Visualization of Traditional and Multimedia Datasets. C. Faloutsos, K. Lin
0.5329	Efficient Maintenance of Materialized Mediated Views. J. J. Lu, G. Moerkotte, J. Schü, V. S. Subrahmanian
0.4068	Broadcast Disks: Data Management for Asymmetric Communications Environments. S. Acharya, R. Alonso, M. J. Franklin, S. B. Zdonik
0.2959	Topological Relations in the World of Minimum Bounding Rectangles: A Study with R-trees. D. Papadias, Y. Theodoridis, T. K. Sellis, M. J. Egenhofer

Table 35: Rank of DBLP components belonging to sigmod 1995 by BEPS

Score	Title
4.082	An Effective Hash Based Algorithm for Mining Association Rules. J. S. Park, M. Chen, P. S. Yu
3.905	View Maintenance in a Warehousing Environment. Y. Zhuge, H. Garcia-Molina, J. Hammer, J. Widom
2.791	Nearest Neighbor Queries. N. Roussopoulos, S. Kelley, F. Vincent
2.596	Incremental Maintenance of Views with Duplicates. T. Griffin, L. Libkin
2.487	Broadcast Disks: Data Management for Asymmetric Communications Environments. S. Acharya, R. Alonso, M. J. Franklin, S. B. Zdonik
2.263	A Critique of ANSI SQL Isolation Levels. H. Berenson, P. A. Bernstein, J. Gray, J. Melton, E. J. O'Neil, P. E. O'Neil
2.167	Balancing Histogram Optimality and Practicality for Query Result Size Estimation. Y. E. Ioannidis, V. Poosala
1.571	Fault Tolerant Design of Multimedia Servers. S. Berson, L. Golubchik, R. R. Muntz
1.386	Adapting Materialized Views after Redefinitions. A. Gupta, I. S. Mumick, K. A. Ross
1.38	FastMap: A Fast Algorithm for Indexing, Data-Mining and Visualization of Traditional and Multimedia Datasets. C. Faloutsos, K. Lin

Table 36: Rank of DBLP components belonging to sigmod 1995 by SCEAS_B1

Score	Title
4.657	An Effective Hash Based Algorithm for Mining Association Rules. J. S. Park, M. Chen, P. S. Yu
4.494	View Maintenance in a Warehousing Environment. Y. Zhuge, H. Garcia-Molina, J. Hammer, J. Widom
3.349	Nearest Neighbor Queries. N. Roussopoulos, S. Kelley, F. Vincent
3.152	Incremental Maintenance of Views with Duplicates. T. Griffin, L. Libkin
3.071	Broadcast Disks: Data Management for Asymmetric Communications Environments. S. Acharya, R. Alonso, M. J. Franklin, S. B. Zdonik
2.781	A Critique of ANSI SQL Isolation Levels. H. Berenson, P. A. Bernstein, J. Gray, J. Melton, E. J. O'Neil, P. E. O'Neil
2.682	Balancing Histogram Optimality and Practicality for Query Result Size Estimation. Y. E. Ioannidis, V. Poosala
2.107	Fault Tolerant Design of Multimedia Servers. S. Berson, L. Golubchik, R. R. Muntz
1.873	FastMap: A Fast Algorithm for Indexing, Data-Mining and Visualization of Traditional and Multimedia Datasets. C. Faloutsos, K. Lin
1.859	Adapting Materialized Views after Redefinitions. A. Gupta, I. S. Mumick, K. A. Ross

Table 37: Rank of DBLP components belonging to vldb 1995 by CC

Score	Title
48	Aggregate-Query Processing in Data Warehousing Environments. A. Gupta, V. Harinarayan, D. Quass
47	Discovery of Multiple-Level Association Rules from Large Databases. J. Han, Y. Fu
39	W3QS: A Query System for the World-Wide Web. D. Konopnicki, O. Shmueli
38	An Efficient Algorithm for Mining Association Rules in Large Databases. A. Savasere, E. Omiecinski, S. B. Navathe
37	Mining Generalized Association Rules. R. Srikant, R. Agrawal
28	Generalized Search Trees for Database Systems. J. M. Hellerstein, J. F. Naughton, A. Pfeffer
24	Sampling-Based Estimation of the Number of Distinct Values of an Attribute. P. J. Haas, J. F. Naughton, S. Seshadri, L. Stokes
20	Fast Similarity Search in the Presence of Noise, Scaling, and Translation in Time-Series Databases. R. Agrawal, K. Lin, H. S. Sawhney, K. Shim
19	Eager Aggregation and Lazy Aggregation. W. P. Yan, P. Larson
16	Estimating the Selectivity of Spatial Queries Using the 'Correlation' Fractal Dimension. A. Belussi, C. Faloutsos

Table 38: Rank of DBLP components belonging to vldb 1995 by BCC

Score	Title
5.877	Discovery of Multiple-Level Association Rules from Large Databases. J. Han, Y. Fu
5.511	W3QS: A Query System for the World-Wide Web. D. Konopnicki, O. Shmueli
5.014	Aggregate-Query Processing in Data Warehousing Environments. A. Gupta, V. Harinarayan, D. Quass
3.919	An Efficient Algorithm for Mining Association Rules in Large Databases. A. Savasere, E. Omiecinski, S. B. Navathe
3.703	Mining Generalized Association Rules. R. Srikant, R. Agrawal
3.086	Generalized Search Trees for Database Systems. J. M. Hellerstein, J. F. Naughton, A. Pfeffer
2.603	Sampling-Based Estimation of the Number of Distinct Values of an Attribute. P. J. Haas, J. F. Naughton, S. Seshadri, L. Stokes
2.12	Redo Recovery after System Crashes. D. B. Lomet, M. R. Tuttle
2.055	Fast Similarity Search in the Presence of Noise, Scaling, and Translation in Time-Series Databases. R. Agrawal, K. Lin, H. S. Sawhney, K. Shim
1.834	Generalizing GlOSS to Vector-Space Databases and Broker Hierarchies. L. Gravano, H. Garcia-Molina

Table 39: Rank of DBLP components belonging to vldb 1995 by PR

Score	Title
3.756	Aggregate-Query Processing in Data Warehousing Environments. A. Gupta, V. Harinarayan, D. Quass
3.64	Sampling-Based Estimation of the Number of Distinct Values of an Attribute. P. J. Haas, J. F. Naughton, S. Seshadri, L. Stokes
3.014	Discovery of Multiple-Level Association Rules from Large Databases. J. Han, Y. Fu
2.91	W3QS: A Query System for the World-Wide Web. D. Konopnicki, O. Shmueli
2.254	Mining Generalized Association Rules. R. Srikant, R. Agrawal
2.138	An Efficient Algorithm for Mining Association Rules in Large Databases. A. Savasere, E. Omiecinski, S. B. Navathe
1.629	Generalized Search Trees for Database Systems. J. M. Hellerstein, J. F. Naughton, A. Pfeffer
1.371	A Data Transformation System for Biological Data Sources. P. Buneman, S. B. Davidson, K. Hart, G. C. Overton, L. Wong
1.305	Fast Similarity Search in the Presence of Noise, Scaling, and Translation in Time-Series Databases. R. Agrawal, K. Lin, H. S. Sawhney, K. Shim
1.302	Eager Aggregation and Lazy Aggregation. W. P. Yan, P. Larson

Table 40: Rank of DBLP components belonging to vldb 1995 by HA

Score	Title
1.163	Generalized Search Trees for Database Systems. J. M. Hellerstein, J. F. Naughton, A. Pfeffer
1.013	Aggregate-Query Processing in Data Warehousing Environments. A. Gupta, V. Harinarayan, D. Quass
0.7487	Estimating the Selectivity of Spatial Queries Using the 'Correlation' Fractal Dimension. A. Belussi, C. Faloutsos
0.7214	Sampling-Based Estimation of the Number of Distinct Values of an Attribute. P. J. Haas, J. F. Naughton, S. Seshadri, L. Stokes
0.6064	Eager Aggregation and Lazy Aggregation. W. P. Yan, P. Larson
0.5647	A Data Transformation System for Biological Data Sources. P. Buneman, S. B. Davidson, K. Hart, G. C. Overton, L. Wong
0.4842	W3QS: A Query System for the World-Wide Web. D. Konopnicki, O. Shmueli
0.4497	Fast Similarity Search in the Presence of Noise, Scaling, and Translation in Time-Series Databases. R. Agrawal, K. Lin, H. S. Sawhney, K. Shim
0.4336	The hBP-tree: A Modified hB-tree Supporting Concurrency, Recovery and Node Consolidation. G. Evangelidis, D. B. Lomet, B. Salzberg
0.409	Near Neighbor Search in Large Metric Spaces. S. Brin

Table 41: Rank of DBLP components belonging to vldb 1995 by P

Score	Title
5.575e-104	Aggregate-Query Processing in Data Warehousing Environments. A. Gupta, V. Harinarayan, D. Quass
3.817e-144	Sampling-Based Estimation of the Number of Distinct Values of an Attribute. P. J. Haas, J. F. Naughton, S. Seshadri, L. Stokes
3.022e-157	W3QS: A Query System for the World-Wide Web. D. Konopnicki, O. Shmueli
2.885e-157	Eager Aggregation and Lazy Aggregation. W. P. Yan, P. Larson
1.411e-158	A Data Transformation System for Biological Data Sources. P. Buneman, S. B. Davidson, K. Hart, G. C. Overton, L. Wong
3.176e-159	Type Classification of Semi-Structured Documents. M. Tresch, N. Palmer, A. Luniewski
7.685e-161	Query Processing in Tertiary Memory Databases. S. Sarawagi
3.113e-161	Benchmarking Spatial Join Operations with Spatial Output. E. G. Hoel, H. Samet
2.286e-161	DB2 Query Parallelism: Staging and Implementation. Y. Wang
5.903e-167	Estimating the Selectivity of Spatial Queries Using the 'Correlation' Fractal Dimension. A. Belussi, C. Faloutsos

Table 42: Rank of DBLP components belonging to vldb 1995 by SA

Score	Title
7.947	Aggregate-Query Processing in Data Warehousing Environments. A. Gupta, V. Harinarayan, D. Quass
7.781	Discovery of Multiple-Level Association Rules from Large Databases. J. Han, Y. Fu
6.457	W3QS: A Query System for the World-Wide Web. D. Konopnicki, O. Shmueli
6.291	An Efficient Algorithm for Mining Association Rules in Large Databases. A. Savasere, E. Omiecinski, S. B. Navathe
6.126	Mining Generalized Association Rules. R. Srikant, R. Agrawal
4.636	Generalized Search Trees for Database Systems. J. M. Hellerstein, J. F. Naughton, A. Pfeffer
3.974	Sampling-Based Estimation of the Number of Distinct Values of an Attribute. P. J. Haas, J. F. Naughton, S. Seshadri, L. Stokes
3.311	Fast Similarity Search in the Presence of Noise, Scaling, and Translation in Time-Series Databases. R. Agrawal, K. Lin, H. S. Sawhney, K. Shim
3.146	Eager Aggregation and Lazy Aggregation. W. P. Yan, P. Larson
2.649	Estimating the Selectivity of Spatial Queries Using the 'Correlation' Fractal Dimension. A. Belussi, C. Faloutsos

Table 43: Rank of DBLP components belonging to vldb 1995 by BHA

Score	Title
0.1373	W3QS: A Query System for the World-Wide Web. D. Konopnicki, O. Shmueli
0.06079	Generalized Search Trees for Database Systems. J. M. Hellerstein, J. F. Naughton, A. Pfeffer
0.05258	Estimating the Selectivity of Spatial Queries Using the 'Correlation' Fractal Dimension. A. Belussi, C. Faloutsos
0.04252	High-Concurrency Locking in R-Trees. M. Kornacker, D. Banks
0.0419	Sampling-Based Estimation of the Number of Distinct Values of an Attribute. P. J. Haas, J. F. Naughton, S. Seshadri, L. Stokes
0.03097	Aggregate-Query Processing in Data Warehousing Environments. A. Gupta, V. Harinarayan, D. Quass
0.02972	A Data Transformation System for Biological Data Sources. P. Buneman, S. B. Davidson, K. Hart, G. C. Overton, L. Wong
0.02787	Type Classification of Semi-Structured Documents. M. Tresch, N. Palmer, A. Luniewski
0.01754	Eager Aggregation and Lazy Aggregation. W. P. Yan, P. Larson
0.0174	Fast Similarity Search in the Presence of Noise, Scaling, and Translation in Time-Series Databases. R. Agrawal, K. Lin, H. S. Sawhney, K. Shim

Table 44: Rank of DBLP components belonging to vldb 1995 by BSA

Score	Title
2.474	Aggregate-Query Processing in Data Warehousing Environments. A. Gupta, V. Harinarayan, D. Quass
1.73	Sampling-Based Estimation of the Number of Distinct Values of an Attribute. P. J. Haas, J. F. Naughton, S. Seshadri, L. Stokes
1.599	Discovery of Multiple-Level Association Rules from Large Databases. J. Han, Y. Fu
1.551	W3QS: A Query System for the World-Wide Web. D. Konopnicki, O. Shmueli
1.373	Generalized Search Trees for Database Systems. J. M. Hellerstein, J. F. Naughton, A. Pfeffer
1.279	Mining Generalized Association Rules. R. Srikant, R. Agrawal
1.232	An Efficient Algorithm for Mining Association Rules in Large Databases. A. Savasere, E. Omiecinski, S. B. Navathe
0.9245	Eager Aggregation and Lazy Aggregation. W. P. Yan, P. Larson
0.8265	Fast Similarity Search in the Presence of Noise, Scaling, and Translation in Time-Series Databases. R. Agrawal, K. Lin, H. S. Sawhney, K. Shim
0.7978	Estimating the Selectivity of Spatial Queries Using the 'Correlation' Fractal Dimension. A. Belussi, C. Faloutsos

Table 45: Rank of DBLP components belonging to vldb 1995 by EPS

Score	Title
1.585	Aggregate-Query Processing in Data Warehousing Environments. A. Gupta, V. Harinarayan, D. Quass
1.386	W3QS: A Query System for the World-Wide Web. D. Konopnicki, O. Shmueli
1.127	Sampling-Based Estimation of the Number of Distinct Values of an Attribute. P. J. Haas, J. F. Naughton, S. Seshadri, L. Stokes
1.089	Discovery of Multiple-Level Association Rules from Large Databases. J. Han, Y. Fu
0.9529	Mining Generalized Association Rules. R. Srikant, R. Agrawal
0.8986	An Efficient Algorithm for Mining Association Rules in Large Databases. A. Savasere, E. Omiecinski, S. B. Navathe
0.6357	Generalized Search Trees for Database Systems. J. M. Hellerstein, J. F. Naughton, A. Pfeffer
0.5317	Eager Aggregation and Lazy Aggregation. W. P. Yan, P. Larson
0.5214	A Data Transformation System for Biological Data Sources. P. Buneman, S. B. Davidson, K. Hart, G. C. Overton, L. Wong
0.4555	Fast Similarity Search in the Presence of Noise, Scaling, and Translation in Time-Series Databases. R. Agrawal, K. Lin, H. S. Sawhney, K. Shim

Table 46: Rank of DBLP components belonging to vldb 1995 by BEPS

Score	Title
4.83	Aggregate-Query Processing in Data Warehousing Environments. A. Gupta, V. Harinarayan, D. Quass
4.747	Discovery of Multiple-Level Association Rules from Large Databases. J. Han, Y. Fu
4.451	W3QS: A Query System for the World-Wide Web. D. Konopnicki, O. Shmueli
3.464	Sampling-Based Estimation of the Number of Distinct Values of an Attribute. P. J. Haas, J. F. Naughton, S. Seshadri, L. Stokes
3.131	Mining Generalized Association Rules. R. Srikant, R. Agrawal
3.124	An Efficient Algorithm for Mining Association Rules in Large Databases. A. Savasere, E. Omiecinski, S. B. Navathe
2.338	Generalized Search Trees for Database Systems. J. M. Hellerstein, J. F. Naughton, A. Pfeffer
1.597	Fast Similarity Search in the Presence of Noise, Scaling, and Translation in Time-Series Databases. R. Agrawal, K. Lin, H. S. Sawhney, K. Shim
1.551	Redo Recovery after System Crashes. D. B. Lomet, M. R. Tuttle
1.466	Generalizing GlOSS to Vector-Space Databases and Broker Hierarchies. L. Gravano, H. Garcia-Molina

Table 47: Rank of DBLP components belonging to vldb 1995 by SCEAS_B1

Score	Title
5.416	Discovery of Multiple-Level Association Rules from Large Databases. J. Han, Y. Fu
5.37	Aggregate-Query Processing in Data Warehousing Environments. A. Gupta, V. Harinarayan, D. Quass
5.1	W3QS: A Query System for the World-Wide Web. D. Konopnicki, O. Shmueli
3.825	Sampling-Based Estimation of the Number of Distinct Values of an Attribute. P. J. Haas, J. F. Naughton, S. Seshadri, L. Stokes
3.712	An Efficient Algorithm for Mining Association Rules in Large Databases. A. Savasere, E. Omiecinski, S. B. Navathe
3.692	Mining Generalized Association Rules. R. Srikant, R. Agrawal
2.903	Generalized Search Trees for Database Systems. J. M. Hellerstein, J. F. Naughton, A. Pfeffer
2.108	Fast Similarity Search in the Presence of Noise, Scaling, and Translation in Time-Series Databases. R. Agrawal, K. Lin, H. S. Sawhney, K. Shim
2.075	Redo Recovery after System Crashes. D. B. Lomet, M. R. Tuttle
1.964	Generalizing GlOSS to Vector-Space Databases and Broker Hierarchies. L. Gravano, H. Garcia-Molina

Appendix B: Ranking Authors Full Results

Here we present the full version of Table 21 to depict the rank positions of un-awarded authors. It is obvious that in the future, the authors with the higher possibility to be awarded are the ones that are highly ranked (especially with the `SCEAS` method) and are not yet awarded. This might be helpful to short-list candidate authors for awarding during the next few years. A final noticeable remark is that Edgar F. Codd remains at the 1st position by almost all ranking methods.

Table 48: Rank of authors by their best 25 publications (part a).

	`CC`	`BCC`	`PageRank`	`SALSA_A`	`BSALSA_A`
1	**D.J. DeWitt**	**Jim Gray**	E.F. Codd	**D.J. DeWitt**	E. F. Codd
2	**M. Stonebraker**	E.F. Codd	**Jim Gray**	**M. Stonebraker**	**M. Stonebraker**
3	**Jim Gray**	**M. Stonebraker**	R.A. Lorie	**Jim Gray**	R.A. Lorie
4	R.A. Lorie	**D.J. DeWitt**	**M. Stonebraker**	R.A. Lorie	**Jim Gray**
5	J.D. Ullman	R.A. Lorie	**D.D. Chamberlin**	J.D. Ullman	**P.A. Bernstein**
6	**P.A. Bernstein**	J.D. Ullman	**P.A. Bernstein**	**P.A. Bernstein**	J.D. Ullman
7	E.F. Codd	**P.A. Bernstein**	J.D. Ullman	E.F. Codd	**D.D. Chamberlin**
8	Won Kim	P.P. Chen	**D.J. DeWitt**	Won Kim	**D.J. DeWitt**
9	**D. Maier**	**D.D. Chamberlin**	E. Wong	**D. Maier**	**R. Fagin**
10	Y. Sagiv	**D. Maier**	**R. Fagin**	Y. Sagiv	E. Wong
11	F. Bancilhon	Won Kim	C. Beeri	F. Bancilhon	C. Beeri
12	C. Beeri	**R. Agrawal**	P.P. Chen	C. Beeri	**D. Maier**
13	**S. Abiteboul**	C. Beeri	**D. Maier**	**S. Abiteboul**	Y. Sagiv
14	**D.D. Chamberlin**	F. Bancilhon	Won Kim	**D.D. Chamberlin**	P. P. Chen
15	**R. Agrawal**	U. Dayal	Y. Sagiv	**R. Agrawal**	Won Kim
16	M. J. Carey	**H. Garcia-Molina**	**R. Agrawal**	M.J. Carey	N. Goodman
17	**R. Fagin**	Y. Sagiv	U. Dayal	**R. Fagin**	F. Bancilhon
18	U. Dayal	**R. Fagin**	F. Bancilhon	U. Dayal	U. Dayal
19	R. Ramakrishnan	**S. Abiteboul**	N. Goodman	R. Ramakrishnan	S.B. Yao
20	N. Goodman	M.J. Carey	**H. Garcia-Molina**	N. Goodman	M.J. Carey
21	**H. Garcia-Molina**	E. Wong	M.J. Carey	**H. Garcia-Molina**	**S. Abiteboul**
22	J. Widom	R. Ramakrishnan	B.G. Lindsay	J. Widom	A.V. Aho
23	E. Wong	J. Widom	**S. Abiteboul**	E. Wong	B.G. Lindsay
24	H. Pirahesh	N. Goodman	**R. Bayer**	H. Pirahesh	R. Ramakrishnan
25	P.P. Chen	H. Pirahesh	H. Pirahesh	P.P. Chen	**P.G. Selinger**
26	C. Faloutsos	J.F. Naughton	R. Ramakrishnan	C. Faloutsos	A.O. Mendelzon
27	B.G. Lindsay	B.G. Lindsay	S.B. Yao	B.G. Lindsay	C. Zaniolo
28	R. Hull	T. Imielinski	**P.G. Selinger**	R. Hull	**R. Agrawal**
29	**C. Mohan**	C. Faloutsos	T. Imielinski	**C. Mohan**	H. Pirahesh
30	N. Roussopoulos	R.H. Katz	J. Widom	N. Roussopoulos	**R. Bayer**
31	A.O. Mendelzon	S.B. Navathe	N. Roussopoulos	A.O. Mendelzon	N. Roussopoulos
32	J.F. Naughton	**C. Mohan**	A.V. Aho	J.F. Naughton	R.H. Katz
33	G. Graefe	R. Hull	R.H. Katz	G. Graefe	R. Hull
34	P. Buneman	N. Roussopoulos	J.F. Naughton	P. Buneman	C. Faloutsos
35	R.H. Katz	A. Shoshani	C. Zaniolo	R.H. Katz	G. M. Lohman
36	S.B. Navathe	G. Graefe	D. McLeod	S.B. Navathe	**H. Garcia-Molina**
37	C. Zaniolo	A.O. Mendelzon	C. Faloutsos	C. Zaniolo	D. McLeod
38	**P. G. Selinger**	H. Kriegel	G.M. Lohman	**P. G. Selinger**	P. Larson
39	M.Y. Vardi	**P.G. Selinger**	S.B. Navathe	M.Y. Vardi	S.B. Navathe
40	H. Kriegel	A.P. Sheth	**C. Mohan**	H. Kriegel	**C. Mohan**
missed	73. R. Bayer	51. R. Bayer		74. R. Bayer	

Table 49: Rank of authors by their best 25 publications (part b).

	PS	BPS	EPS	BEPS	SCEAS
1	E. F. Codd	E. F. Codd	E. F. Codd	**Jim Gray**	**Jim Gray**
2	**M. Stonebraker**	**Jim Gray**	**M. Stonebraker**	E. F. Codd	E. F. Codd
3	**D.D. Chamberlin**	R.A. Lorie	R.A. Lorie	R.A. Lorie	R.A. Lorie
4	R. A. Lorie	**D.D. Chamberlin**	**D.D. Chamberlin**	**M. Stonebraker**	**M. Stonebraker**
5	**P.A. Bernstein**	**M. Stonebraker**	**Jim Gray**	**D.J. DeWitt**	**D.J. DeWitt**
6	**Jim Gray**	**P.A. Bernstein**	**P.A. Bernstein**	J.D. Ullman	J.D. Ullman
7	E. Wong	J.D. Ullman	E. Wong	**D.D. Chamberlin**	**D.D. Chamberlin**
8	J.D. Ullman	**D.J. DeWitt**	**D.J. DeWitt**	**P.A. Bernstein**	**P.A. Bernstein**
9	**R. Fagin**	E. Wong	J.D. Ullman	P.P. Chen	P.P. Chen
10	C. Beeri	**R. Fagin**	C. Beeri	Won Kim	Won Kim
11	**D.J. DeWitt**	C. Beeri	**R. Fagin**	**D. Maier**	**D. Maier**
12	N. Goodman	P.P. Chen	N. Goodman	**R. Agrawal**	**R. Agrawal**
13	Y. Sagiv	Y. Sagiv	**D. Maier**	C. Beeri	C. Beeri
14	**D. Maier**	**D. Maier**	Y. Sagiv	**R. Fagin**	**R. Fagin**
15	S.B. Yao	Won Kim	Won Kim	E. Wong	Y. Sagiv
16	D. McLeod	U. Dayal	S.B. Yao	Y. Sagiv	F. Bancilhon
17	**R. Bayer**	**R. Agrawal**	F. Bancilhon	U. Dayal	E. Wong
18	A. V. Aho	F. Bancilhon	**P.G. Selinger**	F. Bancilhon	U. Dayal
19	**P. G. Selinger**	N. Goodman	U. Dayal	**H. Garcia-Molina**	**H. Garcia-Molina**
20	F. Bancilhon	**R. Bayer**	D. McLeod	M.J. Carey	M.J. Carey
21	Won Kim	**H. Garcia-Molina**	B.G. Lindsay	N. Goodman	**S. Abiteboul**
22	U. Dayal	M.J. Carey	**S. Abiteboul**	**S. Abiteboul**	N. Goodman
23	D. Tsichritzis	B.G. Lindsay	A.V. Aho	R. Ramakrishnan	R. Ramakrishnan
24	P. P. Chen	**S. Abiteboul**	**R. Bayer**	H. Pirahesh	J. Widom
25	H. Schek	S.B. Yao	M.J. Carey	J. Widom	H. Pirahesh
26	**S. Abiteboul**	**P.G. Selinger**	C. Zaniolo	B.G. Lindsay	B.G. Lindsay
27	H.A. Schmid	H. Pirahesh	P.P. Chen	J.F. Naughton	J.F. Naughton
28	B.G. Lindsay	A.V. Aho	H. Schek	T. Imielinski	T. Imielinski
29	C. Zaniolo	N. Roussopoulos	N. Roussopoulos	C. Faloutsos	C. Faloutsos
30	V.Y. Lum	R. Ramakrishnan	G.M. Lohman	R.H. Katz	R.H. Katz
31	M. Schkolnick	T. Imielinski	R.H. Katz	N. Roussopoulos	N. Roussopoulos
32	M.J. Carey	J. Widom	P. Buneman	**P.G. Selinger**	S.B. Navathe
33	N. Roussopoulos	R.H. Katz	R. Hull	S.B. Navathe	**C. Mohan**
34	G.M. Lohman	C. Zaniolo	**C. Mohan**	**C. Mohan**	**P.G. Selinger**
35	P. Buneman	D. McLeod	L.A. Rowe	S.B. Yao	S.B. Yao
36	R. Hull	D. Tsichritzis	M. Schkolnick	**R. Bayer**	A. Shoshani
37	L.A. Rowe	G.M. Lohman	H. Pirahesh	A. Shoshani	G. Graefe
38	R.H. Katz	J.F. Naughton	G. Graefe	G. Graefe	R. Hull
39	A.O. Mendelzon	C. Faloutsos	V.Y. Lum	A.O. Mendelzon	A.O. Mendelzon
40	S. Y.W. Su	S.B. Navathe	G.P. Copeland	R. Hull	**R. Bayer**
missed	41. C. Mohan 69. H. Garcia-Molina 70. R. Agrawal	43. C. Mohan	45. R. Agrawal 50. H. Garcia-Molina		