Automatically Ranking Scientific Conferences using Digital Libraries

Antonis Sidiropoulos Yannis Manolopoulos
emails: {asidirop,manolopo}@delab.csd.auth.gr

Department of Informatics
Aristotle University
Thessaloniki, 54124 Greece

Abstract. Citation analysis is performed in order to evaluate authors and scientific collections, such as journals and conference proceedings. Currently, two major systems exist that perform citation analysis: Science Citation Index (SCI) by the Institute for Scientific Information (ISI) and CiteSeer by the NEC Research Institute. The SCI, mostly a manual system up until recently, is based on the notion of the ISI Impact Factor, which has been used extensively for citation analysis purposes. On the other hand the CiteSeer system is an automatically built digital library using agents technology, also based on the notion of ISI Impact Factor. In this paper, we investigate new alternative notions besides the ISI Impact Factor, in order to provide a novel approach aiming at ranking scientific collections. Furthermore, we present a web-based system that has been built by extracting data from the Databases and Logic Programming (DBLP) website of the University of Trier. Our system, by using the new citation metrics, emerges as a useful tool for ranking scientific collections. In this respect, some first remarks are presented, e.g. on ranking conferences related to databases.

1 Introduction

Currently, two major systems exist that perform citation analysis: Science Citation Index (SCI) by Institute for Scientific Information (ISI) [1] and CiteSeer by the NEC Research Institute [2]. The idea of performing citation analysis has evolved in early 60's. Then, in 1972 it started to be used by the SCI for the evaluation of journals spanning many scientific fields, computer science included. The *ISI Impact Factor* [9, 8] was the main metric used by SCI for journal evaluation and ranking. On the other hand, the CiteSeer system is a modern system and constructs the citation graph from publications acquired from the web [16]. CiteSeer is also based on the ISI Impact Factor for ranking conferences and journals [10].

Citation analysis is based on the notion of *citation graphs*, which are graphs representing papers as nodes, whereas an edge from node x to node y represents a citation from paper x to paper y. Citation graphs can be used to derive useful statistical information related to evaluating and ranking several entities, such as authors, publications in scientific conferences and journals, as well as conferences and journals as scientific collections.

In particular, citation graph analysis is similar to web-graph analysis. Notable is the PageRank algorithm by Brin and Page [6], which is used by the Google search engine. This algorithm computes a score for a page as a summary of the fractions of the scores of the referrer pages. Thus, it ranks the web-pages returned to the user according to their relevance to the user query. It has been derived that the statistical distribution of the PageRank metric follows the familiar inverse polynomial law reported for Webpage degrees [7]. Recently, the PageRank formula has been further analyzed [18].

Except ranking, other operations can also be performed on citation graphs by using graph-theoretic and data-mining techniques [5]. For example, assuming a set of scientific collections, related books, conferences, journals or/and technical reports – with regard to the specific area – can be categorized by using clustering. In an analogous manner, authors can be grouped in clusters in order to find and establish their communities, i.e. authors that co-operate and cite each other and, in an analogous manner, to find *hubs* and *authorities*, i.e. clusters of authors that mostly cite and mostly get cited, respectively. Two important works on this area are the Kleinberg's HITS (Hyperlink Induced Topic Search) algorithm [14, 13], which computes a weighted score for the above notions. A further study of the Web as a graph and the hubs/authorities notions has appeared in [17].

2 Major Systems for Citation Analysis

As mentioned, currently there exist two major systems that perform citation analysis: SCI and CiteSeer. Here, we will examine closer these systems in order to see their "weak" points and motivate the research of this paper.

CiteSeer is a modern system and constructs the citation graph from publications acquired from the web [16]. More specifically, it is an autonomous system that collects computer science papers by crawling the web. Then, from the format each paper is stored (i.e. postscript or pdf), it detects and exports the bibliographic information (e.g. title, authors, etc.) as well as the included citations to construct the underlying citation graph [15].

However, CiteSeer does not really focus on conference or journal evaluation/ranking. There exist only one ranking in CiteSeer [3] which (a) includes data from DBLP only, (b) mixes-up conferences and journals, (c) groups together various scientific areas, and (d) it is based on the notion of ISI Impact Factor

On the other hand, SCI has served the whole academic community for several decades by providing useful information, in the lack of anything better. However, nowadays the system disadvantages and limits are apparent. For example, the main disadvantages of the SCI system are:

1. Each scientific field is divided in certain areas, which remain static over the years and do not reflect the scientific evolution and, in particular, the dramatic evolution of computer science.
2. In each area, only a set of journals is selected for journal evaluation. Thus, the representative value of the selected journals is questionable.

3. Although any such set is dynamic and updated periodically, this update is done in a subjective way, which might also trigger questions about when, why how and by whom.
4. In some cases, irrelevant journals (e.g. technical vs. popular) are grouped in a certain area leading to erroneous results.
5. Scientific conferences, books and technical reports are not taken into consideration.
6. It is/was manually constructed and, therefore is an expensive system to built and maintain.
7. It is not for free neither for libraries nor for individuals.

ISI Impact Factor, although for a long period has played an important role in evaluating journals (and, subsequently academic authors), is quite rigid and cannot be used to perform deeper qualitative analysis. The present work is motivated from the latter point. One could argue against the "flat" nature of the ISI Impact Factor. For example,

- Is it fair to count a citation from Professor "Well-known" as equivalent to a citation from Professor "Unknown"?
- Is it fair to count a citation from Journal "The-top" as equivalent to a citation from Journal "The-bottom"?
- Or, finally, is it fair to count a citation from Paper "The-best" as equivalent to a citation from Paper "The-worst"?

From these simple questions, it is apparent that it is necessary to embed some kind of weighting to answer such questions. In this work we will investigate some new ideas for ranking scientific collections and we will try to put the task of citation analysis and journal or conference evaluation in a more generalized perspective.

3 The SCEAS System

The DBLP website of the University of Trier [4] is a rich digital library focusing in the areas of Data Bases and Logic Programming. More specifically, as of April 2003, the DBLP website contains bibliographic data about 250.000 authors, 1.300 conferences, 310 journals and 370.000 papers, articles or books, with links to personal pages, research groups, publishing houses, etc.

Based on the DBLP database records, we built a system called *Scientific Collection Evaluator by using Advanced Scoring* (SCEAS). Our system imports DBLP XML records into a MySQL database system. We have used the specific software since it is light, fast and fits our needs as the transactions sent to the server are mainly reads rather than updates. The SCEAS system is available on the Internet[1], and thus the user can easily access it, post queries, get answers, and extract useful information.

In our model, the main entities are:

Publication which could be article, in-proceedings, etc.
Collection such as conference, book, journal etc.

[1] http://delab.csd.auth.gr/sceas

Person which could be author or editor.

Each publication belongs to a collection (or more, e.g. conference publications belong to a conference and to a proceedings collection). A collection may be a part of another one, e.g. VLDB'97 is a collection and it is part of the VLDB collection. Persons can be related to publications as authors or to collections as editors (e.g. for proceedings). Finally, publications can be related to each other with the "citation" relation.

Based on the DBLP database, we built the citation graph of the collection, which includes journal as well as conference publications. Using this graph we derived two Collection Citation Graphs, the Conference Citation Graph, and the Journal Citation Graph. In the same way, any other type of semantic grouping of the publications can be used to derive analogous citation graphs (e.g. Book Citation Graph).

Scientific collection evaluation, and conference evaluation in particular, being our first concern, we tried to investigate alternative ways for such a task. The basic idea for the ranking, is that not all the citations should have the same weight. For example, the weight should depend on two factors: (a) the quality of the conference, where a citation to another conference is made, and (b) the scientific domain of the conferences, e.g. if they belong in the same domain. Thus, there exist two tasks that should be done. First, we must specify the scientific domains and then perform the ranking.

In particular, we have performed the following tasks:

1. Cluster the conferences according to Topics, based on the conference citation graph.
2. Cleansing, since the data of many conferences in the DBLP database are not complete. Therefore, we try to exclude the collection subset that inserts noise into our algorithms.
3. Finally, ranking each conference cluster separately. We have performed ranking by taking into consideration the whole lifetime of all conferences and every specific year of each conference. The former does not produce useful results form the statistics point of view, since there exist a lot of factors affecting ranking. Thus, we focused in the latter case and for every distinct year we produced:
   - rankings using Plain Scoring,
   - rankings using Weighted Scoring, and
   - rankings using the Inverted Impact Scoring
   - rankings using Weighted Inverted Impact Scoring

All these new notions introduced above will be explained in the sequel. Please, notice that ranking is performed by using several algorithms, which will be presented in the sequel as well.

3.1 Clustering Conferences According to Topics

Based on the conference citation graph, first we performed a clustering operation. As a utility for clustering the conferences we used the hMetis [11], a hyper-graph partitioning tool for large hyper-graphs. hMetis has been successfully used in applications related to VLSI circuits, data mining and numerical analysis.

In order to help hMetis for better results, we perform some preprocessing on the graph by predefining 4 clusters based on keyword matches in the conference titles:

– Cluster 1: Databases.
– Cluster 2: Logic Programming.
– Cluster 3: Networks and Distributed Systems.
– Cluster 4: Operating Systems, Software Engineering, Compilers and Languages.

After defining the above 4 clusters with some conference members in them, we feed this predefined partitions to hMetis to continue with the unclustered conferences.

3.2 Cleansing the Clusters

As mentioned above, the DBLP database is incomplete. For example, for some conferences or journals, 1 or 2 publications are only included (i.e. the most important ones). This would lead our ranking algorithm to produce erroneous results, since the main metric is the average number of citations per publication.

In this step we exclude from the conference set, which will be used for ranking, the conferences that: (a) contain less than 3 publications, or (b) are held only once, or (c) have average number of publications per year less than 0.5. For these conferences we set a flag meaning that they will not be ranked, but we do not delete them from our database. Thus, any citations included in them do count.

3.3 Definition of the Metrics

Here, we introduce the new metrics in order to establish a new perspective for conference and journal evaluation using the citation graph. These metrics are defined as follows:

Plain Score
If C is the set of all the conferences, then the *Plain Score*, S_c, that is the *Score* for conference c, is defined as:

$$S_c = \frac{1}{P_c} \sum_{\forall i \in C} N_{i \to c} \tag{1}$$

where $N_{i \to c}$ is the number of citations made from conference i to conference c, whereas the normalizing factor P_c is the number of publications in conference c.

The rank is computed by ordering the conferences' scores. In case of a tie, the conference with the fewer publications precedes. This score is the simplest metric, and it is basically used as a first approach for ranking. Actually, although this metric carries some information with respect to ranking, it has the disadvantage that conferences with "long" history are more likely to have more citations. Thus, it can only be used to compare and rank a set of conferences that have exactly the same life-time.

Plain Score per Year
Adapting the notion of the Plain Score in order to rank conferences for each distinct year, we introduce the *Plain Score per Year* metric as:

$$SY_{c,y} = \frac{1}{P_{c,y}} \sum_{\forall i \in C} N_{i \to c,y} \tag{2}$$

where $SY_{c,y}$ is the score for conference c in the year y, $N_{i\to c,y}$ is the number of citations from conference i to conference c that was held in year y and $P_{c,y}$ is the number of publications of conference c during the year y. In particular, a more detailed expression that we used for our computations is:

$$N_{i\to c,y} = \sum_{z=y}^{last\_year} N_{i,z\to c,y} \overset{(2)}{\Rightarrow} SY_{c,y} = \frac{1}{P_{c,y}} \sum_{i}^{\forall i\in C} \sum_{z=y}^{last\_year} N_{i,z\to c,y} \tag{3}$$

where $N_{i,z\to c,y}$ is the number of citations made from conference i in year z to conference c held in year y. The variable *last_year* is set to the maximum valid year in our collection (normally the current year). This ranking can be used to compare conferences that were held in the same year.

Weighted Score
Here we introduce the idea of the weighted ranking. This means that the citations do not count the same. Equation 4 shows abstractly how the *Weighted Score* for conference c, defined as WS_c, can be computed.

$$WS_c \;=\; \frac{1}{P_c}\frac{\sum_i^{\forall i\in C} W_i * N_{i\to c}}{\sum_i^{\forall i\in C} W_i} \tag{4}$$

Where W_i is the weight for conference i.
How can we compute the weights? Which conferences are "The-Top" that should have larger weights and which are the "The-Worst" conferences? Here, arises the need for recursive computing. This computation is performed by using the following formula:

$$\begin{array}{lll} WS_{c,l} &= \frac{1}{P_c}\frac{\sum_i W_{i,l-1} * N_{i\to c}}{\sum_i W_{i,l-1}} & l \geq 1 \\ W_{i,0} &= 1 & \forall i \in C \end{array} \tag{5}$$

Initially all the weights are set equal to 1 (at level 0). Thus, we can compute the ranking for the next level, based on the weights computed in the previous one. The ranking we get at level 1, is equivalent with the Plain Score ranking (since we used weights of 1 for all entities). After computing the scores for level 1, we can compute the weights. This is achieved with another clustering algorithm. In Section 3.4 a detailed discussion on the computation of weights can be found.

After computing the weights for level 1, we continue computing the scores for the next levels by applying the same procedure until the ranking remains unchanged. This is our termination condition. The computation is repeated $\forall l \geq 1$ until L, where the ranking for level L is equivalent with that of the level L–1. Alternatively, if while at level L we get the same weights as in level L–1 ($W_{i,L} = W_{i,L-1}\ \forall i \in C$), then it is obvious that the ranking for L+1 would be the same with the one computed at level L. Thus, an alternative stop condition is a "no change" in the computed weights. Then $\forall l \in \{L..\infty\}$ the condition: $W_{i,l} = W_{i,l-1}$ is true $\Rightarrow WS_{i,l+1} = WS_{i,l}$, and we set:

$$WS_c \;=\; WS_{c,\infty} = WS_{c,L}$$

This type of ranking, similarly to the Plain Score, cannot be used for conference evaluation without risk. Despite the refinement of computing the average score per publication by means of the citations' weights, not all conferences have the same life-time, whereas some are only held only once per two or three years. Therefore, conferences with "longer" history are more likely to have more citations. This ranking can be used only for the conferences that have exactly the same life-time.

Weighted Score per Year
By combining WS (Weighted Score) and SY (Plain Score per Year), for $l \geq 1$ we produce the WSY, i.e. the *Weighted Score per Year* metric:

$$WSY_{c,y,l} = \frac{1}{P_{c,y}} \frac{\sum_{i}^{\forall i \in C} \left(W_{i,y,l-1} * N_{i,y \to c,y} + \sum_{z=y+1}^{last\_year} W_{i,z,\infty} * N_{i,z \to c,y} \right)}{\sum_{i}^{\forall i \in C} \left(W_{i,y,l-1} + \sum_{z=y+1}^{last\_year} W_{i,z,\infty} \right)} \quad (6)$$

The same way as above we set:

$$W_{i,z,0} = 1 \quad \forall i \in C \; and \; \forall z \in \{valid\_years\}$$

The computation is made for each year by starting from the last year in reverse order. Therefore, when computing scores for year Y, all the weights are known for years $\{Y + 1...max\_year\}$. For each year, the procedure is repeated $\forall l \geq 1$ until L, where the ranking does not change or the condition $W_{c,y,L} = W_{c,y,L-1}$ is true $\forall c \in C$. Then $W_{c,y,\infty} = W_{c,y,L}$ and we set:

$$WSY_{c,y} = WSY_{c,y,\infty} = WSY_{c,y,L}$$

Inverted Impact Score per Year
Garfield [8] defined the ISI Impact Factor by using the following example for the 1992 year:

$$\mathbb{A} = total \; cites \; in \; 1992$$
$$\mathbb{B} = 1992 \; cites \; to \; articles \; published \; in \; 1990 - 1991$$
$$\mathbb{C} = number \; of \; articles \; published \; in \; 1990 - 1991$$

then

$$\mathbb{D} = B/C = 1992 \; ISI \; Impact \; Factor \quad (7)$$

If J is the set of journals and j is a specific journal, then Equation (7) is equivalent to (8) in a general form:

$$IF_{j,y} = \sum_{z=y-k}^{y-1} \frac{\sum_{i}^{\forall i \in J} N_{i,y \to j,z}}{P_{j,z}} \quad (8)$$

This metric cannot be applied directly to conferences for the case we want to rank all conferences organized a specific year. This is due to the fact that when we compute the ISI Impact Factor for a conference c for year y, we actually evaluate the events of c that were organized during the previous k years. For example, in order to compute the ISI Impact Factor of VLDB'95, we actually evaluate VLDB'94 and VLDB'93. In order to be able to evaluate a specific conference c held in year y, we "revert" the concept of ISI Impact Factor and instead of counting the citations made to the k previous years, we count the citations made during the next k years (Equation 9). This way we count the "Impact" that has a specific conference during the next 2 years. Let this factor be the "Inverted Impact Factor" or "I-Impact Factor". Actually, this is the reasoning why the VLDB Foundation established the 10-years best-paper award. The *I-Impact Score per Year* is defined as follows:

$$IISY_{c,y} = \frac{1}{P_{c,y}} \sum_{i}^{\forall i \in C} \sum_{z=y}^{y+k} N_{i,z \to c,y} \tag{9}$$

Equations (8) and (9) may be semantically different but they are qualitatively similar as they both count the impact of a collection. In the former, the impact is computed at a specific year, in the latter it is computed for a specific year.

The Inverted Impact Score (Equation 9) metric is a sub-case of the Plain Score per Year (Equation 3) if we set $last\_year = y + k$, where the usual value for k used by ISI is 2 or 5. Since the notion of ISI Impact Factor is widely accepted, we use this metric in our tests as the basic metric to compare with. We cannot use for comparison the ISI Impact Factor as it is defined by Garfield [8] because it is semantically different from the metrics presented here.

Weighted I-Impact Score per Year
The same way, if in Weighted Score per Year (Equation 6) we set $last\_year = y + k$, where k=2 or 5, then we get the I-Impact Score in a weighted manner, let it be $WIISY_{c,y}$. This has the advantages of the I-Impact Score metric, plus the advantages of a weighted metric.

$$WIISY_{c,y,l} = \frac{1}{P_{c,y}} \frac{\sum_{i}^{\forall i \in C} \left(W_{i,y,l-1} * N_{i,y \to c,y} + \sum_{z=y+1}^{y+k} W_{i,z,\infty} * N_{i,z \to c,y} \right)}{\sum_{i}^{\forall i \in C} \left(W_{i,y,l-1} + \sum_{z=y+1}^{y+k} W_{i,z,\infty} \right)}$$

$$WIISY_{c,y} = WIISY_{i,y,\infty} = WIISY_{c,y,L}$$

3.4 The Weight Set

For every distinct ranking[2] we need to compute a set of sets:

$$G = \{G_1, G_2, ..G_n\} \tag{10}$$

where

$$\begin{aligned} G_i &= \{cluster\ of\ conferences\} && \text{for } 1 \leq i \leq n \\ G_1 \cup G_2 \cup ... \cup G_n &= C \\ G_i \cap G_j &= \emptyset && \text{for } 1 \leq i, j \leq n, \quad i \neq j \end{aligned}$$

[2] At level l and for year y, or just at level l when computing ranking for all years.

We have to assign a specific weight value W^i to each set G_i, for $1 \leq i \leq n$. Thus, we have to define the set:

$$W = \{W^1, W^2, .., W^n\}$$

At this point, it is necessary to introduce two important parameters. The *number of clusters* ($\equiv n$) and the *range for the weights*. For our tests, we have set the number of clusters equal to 5 (meaning: very strong, strong, average, weak, very weak). This leads us to have 5 distinct weights and 5 clusters in the conference ranking.

The selection of the weight range is also important as it affects the results in the sense that it tunes the importance of a citation from a "very strong" conference in comparison to the importance of a citation from a "very weak" conference. We decided to use the range 1-5 and specifically the weights W^1=1, W^2=2, W^3=3, W^4=4, W^5=5, in order to emphasize the difference to the Plain Score. For instance, selecting the range 1-2 (i.e. 1, 1.2, 1.4, 1.6, 1.8, 2) would not make any difference.

Actually, since the scores are normalized by dividing with the sum of weights, the important factor is the fraction of the weights divided with the minimum one, and not the absolute values. Thus, it is safe to accept a minimum weight of 1. It is obvious that there is no sense in using a negative or zero weight[3].

Also, we defined that the conferences which belong in a different scientific domain than the one the ranking is computed for, to be members of the G_1 set. In addition to that, conferences that have zero score ($\Leftarrow$ 0 citations to them), are set by the classification algorithm in group G^1, as well.

3.5 Clustering Conferences According to Citations

The clustering algorithm is a hierarchical clustering algorithm applied on one-dimension points [12]. Initially, a number of clusters N is defined, where N is the number of conferences for ranking.

$$G_1 = \{S_1\} \quad G_2 = \{S_2\} \quad ... \quad G_N = \{S_N\}$$

For each cluster we set G_x^A as the average value of G_x members. In every step of the algorithm, we find two sets G_i and G_j, for which the difference of their average values ($|G_i^A - G_j^A|$) is the minimum of any other pair. We define a new set $G_k = G_i \cup G_j$ and we delete the sets G_i and G_j. The procedure is repeated until the number of groups reaches n. If, when reaching n, there exists a pair with zero difference of their average values ($|G_i^A - G_j^A| = 0$)[4], we continue joining the clusters, until we get $|G_i^A - G_j^A| > 0 \;\; \forall G_i, G_j \in G$.

[3] A weight can be set to zero, iff the appropriate conference does not exist. This happens in the case of computing the rankings per year, where not all conferences are present within a specific year, either because they were not organized or they are just absent from our database.

[4] This occurs only when ALL the members of the two groups have exactly the same score.

3.6 Weight Refinement

The Weighted Score algorithm, as described above, is open to deadlocks. This is due to the fact that there is no guarantee that a conference will not move from one cluster to another at some point during the algorithm execution. We illustrate this situation with a simple example of two conferences A and B for which:

$$P_A = P_B = x(= 10)$$
$$N_{A\to B} = 4, N_{B\to B} = 0, N_{B\to A} = 3, N_{A\to A} = 0$$

In this case:

$$\begin{array}{ll} level: 1 & \left.\begin{array}{l} WS_{A,1} = 0.15 \\ WS_{B,1} = 0.2 \end{array}\right\} \Rightarrow \left.\begin{array}{l} W_{A,1} = 1 \\ W_{B,1} = 2 \end{array}\right\} \Rightarrow \\ level: 2 & \left.\begin{array}{l} WS_{A,2} = 0.2 \\ WS_{B,2} = 0.13 \end{array}\right\} \Rightarrow \left.\begin{array}{l} W_{A,2} = 2 \\ W_{B,2} = 1 \end{array}\right\} \Rightarrow \\ level: 3 & \left.\begin{array}{l} WS_{A,3} = 0.1 \\ WS_{B,3} = 0.26 \end{array}\right\} \Rightarrow \begin{array}{l} W_{A,3} = 1 \\ W_{B,3} = 2 \end{array} \end{array}$$

This leads to an infinite loop since at level 4 we will get exactly the same results as at level 2. In order to avoid this case, after the computation of the clusters $G_1, G_2, ..., G_n$ at level l, and before we assign the weights $(W^1, W^2, ..., W^n)$ for each conference, we check if the same condition has been raised in a previous level d. If there is a level $d < (l-1)$ for which $W_{c,d} = W^k$ $(c \in G_k)$, $\forall c \in C$[5], then we do not set $W_{c,l} = W^k$ (as we should do), but instead we set:

$$W_{c,l} \;=\; avg(W_{c,d}...W_{c,l-1}) \;=\; \frac{\sum_{p=d}^{l-1} W_{c,p}}{l-d}$$

This way, $W_{c,l} \stackrel{l\to\infty}{\longrightarrow} x$, where x is a real number. Actually, since we have distinct weights, we reach x very fast.

In the previous example the next steps should be:

$$\begin{array}{ll} level: 3 \;\Rightarrow & \left.\begin{array}{l} W_{A,3} = 1.5 \\ W_{B,3} = 1.5 \end{array}\right\} \Rightarrow \\ level: 4 & \left.\begin{array}{l} WS_{A,4} = 0.15 \\ WS_{B,4} = 0.2 \end{array}\right\} \Rightarrow termination \end{array}$$

4 Results

First, we note that the ranking is made for only one out of the four clusters that have been presented in Section 3.1. That is, we focus in the Database cluster as it is the most complete cluster in the DBLP database. The database contains conferences from 1959 to 2002 (but "complete" data for these conferences exist only for the year 1980 and afterwards). Thus, we find the ranking for each year separately by using: the Plain Score per Year, the Weighted Score per Year, the Plain I-Impact Score per Year, the Weighted I-Impact Score per Year.

[5] If d=l–1 then the termination condition is held.

For all runs, the same algorithm is used. The Plain Score is the result of the algorithm at level 1 (all the weights at level 0 are equal to 1). The results of a weighted ranking are the results of the last reached level. The I-Impact Score is a sub-case of the previous ones and, therefore, we reach it if we set the variable *last_year* equal to $y+2$ [6].

4.1 Rank Comparisons

In order to visualize the comparison of the various ranking results, we use q-q plots (quantile-quantile plots), which illustrate the quantiles of one univariate distribution against the corresponding quantiles of another (in our case, we compare the rankings). Therefore, for comparing the type A ranking to type B ranking, for each conference c in our rank table, we put a dot in the graph at point (x, y) where x is the position of c by using type A, whereas y is the position of c by using type B. Thus, the x-axes represent positions computed by A and y-axis positions computed by B. The two rankings would be equivalent iff $y=x$ for every point in the graph. It is easy to notice from the q-q plots (Figures 1,2), that the various results do not differ substantially for the "very strong" or "very weak" conferences but mainly for the "average" cluster.

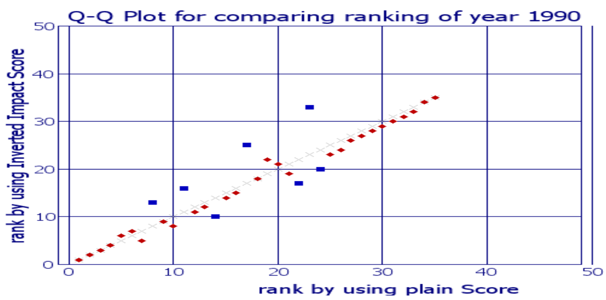

Fig. 1. Comparison of the ranking for year 1990.

In all q-q plots that compare *I-Impact Score* ranking (either Weighted or Plain) and *Score* ranking (either Weighted or Plain) (Figures 1, 2a, 2b), there are some outliers (marked as blue squares in the graphs), for which $x \ll y$, meaning that they have much better rank position by using Score than I-Impact Score. This is due to the nature of the Impact Score notion, where only citations made in the next k (2 in our tests) years are taken into account. Thus, these specific conferences do not have big "Impact", meaning that do not have a lot of citations during the next 2 years, but they have citations until "now".

[6] All results are accessible at the web location: http://delab.csd.auth.gr/sceas

Specifically in Figures 2(a) and (b) the outlier that lies above the line y=x is the CPM'96 Conference (Combinatorial Pattern Matching[7]). The specific conference does not get any citations during the next 2 years, so it's I-Impact Score is low. However, if we have to evaluate its overall contribution to the academic community, we have to see the *Score* Ranking.

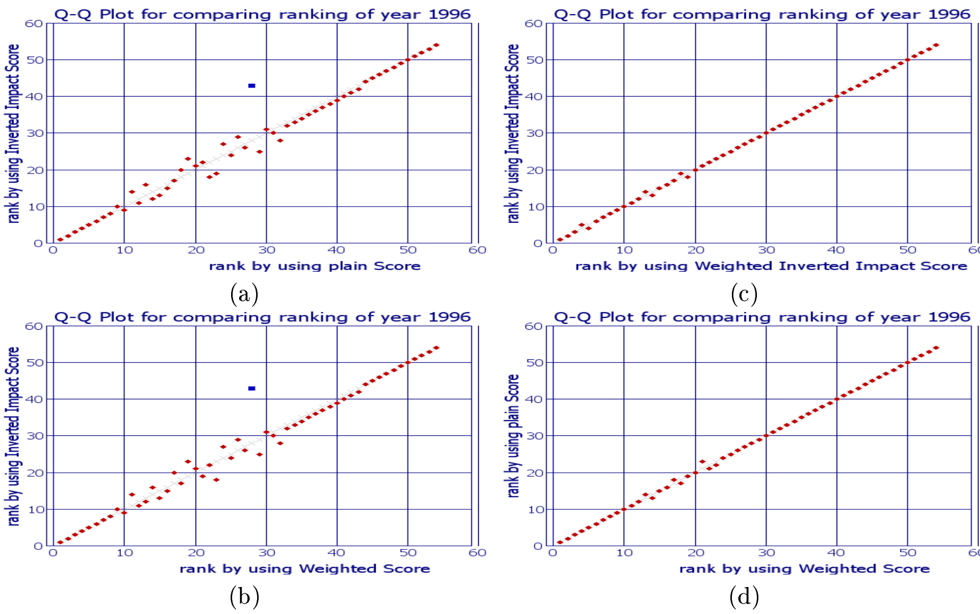

Fig. 2. Comparison of the ranking for year 1996.

In an analogous way, the outliers for which $y \ll x$, are conferences with big *I-Impact Score* but low *Score*. These conferences get a lot of citations during the next k years, but the citations are decreasing over time, meaning that they do not contain "citation classics" papers.

In cases of Figures 2c, 2d, the outliers are very close to the line y=x and quantitatively few. This means, that there is no radical repositioning in the plain ranking by adding the notion of weight, although the fraction W^5/W^1 we have used is high (=5). There are some re-orderings, which help in refining the ranking. The conferences of Figure 2c, for which $x \neq y$, are shown in detail in Table 4.1. We see that the HT Conference (= ACM Conference on Hypertext) and the SPIESR Conference (= Storage and Retrieval for Image and Video Databases) have swapped positions after computing the weighted score. In the Plain Score the scores for these two conferences are very

[7] The specific conference does not strictly belong to the database cluster, but it has been placed in this cluster by the algorithm as we have not defined deliberately any other closer scientific domain.

Weighted Score		Plain Score		
pos	score	pos	score	conference
13	0.404208	14	0.287565	ACM Conference on Hypertext
14	0.382026	13	0.287772	Storage and Retrieval for Image and Video Databases (SPIE)
17	0.236508	18	0.168072	Database and Expert Systems Applications (DEXA)
18	0.223758	17	0.168552	Digital Libraries
21	0.160647	23	0.110927	Advances in Databases and Information Systems (ADBIS)
22	0.158213	21	0.119178	Australasian Database Conference (ADC)
23	0.154697	22	0.116530	British National Conference on Databases (BNCOD)

Table 1. Detailed Comparison of Plain Score vs. Weighted Score for the 1996 year.

pos	score	#papers	weight	conference
1	10.82818	64	5	ACM SIGMOD Conference
2	7.52698	72	4	Very Large Data Bases (VLDB) Conference
3	6.14474	26	4	Symposium on Principles of Database Systems (PODS)
4	3.71273	27	3	Conf. on Parallel and Distributed Information Systems (PDIS)
5	3.61389	81	3	Int. Conference on Data Engineering (ICDE)
6	2.60681	47	3	Int. Conf. on Extending Database Technology (EDBT)
7	1.65844	17	2	Research Issues in Data Engineering (RIDE)
8	1.19002	74	2	Knowledge Discovery and Data Mining (KDD)
9	0.68100	23	2	Int. Conf. on Cooperative Information Systems (CoopIS)
10	0.59669	28	2	Statistical and Scientific Database Management (SSDBM)

Table 2. Rank with Weighted Score for the 1996 year.

close. The weighted score for “HT” is greater than the one of “SPIESR”, as it has more citations from “very strong” conferences.

4.2 Rank Results

Besides the introduction of the new metrics for citation analysis, in this paper we report some first rankings of database conferences.[8] The presentation of the ranking results derived from the SCEAS system are made by using two ways:

– by a Rank table, assuming a specific type of ranking and a selected year. For example, in Table 2 we present the ranking by using the Weighted Score for year 1995.
– by a Historical chart, where we can view the whole history of a conference for any specific type of ranking.

In Figure 3, the history of ranking of VLDB conference is presented, according to all types rankings. Each bar consists of three parts. The bottom part (blue color) gives the percentage of conferences that have a lower ranking, the top part (red) gives the percentage of conferences that have a higher ranking, whereas the middle part (green) gives the percentage of conferences that have equal ranking. In addition, the ratio below the x-axis gives the relative rank for each year. A different position occurs for several

[8] Full presentation of the results is available at the web location http://delab.csd.auth.gr/sceas/ . However, the results shown there may be slightly different than the ones presented here since the database is continuously updated.

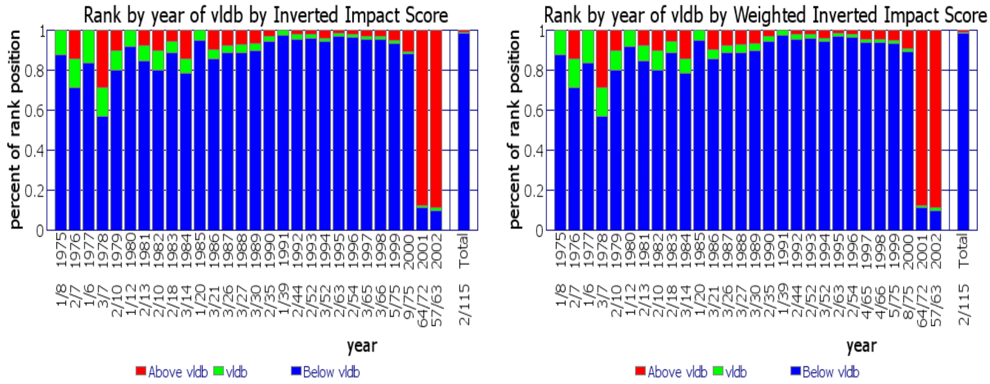

Fig. 3. The ranking history of VLDB.

years, but all the graphs are very similar since the specific conference is clearly a “very strong” one during all years[9].

Note that the last 2 years ranking (i.e. 2001 and 2002) could not be considered as reliable, since there are no citations in our database to conferences organized during these years.

5 Conclusion

In this paper we first presented an overview of two major current systems for conference and journal ranking by using citation analysis, CiteSeer and SCI. A weak point of these systems is that they are based on the ISI Impact Factor, thus, considering citations in a flat way, e.g. without paying attention to the quality of the respective publication. Therefore, we introduced four new metrics in order to cure this deficiency, which are suitable for considering both journal and conference publications. These new metrics are used by a system that we have built, the SCEAS system (Scientific Collection Evaluator by using Advanced Scoring). The system is autonomous and has the following characteristics:

– it imports the DBLP bibliography records into a local database (it could be extended to import any other scientific collection of publications),
– it partitions the imported collection into clusters according to the topic of the conference and performs a cleansing step to provide reliable information.
– it performs the ranking by all four metrics for the conferences that focus in databases.

The web user of SCEAS system has access to all the results produced at any stage of the rank process, can compare the various rank metrics and can study the rank results in order to derive useful information regarding the quality of the database conferences.

In the future we plan to extend the system by:

[9] VLDB and SIGMOD are not directly comparable to the other references since they includes “industrial” publications, which can be excluded only in manually. In other words, in reality the distance between these two conferences from the third one is greater.

– Computing more variations of the weighted metrics in which self-citations of the collection could be exludeded or taken into account multiplied with a smaller weigth.
– Extenting the ranking in more collections/sientific domains. This could give us the ability when ranking a cluster (e.g. DB conferences) to take into account "weighted" citations from other type of collections belonging in the same scientific field (e.g. DB Journals and Books).

References

1. Science Citation Index by the Institute for Scientific Information: http://www.isinet.com/isi/ (valid as of 1/4/2003).
2. Citeseer by the NEC Research Institute: http://citeseer.nj.nec.com/ (valid as of 1/4/2003).
3. Citeseer Ranking page: http://citeseer.nj.nec.com/impact.html (valid as of 1/4/2003).
4. Data Bases and Logic Programming (DBLP) of the University of Trier: http://www.informatik.uni-trier.de/~ley/db/ (valid as of 1/4/2003).
5. Y. An, J. Janssen, and E. Milios. Characterizing and Mining the Citation Graph of the Computer Science Literature. *Knowledge and Information Systems,*. to appear.
6. S. Brin and L. Page. The Anatomy of a Large-scale Hypertextual Web Search Engine. In *Proceedings 7th WWW Conference*, pages 107–117, 1998.
7. D. Dhyani, S. Bhowmick, and W. K. Ng. Deriving and Verifying Statistical Distribution of a Hyperlink-based Web Page Quality Metric. In *Proceedings 13th DEXA Conference*, pages 19–28, 2002.
8. E. Garfield. The Impact Factor. http://www.isinet.com/isi/hot/essays/ journalcitation-reports/7.html (valid as of 1/4/2003).
9. E. Garfield. Citation Analysis as a Tool in Journal Evaluation. *Essays of an Information Scientist*, 1:527–544, 1972.
10. A. Goodrum, K. McCain, S. Lawrence, and C. L. Giles. Scolarly Publication in the Interent Age: a Citation Analysis of Computer Science Literature. *Information Processing and Management*, 37(2):661–675, 2001.
11. E. Han, G. Karypis, V. Kumar, and B. Mobasher. Clustering based on Association Rule Hypergraphs. In *Proceedings DMKD'97 Workshop*, 1997.
12. A. K. Jain, M. N. Murty, and P. J. Flynn. Data Clustering: a Review. *ACM Computing Surveys*, 31(3):264–323, 1999.
13. J. M. Kleinberg. Authoritative Sources in a Hyperlinked Environment. *Journal of the ACM*, 46(5):604–632, 1999.
14. 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.
15. S. Lawrence, K. D. Bollacker, and C. L. Giles. Indexing and Retrieval of Scientific Literature. In *Proceedings 1999 ACM CIKM Conference*, pages 139–146, 1999.
16. S. Lawrence, C. L. Giles, and K. Bollacker. Digital Libraries and Autonomous Citation Indexing. *IEEE Computer*, 32(6):67–71, 1999.
17. G. Meghabghab. Discovering Authorities and Hubs in Different Topological Web Graph Structures. *Information Processing and Management*, 38(1):111–140, 2002.
18. L. Pretto. A Theoretical Analysis of Google's PageRank. In *Proceedings SPIRE Symposium*, pages 131–144, 2002.