Toward Alternative Measures for Ranking Venues: A Case of Database Research Community

Su Yan, Dongwon Lee
College of Information Sciences and Technology
The Pennsylvania State University
University Park, PA 16802
syan@ist.psu.edu, dongwon@psu.edu

ABSTRACT

Ranking of publication venues is often closely related with important issues such as evaluating the contributions of individual scholars/research groups, or subscription decision making. The development of large-scale digital libraries and the availability of various meta data provide the possibility of building new measures more efficiently and accurately. In this work, we propose two novel measures for ranking the impacts of academic venues – an easy-to-implement *seed-based measure* that does not use citation analysis, and a realistic *browsing-based measure* that takes an article reader's behavior into account. Both measures are computationally efficient yet mimic the results of the widely accepted Impact Factor. In particular, our proposal exploits the fact that: (1) in most disciplines, there are "top" venues that most people agree on; and (2) articles that appeared in good venues are more likely to be viewed by readers. Our proposed measures are extensively evaluated on a test case of the Database research community using two real bibliography data sets – ACM and DBLP. Finally, ranks of venues by our proposed measures are compared against the Impact Factor using the Spearman's rank correlation coefficient, and their positive rank order relationship is proved with a statistical significance test.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; I.7.4 [**Document and Text processing**]: Electronic Publishing

General Terms

Algorithms, Design, Experimentation, Measurement

Keywords

Bibliometrics, Citation Analysis, Venue Ranking, Impact Factor, Digital Library

JCDL'07, June 18–23, 2007, Vancouver, British Columbia, Canada.

1. INTRODUCTION

Ranking publication venues based on various bibliometrics is an important and oft-studied field. Proper venue rankings offer objective methods to evaluate the relative intellectual influence of a particular venue. Thus, the impacts of venues are often related with many decision scenarios – e.g., deciding which venues to subscribe in a library or to evaluate the performance of scholars or research groups [15].

The questions such as "How good is a journal X?" or "Is a conference X better than Y?" are inherently difficult to answer since they involve subjective measures. Nevertheless, to answer these questions, people have proposed many methods (e.g., [1, 15, 7]). However, in this paper, we argue that existing methods are not sufficient. Let us elaborate this point in the motivation below.

1.1 Motivation

By and large, most of existing works on venue ranking perform some kinds of *citation analysis* to evaluate the impact of publication venues. For example, the Thomson ISI *Impact Factor* (IF hereafter) [1] is one of the well-accepted methods of this kind. In that measure, the citation of "the article a cites the article b" is treated as the endorsement from a to b. Although being intuitive, IF is *not* without problems, limiting its validity and applicability (e.g., [14, 2]). For instance,

1. One of the common issues of the citation-based methods is the hardness of the citation meta data extraction and parsing from articles. An accurate citation analysis depends on clean and comprehensive citation information. In general, there are two kinds of citation extraction and parsing methods – manual (as being used in DBLP, ISI JCR) and automatic methods (as being used in Google Scholar, CiteSeer). The manual methods can provide cleaner (or more accurate) citations, but demand for extensive manual labor. Thus the application of the manual methods is restricted to a manageable subset of all the available publications. On the other hand, the automatic methods use computer programs to automatically extract and parse citation meta data. However, since computer programs cannot handle the diverse cases in the citation format, the qualities of the extracted citations are often less than satisfactory. Often, the automatically extracted citation data cannot be used to evaluate the quality of publication venues or scholars.

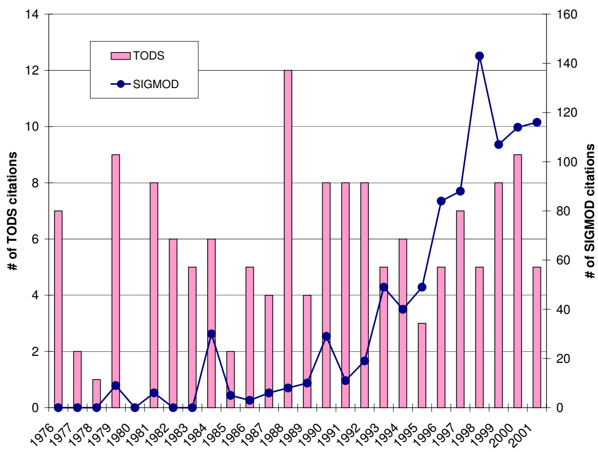

Figure 1: Yearly distribution of # of TODS/SIGMOD papers being cited in 2002.

2. Existing citation-based venue ranking methods tend to consider only the explicit citation relationships as indicated in the reference parts of academic articles. However, when an author writes an article, it is reasonable to expect for the author to read a large number of articles, but cite only a fraction of them. The final decision of making references in the article depends on many factors. Further more, it has been shown that citations tend to have problems like biased-citation, self-citation, or positive vs. negative citation [9].

3. Most of the citation-based methods focus on the ranking of "journals", excluding "conferences" or "workshops" [1]. Conferences are becoming more and important publication outlets in Computer Science. However, their publication and citation behaviors do not parallel with those of journals. For instance, [12] studies the IF of 5 major database publication venues (2 journal and 3 conferences) and concludes that two major database conferences have higher IF values than the other two database journals. An implicit hypothesis of the work is that journals and conferences have the similar citation patterns so that they can be compared uniformly. However, as shown in Figure 1, the hypothesis does *not* hold sometimes. For instance, TODS and SIGMOD are two top database journal and conference respectively. Yet, they show very different citation patterns. That is, while people cite old TODS papers as frequently as they cite new TODS papers, people seem to prefer to cite newly published SIGMOD papers. Such difference in citation patterns may indicate the fundamental difference in publication patterns between journals and conferences.

1.2 Our Approach

To remedy the aforementioned issues of citation-based methods, in this paper, we argue that:

1. Instead of relying on citation analysis, other kinds of meta data may also be useful. For instance, extracting and parsing the title or authors of an article is much easier and more accurate than extracting and parsing citations from the reference section of an article. Therefore, we propose to exploit author information of articles in measuring the impact of venues.

2. We do not desert the citation meta data, since has the citation has been widely acknowledged as a kind of informative meta data. However, in order to use the citation data more properly, we need to define new metrics to take into account the differences in citation patterns among various publications venues.

3. Although it is difficult to rank venues in general, in practice, people are more interested in the question: "What are the *top-k* venues in the field f?". Furthermore, note that this question can be solved if the following two sub-questions can be answered:

   S1: What is the set of good articles, $Seed_P$?

   S2: What are the *top-k* venues that are most *similar* in their qualities to $Seed_P$?

In this paper, we answer the two sub-questions S1 and S2 by proposing several new bibliometrics that can be used in ranking publication venues. The new bibliometrics, which have the properties of easier-to-implement and more accurate, provide new perspectives and alternative ways in evaluating publication venues, which supplement existing ranking methods to provide a more comprehensive estimation of the scientific qualities of publication venues.

2. MAIN PROPOSAL

We first define the *goodness* of a venue that is used throughout the paper. The goodness of a venue is defined as the `sum` of the goodness of articles in it – i.e., a venue a is "better" than a venue b if a has more "good" articles than b has[2]. Note that the definition of the goodness of a venue is intentionally *recursive* by using the goodness of an article, which is to be defined later.

Definition 1 (Goodness of Venue *B*) *Let P be the set of articles in a venue B, $\mathbf{G}(p)$ be the goodness of an article p ($p \in P$), and $\triangle$ be the normalization factor to make $\mathbf{G}(B)$ constant. Then, the goodness of venue B is:*

$$\mathbf{G}(B) = \triangle \sum_{p \in P} \mathbf{G}(p)$$

Further, venue B_1 is said to be better *than venue B_2 if $\mathbf{G}(B_1) > \mathbf{G}(B_2)$.*

where, $0 \leq \mathbf{G}(B) \leq 1$ □

In the following sections, we propose several alternative definitions of the "good article" – $\mathbf{G}(p)$, but the definition of the "good venue" remains the same.

[1]We treat conferences, symposiums, and workshops equally in this paper – i.e., they are referred to as conferences.

[2]Other definitions of the goodness of a venue are also possible, such as using `avg` or `max`. For instance, using `avg`, the semantics is changed to: a venue a is "better" than a venue b if the articles in a are better than those in b on average.

2.1 Sub-question S1

Sub-question S1 asks for a collection of good articles, denoted as $Seed_P$, which acts as a seed for sub-question S2. The basic hypothesis is:

Hypothesis 1 *There are a number of good articles in each subject field that most people agree on (denoted as $Seed_P$).* □

Using this Hypothesis, to identify $Seed_P$, we consider two solutions to the sub-question S1 as follows:

1. Users may provide a number of good articles that collectively can be used as a seed. For instance, all the articles of certain prestigious journals or conferences may be used as seed articles $Seed_P$.
2. Seed articles, $Seed_P$, may be decided via citation count. For instance, those articles whose accumulated citations are above a threshold may be considered as seed articles.

Note that both approaches to S1 are very simple yet intuitive. For instance, if people agree that JCDL is one of the top venues in the digital library community, then one can answer the sub-question S1 by using all the articles of JCDL as seed articles, $Seed_P$.

2.2 Sub-question S2

From here forward, we assume that the sub-question S1 is resolved somehow – i.e., $Seed_P$ is known, and focus on solving the sub-question S2, which is much harder to solve and evaluate. In particular, we propose two novel measures – one based on seed articles directly and the other exploiting the user-browsing model.

2.2.1 Seed-Based Measure

Hypothesis 2 *Authors of seed articles, $Seed_P$, are authoritative authors (denoted as $Seed_A$) and are likely to produce good quality articles.* □

In order words, Hypothesis 2 assumes that each community or subject field has a certain number of "top" scholars who are known for writing good quality articles. Using the given $Seed_P$ as the starting seed, therefore, one can test the quality of an article p by checking if p is contributed by any subset of authors in $Seed_A$ or not. Now, based on Hypothesis 2, we consider three variations of seed-based goodness measures for an article – *naive*, *fair*, and *unfair* policies.

Definition 2 (Seed-based Goodness of an Article) *Suppose that an article p has n co-authors a_i, $(1 \leq i \leq n)$, P_a^i is the set of articles in $Seed_P$ to which an author a_i has contributed, W_i is a weight factor with the the constraint $\sum_{i=1}^{n} W_i = n$ ($W_i = 1$ by default), and $\triangle$ is a normalization factor. Then,*

- *Naive:* $\mathbf{SG}_n(p) = \begin{cases} 1, & \text{if } a_i \in Seed_A \\ 0, & \text{otherwise} \end{cases}$
- *Fair:* $\mathbf{SG}_f(p) = \frac{1}{n}\sum_{i=1}^{n} W_i X_i$, *where* $X_i = 1$ *if* $a_i \in Seed_A$*, and 0, otherwise.*
- *Unfair:* $\mathbf{SG}_u(p) = \frac{1}{n}\sum_{i=1}^{n} W_i X_i$, *where* $X_i = \triangle \frac{|P_a^i|}{|Seed_P|}$ □

Given an article p, $\mathbf{SG}_n(p)$ considers p as "good" as long as p is contributed by any author in $Seed_A$. Therefore, $\mathbf{SG}_n(p)$ does not differentiate the quality of two articles to which different number of $Seed_A$ have contributed. On the other hand, $\mathbf{SG}_f(p)$ views the quality of an article in terms of the number of co-authors in $Seed_A$. Therefore, an article with n "top" co-authors are considered to be n times better than an article with n co-authors where only one is "top" author. Obviously, $\mathbf{SG}(p)$ has the range of $0 \leq \mathbf{SG}(p) \leq 1$. The weight factor W_i can be used to give relative importance among co-authors. Consider two articles p_1 and p_2, where only the first co-author of p_1 is $Seed_A$ while only the second co-author of p_2 is $Seed_A$. If one wants to apply an arbitrary logic such as "the first author counts three times more than the second author", then one can give non-uniform weights like $W_1 = 1.5$ and $W_2 = 0.5$. Then, $\mathbf{SG}_f(p_1) = \frac{1.5*1+0.5*0}{2} = \frac{1.5}{2} = 0.75$ and $\mathbf{SG}_f(p_2) = \frac{1.5*0+0.5*1}{2} = \frac{0.5}{2} = 0.25$. For our current work, we use uniform weights, ignoring the order information among author names.

The third variation that we consider is to give different influence among authors in $Seed_A$. This mimics the fact that "if an author a_1 has written 50 articles in $Seed_P$ while another author a_2 has written only 1 article in $Seed_P$, then a_1 has a higher probability of producing top quality articles than a_2 has". Therefore, experienced authors (e.g., professors) would have more impact toward the goodness of an article than amateur authors (e.g., graduate students) would. Since this scheme is not fair to amateur authors, we name it as *unfair* while we name the other case as $fair$. Note that, more complex methods can also be used to assign influence to authors, such as computing PageRank for the authors [8]. However, we stick to our simple scheme in this work.

The goodness of venue $\mathbf{G}(B)$ in Definition 2 is then defined accordingly for the three variations. "$\mathbf{G}(B) = 1$" roughly implies that "the quality of venue B is as good as that of $Seed_P$". The overall accuracy of the seed-based measures heavily depends on the correct choice of the initial $Seed_P$.

2.2.2 Browsing-Based Measure

In defining the above measures, we only consider the static bibliographic properties of the articles without considering article readers' behaviors. We believe that a realistic venue-ranking scheme should match real readers' evaluations to the venues. Therefore, we argue that incorporating readers' behaviors into a venue-ranking scheme is a reasonable choice. In this section, we mathematically model a common reader's behavior when browsing through a collection of articles in digital libraries, denoted as the the *article-browsing model*, and derive a *browsing-based measure* accordingly.

Suppose a reader is reading an interesting article p_i in the field f. If the reader wants to find more interesting articles about field f to read, he/she may achieve the goal by following two paths among others: (1) select more articles from the reference part of the article p_i, or (2) find articles authored by the same author of p_i. When the reader has

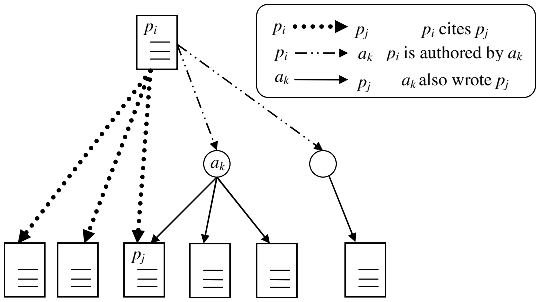

Figure 2: The article browsing model.

switched to the newly found article p_j, he/she may follow either of the two paths to further select more relevant articles to read. Figure 2 illustrates this paper browsing process conceptually. The reader can follow the paths indicated by different kinds of lines to select the next article to read.

Based on the above observation of a reader's article browsing behavior, we propose the *article-browsing model* as follows. The theory of the model is a modification of the well-known PageRank model by using the personalization vector to change the teleportation matrix. However, the new model is different than computing PageRank on the citation graph. Intuitively, a reader's article browsing pattern is incorporated in the new model, and not only the references but also the authors of a paper will influence the reader's choice of articles. Google's original intent in introducing the personalization vector is to deal with different classes of surfers [6]. [4] extends this idea to improve the ranking of search-query results by allowing query-time information to influence the link-based score. Using the matrix representation, the basic PageRank is the solution to

$$\mathbf{r} = M \times \mathbf{r} \tag{1}$$

where M is the matrix corresponding to the directed Web graph G, and $\mathbf{r}$ is the vector that contains the PageRank value of all of the web pages. To ensure PageRank convergence, M must be stochastic, irreducible and aperiodic [6]. The last one is guaranteed in practice. M is modified to be stochastic by assigning artificial out-links to the *dangling nodes* to all other nodes in G, and be irreducible by damping the rank propagation by a factor of α.

Let $\mathbf{p}$ be an n-dimensional column vector representing a uniform probability distribution over all the nodes:

$$\mathbf{p} = \frac{1}{n}\mathbf{e}. \tag{2}$$

where $\mathbf{e}$ is the column vector of all ones. Let $\mathbf{d}$ be the n-dimensional column vector identifying the nodes with out-degree 0:

$$d_i = \begin{cases} 1, & \text{if } deg(i) = 0 \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

Then the stochastic, irreducible $\overline{M}$ is constructed as follows:

$$\begin{aligned} D &= \mathbf{p} \times \mathbf{d}^T \\ E &= \mathbf{p} \times \mathbf{e}^T = [\frac{1}{n}]_{n\times n} \\ \overline{M} &= (1-\alpha)(M + D) + \alpha E. \\ \mathbf{r} &= \overline{M} \times \mathbf{r} \end{aligned} \tag{4}$$

To model a reader's article browsing pattern, we use a citation graph C in place of the Web graph G. Then M is the matrix corresponding to the directed citation graph. Let P be the set of all articles in the data collection, P_A be the set of articles written by authoritative authors, and function $author(p_j)$ stand for the authors of article p_j. We substitute the uniform vector $\mathbf{p} = \frac{1}{n}\mathbf{e}$ with the non-uniform vector $\mathbf{w}$ into Equation 4, where

$$w_{ji} = \begin{cases} \frac{1}{|P_A|} & \text{if } author(p_j) \cap Seed_A \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

Then the *article-browsing model* becomes:

$$\mathbf{r} = (1-\alpha)(M + \mathbf{w} \times \mathbf{d}^T) \times \mathbf{r} + \alpha\mathbf{w} \tag{6}$$

Since $\mathbf{w}^T > 0$ is a probability vector, every node (i.e., article) is still directly connected to every other node, forcing Equation 6 to converge.

Roughly speaking, the *article-browsing model* ranks each article with respect to the probability that a reader choose the article when browsing through a collection of articles. Then, using this model, we define the *browsing-based goodness* as follows:

Definition 3 (Browsing-based Goodness of an Article) *The browsing-based goodness of an article p is:*

$$\mathbf{BG}(p) = r_p$$

where r_p is determined by the article-browsing model of Equation 6. □

This measure indicates that articles browsed by readers with higher probability are likely to be of better quality than otherwise. Obviously, $\mathbf{BG}(p)$ has the range $0 \leq \mathbf{BG}(p) \leq 1$.

3. EMPIRICAL EVALUATION

In this section, we empirically evaluate the two proposed measures. Since the goodness of venues is a quite subjective matter, we do not seek to conduct a user study here. Rather, we test if the two proposed measures can produce ranking results comparable to some known methods. That is, we compare our results to those of IF since it is one of the most adopted ones.

3.1 Set-Up

To rank academic publication venues, we used two data sets – the ACM data set (ACM hereafter) and the DBLP data set. Detailed data characteristics are summarized in Tables 4 and Table 5. Like CiteSeer, ACM also extracts meta data from each paper automatically. In recent years, ACM begins to require authors to submit meta data together with the published articles, improving the quality of meta data

	Naive		Fair		Unfair		Browsing	
Rank	Venue	Score	Venue	Score	Venue	Score	Venue	Score
1	VLDB	1.000	VLDB	1.000	VLDB	1.000	TODS	3.161
2	EDBT	0.755	VLDB-J	0.519	VLDB-J	0.689	VLDB	3.025
3	VLDB-J	0.750	DBPL	0.504	SIGMOD	0.589	SIGMOD	2.597
4	DBPL	0.721	EDBT	0.498	EDBT	0.562	WebDB	2.324
5	WebDB	0.679	WebDB	0.493	WebDB	0.526	EDBT	2.169
6	SIGMOD	0.654	TODS	0.401	TODS	0.397	VLDB-J	1.955
7	DS	0.618	SIGMOD	0.393	ICDE	0.364	DBPL	1.954
8	IQIS	0.600	DS	0.387	PODS	0.332	PODS	1.809
9	SSD	0.597	ICDT	0.382	DBPL	0.330	ICDE	1.754
10	TODS	0.540	SSD	0.378	FODO	0.323	ICDT	1.747
11	CoopIS	0.539	FODO	0.364	DPD	0.276	SSD	1.687
12	DPD	0.536	ICDE	0.359	SSD	0.270	DNIS	1.583
13	ICDE	0.530	DPD	0.320	DNIS	0.267	DS	1.565
14	PODS	0.521	CoopIS	0.299	DKD	0.254	IQIS	1.528
15	ICDT	0.517	SIGMOD Rec.	0.276	ICDT	0.242	DASFAA	1.512
16	FODO	0.513	DASFAA	0.269	SIGMOD Rec.	0.237	CoopIS	1.406
17	DASFAA	0.488	BNCOD	0.259	DS	0.228	SSDBM	1.403
18	SIGMOD Rec.	0.418	PODS	0.249	CoopIS	0.212	DAWAK	1.382
19	BNCOD	0.415	Inf. Syst.	0.244	DAWAK	0.192	FODO	1.354
20	DAWAK	0.413	DAWAK	0.242	DASFAA	0.176	RIDE	1.325

Table 1: Ranking results (Seed = VLDB conference)

Statistics	Value
# of papers	842,422
# of papers in G	281,904
# of papers being cited	206,783
# of papers citing other papers	138,638
# of papers in G without authors	3,799
# of authors in G	258,421

Table 4: Statistics of ACM data set and their citation graph G.

Statistics	Value
# of distinct conf., symp., and workshop	2,530
# of distinct journals	438
# of papers	500,462
# of distinct venues matching ACM	2,385

Table 5: Statistics of DBLP data set.

substantially. ACM also gets meta data information from affiliated publishers. However, errors such as missing the data for a certain year of a publication venue (e.g., ICDE 2003) or missing the author information (e.g., SSDBM 2004 has no author information) are found in ACM dataset. Since DBLP manually extracts (by human editors) meta data from each publication, its data can be deemed to be cleaner (although there are still some errors). However, DBLP does not have the paper citation information.

Since neither the ACM nor the DBLP data set is complete and error-free, we pre-processed and consolidated them as follows. We first linked DBLP and ACM using titles (and ISBN if exists) of articles, remove all conflicting authors and venues, and form the DBLP-ACM data set – a clean data set with citation information. After linking DBLP and ACM data, we hand-picked publication venues (journals, conferences, symposiums and workshops, 86 in all) that we believe to be closely-related to the Database research community in Computer Science. Note that we intentionally excluded venues that have some database papers but also have papers from other fields (e.g., J. ACM, ACM Comm. ACM, and WWW). Hereafter, we will refer to this collection of 86 venues as **DBLP-ACM**, as shown below. At the end, the DBLP-ACM data set contains 32,192 papers and 34,216 authors in the Database community.

ADBIS, ADC, ARTDB, BNCOD, CDB, CIKM, CoopIS, DANTE, DASFAA, DAWAK, DB, DBPL, DBSEC, DEXA, DKD, DKE, DL, DMKD, DNIS, DOLAP, DOOD, DPD, DPDS, DS, EDBT, ER, FODO, FOIKS, FQAS, GIS, HPTS, ICDE, ICDM, ICDT, ICIS, IDA, IDEAL, IDEAS, IGIS, Inf. Process. Lett., Inf. Sci., Inf. Syst., IPM, IQIS, ISF, ISR, IW-MMDBMS, IWDM, JDM, JIIS, JMIS, K-CAP, KA, KDD, KER, KIS, KR, MDA, MFDBS, MLDM, MMDB, MSS, NLDB, OODBS, PAKDD, PKDD, PODS, RIDE, RIDS, SIGKDD Exp., SIGMOD, SIGMOD Rec., SSD, SSDBM, TKDE, TODS, TOIS, TSDM, UIDIS, VDB, VLDB, VLDB-J, WebDB, WIDM, WISE, XMLEC

3.2 Result

First, to solve the sub-question S1, we use SIGMOD and VLDB as seed venues after consulting colleagues in the Database community. That is, all articles from either SIGMOD or VLDB are considered to be "good" articles. Second, we use the threshold 10%, and all articles whose numbers of accumulated citations are above top 10% are considered to be "good" articles. After finding seed papers, $Seed_P$, we apply the seed-based measure and the browsing-based measure to solve the sub-question S2. Both measures are applied to the three different $Seed_P$, which are the VLDB papers, the SIGMOD papers, and the upper 10% highest cited papers.

	Naive		Fair		Unfair		Browsing	
Rank	Venue	Score	Venue	Score	Venue	Score	Venue	Score
1	SIGMOD	1.000	SIGMOD	1.000	SIGMOD	1.000	TODS	3.873
2	VLDB-J	0.758	VLDB-J	0.513	VLDB-J	0.712	SIGMOD	3.496
3	VLDB	0.709	TODS	0.503	VLDB	0.710	VLDB	2.635
4	DBPL	0.658	VLDB	0.496	EDBT	0.496	PODS	2.227
5	DMKD	0.654	DBPL	0.496	TODS	0.474	WebDB	2.174
6	TODS	0.647	wedDB	0.432	WebDB	0.471	VLDB-J	2.008
7	PODS	0.642	EDBT	0.419	PODS	0.462	EDBT	1.998
8	EDBT	0.635	ICDT	0.410	DKD	0.385	DBPL	1.967
9	WebDB	0.607	SSD	0.361	DBPL	0.379	ICDT	1.917
10	DPD	0.562	SIGMOD Rec.	0.354	ICDE	0.348	DMKD	1.668
11	ICDT	0.557	FODO	0.352	FODO	0.316	SSD	1.654
12	SSD	0.554	PODS	0.343	DMKD	0.308	ICDE	1.650
13	SIGMOD Rec.	0.490	DPD	0.342	SIGMOD Rec.	0.306	DNIS	1.599
14	ICDE	0.473	ICDE	0.319	SSD	0.297	SIGMOD Rec.	1.465
15	FODO	0.470	DMKD	0.284	ICDT	0.289	DASFAA	1.415
16	CoopIS	0.435	DS	0.257	DPD	0.287	FODO	1.405
17	DS	0.404	CoopIS	0.230	KDD	0.268	DPD	1.380
18	DASFAA	0.383	Inf. Syst.	0.229	DAWAK	0.225	CoopIS	1.377
19	DKD	0.371	DKD	0.228	DNIS	0.197	RIDE	1.342
20	CIKM	0.370	DNIS	0.218	Sigkdd Exp.	0.191	DKD	1.305

Table 2: Ranking results (Seed = SIGMOD conference)

3.2.1 Seed-Based Measure

Three approaches of the seed-based goodness measure – naive, fair, and unfair – are tested. According to the seed-based measure, seed papers are "good" articles that form the set $Seed_P$, and authors of seed papers are considered as authoritative authors to form the set $Seed_A$. Corresponding venue ranking results are shown in Table 1 for the VLDB case, Table 2 for the SIGMOD case, and Table 3 for the top-10% case. Note that, in the naive and the fair version of the seed-based approaches, the goodness score of $Seed_P$ (e.g., VLDB in Table 1) always equals to 1. However, this is not true for the unfair version. By using the normalization factor $\triangle$ defined in Definition 2, we normalize scores for all the venues so that the goodness score for the seed venue $Seed_P$ equals to 1.

3.2.2 Browsing-Based Measure

For a given $Seed_P$, we extract all the authors of seed papers and set the value of the non-uniform vector $\mathbf{w}$ in the paper browsing model according to Equation 5. The goodness for each article is then calculated based on the browsing pattern model according to Equation 6. The final goodness value of a publication venue is determined according to the definition of the venue goodness (Equation 1). Experimental results are shown in Tables 1, 2, and 3 (all the 4-th columns).

The browsing-based goodness scores shown in tables are after normalization as follows: we first suppose that a reader randomly selects an article in the collection to read. The probability that an article is being randomly selected is calculated as $\frac{1}{N}$, where N is the total number of articles in our collection. Then, the raw score of each venue is divided by this randomly selected probability and the results are shown in tables. After the normalization, the score of a venue stands for the ratio of the probability that an article in the venue being selected by a reader according to the paper-browsing model to the probability that the paper being randomly selected. The bigger the score of a venue is, the larger probability that a reader will select an article from the venue.

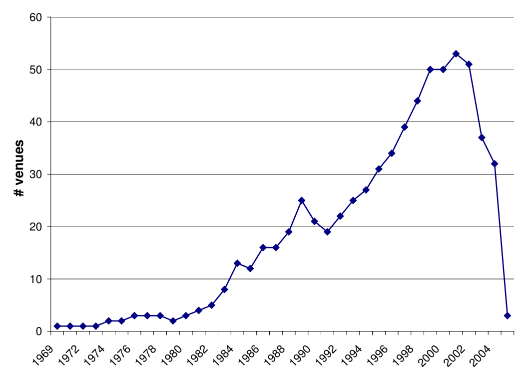

Figure 3: # of venues per year in DBLP-ACM.

The browsing-based goodness score is entirely determined by the quality of the articles in a venue in terms of the reputation of authors and the number of citations. Note that the meaning of the score is different from the probability of a "venue" being selected by a reader. The probability of a venue being selected is influenced by the number of articles contained in the venue. In other words, the probability of a venue being selected by a reader is going to bias towards the venues that contain larger number of articles, and may not truly reflect the quality of a venue.

3.2.3 Comparison to Impact Factor

We rank the 86 DBLP-ACM venues according to ISI IF method. The IF is calculated based on a three-year period.

	Naive		Fair		Unfair		Browsing	
Rank	Venue	Score	Venue	Score	Venue	Score	Venue	Score
1	TODS	0.764	TODS	0.666	SIGMOD	0.706	TODS	3.383
2	PODS	0.759	PODS	0.634	VLDB	0.702	SIGMOD	2.425
3	VLDB-J	0.758	SIGMOD	0.578	VLDB-J	0.636	VLDB	2.152
4	SIGMOD	0.749	WebDB	0.574	PODS	0.622	PODS	1.993
5	VLDB	0.724	VLDB	0.552	WebDB	0.592	WebDB	1.912
6	WebDB	0.714	DBPL	0.549	TODS	0.574	TOIS	1.769
7	DBPL	0.705	VLDB-J	0.508	EDBT	0.549	EDBT	1.740
8	EDBT	0.688	EDBT	0.473	DBPL	0.518	ICDT	1.655
9	IQIS	0.625	ICDT	0.469	DKD	0.466	DBPL	1.646
10	ICDT	0.603	TOIS	0.444	ICDT	0.405	VLDB-J	1.638
11	SSD	0.590	SSD	0.406	ICDE	0.401	DKD	1.496
12	TOIS	0.544	DKD	0.398	SIGMOD Rec.	0.385	ICDE	1.441
13	DKD	0.543	IQIS	0.375	FODO	0.335	SSD	1.379
14	DPD	0.531	ICDE	0.348	SSD	0.283	FODO	1.230
15	FODO	0.513	DPD	0.335	DPD	0.278	RIDE	1.193
16	ICDE	0.501	SIGMOD Rec.	0.328	KDD	0.276	DAWAK	1.184
17	KDD	0.488	FODO	0.327	TOIS	0.273	SIGMOD Rec.	1.170
18	DS	0.483	DS	0.301	DAWAK	0.271	DS	1.150
19	SIGMOD Rec.	0.464	KDD	0.295	DMKD	0.232	SSDBM	1.136
20	CIKM	0.453	TISSEC	0.286	TKDE	0.225	DPD	1.121

Table 3: Ranking results (Seed = top-10% papers)

For example,

Definition 4 (2003 IF of Journal X) *The IF of a journal X is: $IF_X = \frac{A}{B}$, where A is the number of times that articles (that were published in 2001-2002) were cited in indexed journals during 2003, and B is the total number of articles published in 2001-2002.* □

The IF is only calculated for journals, excluding other outlets such as conferences. Since DBLP-ACM contains substantial number of non-journal venues, their IF values were not readily available. Therefore, using the same formula, we calculated IF values for all non-journal venues in DBLP-ACM as follows: According to Figure 3, year 2001 contains the largest number (53) of unique venues. We then calculate the 2002 IF which will cover the largest number of venues in DBLP-ACM. Ranking result is show in Table 6. Notice that, there is a big drop in number of venues in the graph of Figure 3 after 2002. This is because the DBLP data we use are up to early 2005. It takes time for DBLP to update and add new publications into its database. However, we can trust the data before 2002 as shown in the graph.

Although ranks in Table 6 look reasonable, note that IF ranks venues according to years. However, none of our proposed measures consider time. Therefore, if we compare the ranking result of our proposed measures with that of IF, we would expect substantial difference. For example, consider the case of ACM Transaction of Databases (TODS), a premier journal in Database community. Given three different seeds, all results of the browsing-based measure show that TODS is ranked as the first. However, IF ranks TODS 9-th in Table 6. In order to mitigate this variation, we repeat the experimentation using a slightly-modified IF formula such that:

Definition 5 (Modified IF of Journal X) *The modified IF of a journal X is: $IF_X = \frac{C}{D}$, where C is the number of times that articles of X were cited, and D is the total number of articles in X.* □

The new ranking using this modified formula is shown in Table 7. The table is created using 2002 as the start point and includes all data (not just 2000 and 2001). It appears that TODS has been cited a lot. However, many of these citations happen after the cited TODS papers have been published for more than two years. Therefore, if we do not consider time constraints, TODS is ranked No. 1. However, if we only count the citations within the short period of 2 years after publishing, TODS is ranked much lower.

In order to objectively show that ranks in Tables 1, 2, and 3 are meaningful, we use Spearman's rank correlation coefficient ρ, and access the relationship between two ranking results. Let X and Y be two ranked sequences. Then, ρ is given by:

$$\rho = 1 - \frac{6\Sigma D^2}{N(N^2-1)}$$

where D is the difference between the ranks of corresponding values of X and Y, and N is the number of pairs of values.

To test if the comparison results are significant enough to make any conclusion, we did the significance test. The Spearman critical value table ends with N = 30. For number of pairs of data larger than 30, two ways can be taken to approximately test the significance.

1. For $N > 30$, use the critical value from Pearson's table as an approximation;
2. Use t test, where

$$t = \frac{\rho}{\sqrt{(1-\rho^2)/(N-2)}}$$

Rank	Venue	Score
1	SIGMOD	1.023121387
2	VLDB	0.852071006
3	PODS	0.833333333
4	ICDT	0.827586207
5	KDD	0.616766467
6	TOIS	0.571428571
7	DBPL	0.5
8	DL	0.342592593
9	TODS	0.32
10	ICDE	0.316666667
11	WebDB	0.307692308
12	FOIKS	0.277777778
13	SSDBM	0.274193548
14	CIKM	0.260869565
15	SSD	0.260869565
16	EDBT	0.243243243
17	DOLAP	0.230769231
18	DKD	0.225806452
19	VLDB-J	0.2
20	SIGKDD Exp.	0.166666667

Table 6: Venues of DBLP-ACM sorted by their 2002 Impact Factors.

We implemented the significance test using both methods. For the t test, we use $\alpha = 0.01$, and set the null hypothesis H_0 as "*There is no strong positive rank order relationship between the naive/fair/unfair seed-based/browsing-based measure result and the impact factor result.*"

The first significance test results against the ranking by IF measure of Table 6 are shown in Table 8. Here, ρ_s is the Spearman coefficient, t_s is the calculated t coefficient, $t = 2.396$ is the one tail test t value where the degree of freedom is 54 according to the t-distribution table, and ρ_p is the critical value for Pearson's coefficients where the degree of freedom is 54. According to experimental results, we reject the null hypothesis H_0 and accept that there is a *strong* positive rank order correlation between the venue ranking results by our proposed measures and those by the IF method.

Next, the results of the same significant test against the *modified* IF measure are shown in Table 9. From the table, again, we can conclude that the ranking results by all our proposed measures are positively correlated with the ranking result by the modified IF method.

3.2.4 Comparison to Known Ranks

Finally, we compare the ranking results by the proposed measures to the known ranks in the community. We used the CS conference ranking from the following site:

`http://www.cs-conference-ranking.org/`

They used the so-called "Estimated Impact of Conference (EIC)" measure that consists of:

1. CP: 40% citation of papers

Rank	Venue	Score
1	TODS	7.835294118
2	SIGMOD	5.870910173
3	VLDB	4.892446634
4	PODS	4.722077922
5	WebDB	2.833333333
6	TOIS	2.00660793
7	ICDT	1.975609756
8	EDBT	1.81232493
9	ICDE	1.772783251
10	VLDB-J	1.523972603
11	DBPL	1.48241206
12	SSD	1.463576159
13	DKD	1.459302326
14	KDD	1.447080292
15	TISSEC	1.18487395
16	TKDE	1.025936599
17	SIGMOD Rec.	0.883399209
18	CIKM	0.756410256
19	FODO	0.752066116
20	JIIS	0.750778816

Table 7: Venues of DBLP-ACM sorted by their *modified* Impact Factors.

Seed	Pair	ρ_s	t_s	Conclusion
VLDB	(naive, IF)	0.5081	4.34	reject H_0
	(fair, IF)	0.4914	4.15	reject H_0
	(unfair, IF)	0.5765	5.18	reject H_0
	(browsing, IF)	0.5684	5.08	reject H_0
SIGMOD	(naive, IF)	0.5964	5.4597	reject H_0
	(fair, IF)	0.56992	5.10	reject H_0
	(unfair, IF)	0.5995	5.5	reject H_0
	(browsing, IF)	0.6154	6.15	reject H_0
Top-10%	(naive, IF)	0.7126	0.71	reject H_0
	(fair, IF)	0.7204	7.63	reject H_0
	(unfair, IF)	0.7075	7.36	reject H_0
	(browsing, IF)	0.7244	8.05	reject H_0

Table 8: Significance test: $\alpha = 0.01, t = 2.396, \rho_p = 0.354$ (56 venues, using t test and Pearson's critical value). Note that this comparison is made against the rank by IF measure of Table 6.

2. RR: 20% quality of referees' reports
3. RS: 20% availability of resources for students (funds for travel, fees, hotel)
4. IN: 10% indexing
5. JA: 10% percentage of conference papers accepted and appeared in reputable journals

The Top-80 ranks of Database community using the EIC measure is available in the web site. From the top-80, we selected the top-20 venues, which are overlaps with DBLP-ACM, for experiments, as shown in Table 10. Table 11 summarizes the results of the significance test using the ranks in Table 10. Unlike the previous two significance tests, here, we get quite different results. That is, measures using VLDB or SIGMOD as the seed did *not* show strongly correlated ranks to the EIC ranking. However, three (naive, unfair,

Seed	Pair	ρ_s	t_s	Conclusion
VLDB	(naive, IF')	0.70589	7.32	reject H_0
	(fair, IF')	0.75501	8.46	reject H_0
	(unfair, IF')	0.80553	9.99	reject H_0
	(browsing, IF')	0.73262	7.91	reject H_0
SIGMOD	(naive, IF')	0.794778	9.62	reject H_0
	(fair, IF')	0.815609	10.36	reject H_0
	(unfair, IF')	0.812581	10.24	reject H_0
	(browsing, IF')	0.818105	10.45	reject H_0
Top-10%	(naive, IF')	0.84461	11.57	reject H_0
	(fair, IF')	0.91640	16.82	reject H_0
	(unfair, IF')	0.88301	13.83	reject H_0
	(browsing, IF')	0.88930	14.29	reject H_0

Table 9: Significance test: $\alpha = 0.01, t = 2.396, \rho_p = 0.354$ (56 venues, t test and using Pearson's critical value). Note that this comparison is made made against the rank by the modified IF measure, IF', of Table 7.

Rank	Venue	EIC Score
1	SIGMOD	0.99
2	VLDB	0.99
3	ICDE	0.97
4	ICDT	0.94
5	PODS	0.94
6	FODO	0.92
7	ER	0.91
8	CIKM	0.90
9	DOOD	0.90
10	DEXA	0.90
11	SSDBM	0.90
12	COMAD	0.90
13	EDBT	0.90
14	VDB	0.88
15	SSD	0.88
16	CoopIS	0.88
17	DS	0.86
18	DAWAK	0.86
19	MDM	0.83
20	ARTDB	0.83

Table 10: Top-20 conferences by the EIC measure.

and browsing) out of four measures using top-10% articles as the seed shows strong correlation to the EIC ranking by rejecting the null hypothesis. Both facts – top-10% articles with more number of citations are used as the seed as well as that EIC counts 40% weight of the citations of papers – may be attributable to this result [3]

4. RELATED WORK

ISI's Impact Factor has been used in many applications – journal quality estimation, promotion and tenure of scholars, etc. Since the introduction of the IF, however, it has been heavily criticized (e.g., its sole dependency on citation counts [13]). To remedy those issues, many alternatives have been proposed such as H-index to measure the impact of individual scholars [5], the measure for ranking documents retrieved from a digital library [7], case study in database field [12], or the PageRank-like measure [3].

[3]We thank the anonymous reviewer for his/her point on this matter. However, more study needs to be done to draw a conclusion.

Seed	Pair	ρ_s	Conclusion
VLDB	(naive, EIC)	0.2632	accept H_0
	(fair, EIC)	0.2797	accept H_0
	(unfair, EIC)	0.4964	accept H_0
	(browsing, EIC)	0.5046	accept H_0
SIGMOD	(naive, EIC)	0.5273	accept H_0
	(fair, EIC)	0.6120	reject H_0
	(unfair, EIC)	0.5893	reject H_0
	(browsing, EIC)	0.5294	accept H_0
Top-10%	(naive, EIC)	0.5439	reject H_0
	(fair, EIC)	0.5273	accept H_0
	(unfair, EIC)	0.6471	reject H_0
	(browsing, EIC)	0.5501	reject H_0

Table 11: Significance test with the EIC ranking (critical value of Spearman's rank correlation coefficient $\rho_s = 0.534$ for 20 pairs).

In particular, [3] proposes to use the PageRank algorithm [11] to distinguish the "quality" of citations and hence improve the IF calculation. However, the improved IF still does not consider the different citation patterns between journal and conference types. [2] generates networks of journal relationships from the citation and download data, and determines the journal impact ranking from the networks using social network centrality measures. For instance, according to their scheme, a more frequently downloaded article is likely to have more impact. Again, they do not consider the application scope problem of IF. However, they draw an interesting conclusion that "a unique aspect of general journal impact is not captured by the Impact Factor" and further question the validity of the IF as the sole assessment of the journal impact.

Finally, recent study such as [10] introduces the topic modeling to further complement the citation-based bibliometric indicators, producing more fine-grained impact measures.

5. CONCLUSION

Although having many benefits, the ISI's Impact Factor measure is not suitable for cases like:

- Young or emergent journals whose historical citation statistics are not readily available or mature, rendering citation-based metrics inapplicable.
- Well-established conferences whose citation statistics takes time to accumulate. A recent study of the major database conferences and journals between 1994 and 2003 shows many of the citations reach back five and more years [12], rendering three-year window of IF inadequate.
- Data sets from digital libraries whose quality of meta data extraction mechanism is less than perfect.

In this paper, toward these issues of IF, we proposed an array of alternative measures to judge the goodness of venues. The seed-based measure uses the simple author meta data which are much more accurate and easier to extract and parse than the citation meta data. Experimental results show that our measure using the simple author meta data produces positively correlated results with that of the IF method, however, the drawbacks caused by citation-based metrics are eliminated. Three variations of the seed-based measure provide a broader applicability of the measure. The browsing-based measure, which uses both the citation and author meta data, takes an article reader's behavior into account. Thus this measure yields more realistic rankings of venues from the reader's point of view.

During the research work, we are inspired to further extend the venue ranking measures in several directions. First, our current work proves the applicability and efficacy of the new measures in the Database domain. We plan to apply the measures to other domains and check their performances. Second, since venue ranking is a quite subjective problem, it is interesting to do user evaluations and find out how well the objective measures match with the subjective evaluations. In other words, we want to find out how well the objective measures reveal the reality. Last, with the existence of various meta data, we envision a unified model, which properly combine various information, to yield more realistic ranking results.

6. ACKNOWLEDGEMENT

The research of Dongwon Lee was partially supported by Microsoft SciData Award (2005) and IBM Eclipse Innovation Award (2006). We sincerely thank for constructive comments of anonymous reviewers.

7. REFERENCES

[1] M. Amin and M. Mabe. Impact factors: Use and abuse. *Perspectives in Publishing*, October 2000.

[2] J. Bollen, H. V. de Sompel, J. A. Smith, and R. Luce. Toward alternative metrics of journal impact: A comparison of download and citation data. *Inf. Process. Manage.*, 41(6):1419–1440, 2005.

[3] J. Bollen, M. A. Rodriguez, and H. Van de Sompel. Journal status. *ArXiv Computer Science e-prints.*, January 2006.

[4] T. H. Haveliwala. Topic-sensitive pagerank. In *WWW '02: Proceedings of the 11th international conference on World Wide Web*, pages 517–526, New York, NY, USA, 2002. ACM Press.

[5] J. E. Hirsch. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102(46):16569–16572, 2005.

[6] A. N. Langville and C. D. Meyer. Deeper inside pagerank. *Internet Mathematics*, 1(3):335–380, 2003.

[7] B. Larsen and P. Ingwersen. Using citations for ranking in digital libraries. In *ACM/IEEE joint conference on Digital libraries (JCDL)*, pages 370–371, Chapel Hill, NC, USA, 2006.

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

[9] M. H. MacRoberts and B. R. MacRoberts. Problems of citation analysis: A critical review. *Journal of The American Society for Information Science.*, 40(5):342–329, 1989.

[10] G. Mann, D. Mimno, and A. McCallum. Bibliometric impact measures leveraging topic analysis. In *ACM/IEEE joint conference on Digital libraries (JCDL)*, pages 65–74, Chapel Hill, NC, USA, 2006.

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

[12] E. Rahm and A. Thor. Citation analysis of database publications. *SIGMOD Rec.*, 34(4):48–53, 2005.

[13] S. Saha, S. Saint, and D. A. Christakis. Impact factor: a valid measure of journal quality? *Journal of the Medical Library Association*, 62(1):42–46, 2003.

[14] P. O. Seglen. Why the impact factor of journals should not be used for evaluating research. *British Medical Journal*, 314(7079):498–502, February 1997.

[15] A. F. J. vanRaan. Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics*, 62(1):133–143, January 2005.