Annales Univ. Sci. Budapest., Sect. Comp. 42 (2014) nn–nnn

Evaluate scientific publications by N-linear ranking model

Vu Le Anh (Ho Chi Minh city, Vietnam)
Hai Vo Hoang (Ho Chi Minh city, Vietnam)
Hieu Le Trung (Da Nang, Vietnam)
Kien Le Trung (Thua Thien Hue, Vietnam)
Jason J. Jung(Gyeongsan, Korea)

Communicated by ???

(Received ???; accepted ???)

Abstract. Ranking has been applied in many domains using recommendation systems such as search engine, e-commerce, and so on. We will introduce and study N-linear mutual ranking, which can rank n classes of objects at once. The ranking scores of these classes are dependent to the others. For instance, PageRank by Google is a 2-linear ranking model, which ranks the web-pages and links at once. Particularly, we focus to N-star ranking model and demonstrate it in ranking conference and journal problems. We have conducted the experiments for our models versus classical ones. The experiments are based on the DBLP dataset, which contains more than one million papers, authors and thousands of conferences and journals in computer science. The experimental results show that N-star ranking model evaluates everything much more detail based on the context of their relationships.

Key words and phrases: N-star ranking, Markov chain, PageRank, Academic ranking, Conference ranking, Ranking algorithms, Prolific ranking, Recommendation systems, Bibliographical database, DBLP
2010 Mathematics Subject Classification: Please use the *2010 Mathematics Subject Classification*
This work is the extend version of the paper "A General Model for Mutual Ranking Systems" in *Intelligent Information and Database Systems Lecture Notes in Computer Science* **Volume 8397**, 2014, pp 211-220

1. Introduction

Ranking is an interesting but difficult problem on many information processing systems. With a large amount of information, the systems need to adapt efficient ranking schemes to sort out (or to select) only the information which are highly relevant to the users' contexts. Particularly, in the context of *bibliometrics*, a set of given entities can be quantified to compare several evaluation indicators (e.g., popularity and reputation). For example, impact factors (IF) of international journals can be measured by taking into account how many times the papers in the corresponding journals have been cited.

In this work we focus on a system of ranking classes. Their ranking scores have mutual dependencies, which can expressed by a system of linear equations. Let us explain the ideas by two examples.

PageRank. PageRank is very well-known ranking for website [14], which were applied in Google search engine. We rewrite the original formula by a system of two generic linear equations describing the mutual dependency of ranking of two classes, Web and $Link$.

$$Link \longleftarrow 100\% \times Web \tag{1.1}$$

$$Web \longleftarrow 85\% \times Link + 15\% \times Random \tag{1.2}$$

Equation 1.1 says that the rank score of a link is determined by the rank score, which is the link from. Equation 1.2 says that the rank score of a web is determined by 85% from the rank score of links, which refers to the web; and 15% from randomness. Thus the rank scores of webs and links are mutual dependent. Moreover, we prove that there exists only one rank scores satisfy the above system of linear equations.

Ranking scientific publication. We propose a model for ranking 4 classes Authors ($Author$), Publications (Pub), Conference ($Conf$) and Citations ($Cite$). Their relationships are described by following a system of four generic linear equations.

$$Author \longleftarrow 100\% \times Pub \tag{1.3}$$

$$Conf \longleftarrow 100\% \times Pub \tag{1.4}$$

$$Cite \longleftarrow 100\% \times Pub \tag{1.5}$$

$$\begin{aligned} Pub \longleftarrow\ & 30\% \times Author + 30\% \times Conf \\ & + 30\% \times Cite + 10\% \times Random \end{aligned} \tag{1.6}$$

Equation 1.3 says that the rank score of each author is determined by the rank scores of his publications. Equation 1.4 says that the rank score of each conference is determined by the rank scores of it publications. Equation 1.5 says that the rank score of each citation is determined by the rank score of the publication, which is the owner of the citation. Equation 1.6 says that the rank score of each publication is determined by 30% from the rank score of its authors; 30% from the rank score of its conference; 30% from the rank scores of the cited-to citations; and 10% from randomness.

Both of above ranking systems are described by systems of linear equations called *N-linear ranking models*. Here are the key questions: *Does the system of linear equations have a unique solution? How can we compute the solution?* And *how do the models work in realistic ranking systems?* We solve only a part of problems by studying a special case of N-linear ranking model, *N-star ranking model*. Both of two examples are N-star ranking models and there exists unique solutions. Moreover we can estimate it by a loop of computing the linear function.

The main contribution and outline of this paper are as follows.

N-linear ranking model. We describe the background of the N-linear ranking model (Section 2). N-linear ranking model is the system of N ranking scores of N classes. The rank scores are depend on others by a linear constraint system (Subsection 2.1). We introduce the affect and reflect relation between classes (Subsection 2.1). We explain these definitions in detail by the case study of PageRank (Subsection 2.2).

N-star ranking model. We define the N-star ranking model as a N-linear ranking model in which there exists a core class (Section 3). We prove that there exists unique N-star ranking model which satisfy a given linear constraint system (Proposition 3.2). We show that PageRank is a 2-star ranking model (Proposition 3.1). Finally, we describe the algorithm to compute scores of classes based on the linear constraint system.

Ranking bibliographical database. We study two N-star ranking models for the author, publication and conference ranking problem in different contexts (Section 4). The first model is general N-star ranking model for 4 classes: authors, publications, conferences, citations (Definition4.1). In the second model, we simplify the conditions by the assumption that everything is equal (Definition 4.2).

Experiments. We do the experiments for the simple N-linear ranking model of authors, publications and conference ranking (Section 5). We have designed the three different datasets to adapt the limit of computing resources (Subsection 5.1). The datasets are classified into two contexts: with/ without citations. We propose the models and the measurements for comparing different ranking scores in both contexts (Subsection 5.2). We show the results and have discussions over datasets (Subsection 5.3). Our results are quite different from the naive one's and provide us some interesting things.

Related works and conclusion. We discuss the related works of N-linear ranking model (Section 6). We discuss the power and applicable ability of N-linear ranking model (Section 7).

2. Backgrounds

2.1. N-linear mutual ranking system

The couple $(\mathcal{A}, R)$ is called a *ranking system* if (i) $\mathcal{A} = \{a_1, \ldots, a_n\}$ is a finite set, and (ii) $R : \mathcal{A} \to [0; +\infty)$. $\mathcal{A}$ is called a *class*, $a \in \mathcal{A}$ is called an *object of the class* $\mathcal{A}$, and R is called a *score* of $\mathcal{A}$. R is *positive* if $R(a) > 0 \ \ \forall a \in \mathcal{A}$. $n = |\mathcal{A}|$ is the *size* of $\mathcal{A}$.

Definition 2.1. $\Omega = \{(\mathcal{A}_i, R_i)\}_{i=1}^N$ *is called a* N-linear mutual ranking system *described by a system* $\{\alpha_{ij}, \beta_i, I_i, W_{ij}\}$ *if* $(\mathcal{A}_i, R_i)$ *is a ranking system and* $\alpha_{ij}, \beta_i \in [0; +\infty)$, $I_i = (t_u^{(i)})_{n_i}$, $n_i = |\mathcal{A}_i|$, *is a* n_i*-dimensional normalized nonnegative real number vector ,* $W_{ij} = (\omega_{kl}^{(ij)})_{n_i \times n_j}$ *is a nonnegative real number and normalized columns matrix such that for all* $i = 1, \ldots, N$,

$$\sum_j \alpha_{ij} + \beta_i = 1 \qquad \textit{and} \qquad R_i = \sum_{j=1}^N \alpha_{ij} W_{ij} R_j + \beta_i I_i.$$

$\{\alpha_{ij}, \beta_i, I_i, W_{ij}\}$ is called a *linear constraint system* of Ω.

Note that, generally since $\sum_i \alpha_{ij} + \frac{1}{N} \sum_j \beta_j$ is different one, a N-linear mutual ranking system is not a Markov chain. Let a_{iu}, a_{jv} be objects in $\mathcal{A}_i, \mathcal{A}_j$ respectively. Suppose $\mathcal{C}^*(a_{iu}, a_{jv}) = \alpha_{ij}\omega_{uv}^{(ij)}$. From the definitions, we have:

$$R_i(a_{iu}) = \sum_{j=1}^N \sum_{v=1}^{n_j} \alpha_{ij}\omega_{uv}^{(ij)} R_j(a_{jv}) + \beta_i t_u^{(i)} = \sum_{j=1}^N \sum_{v=1}^{n_j} \mathcal{C}^*(a_{iu}, a_{jv}) R_j(a_{jv}) + \beta_i t_u^{(i)}$$

a_{jv} is called *affect to* a_{iu} (denoted by $a_{jv} \to a_{iu}$) if $\mathcal{C}^*(a_{iu}, a_{jv}) > 0$. Class $\mathcal{A}_i$ is called *total affect and reflect directly to* class $\mathcal{A}_j$ (denoted by $\mathcal{A}_i \to \mathcal{A}_j$) if $\forall a_{jv} \in \mathcal{A}_j$: $\exists a_{iu_1}, a_{iu_2} \in \mathcal{A}_i$: $a_{jv} \to a_{iu_2} \wedge a_{iu_1} \to a_{jv}$.

Definition 2.2. *Class* $\mathcal{A}_i$ *is called* total affect and reflect to *class* $\mathcal{A}_j$*, denoted by* $\mathcal{A}_i \rightsquigarrow \mathcal{A}_j$*, if* $\mathcal{A}_i \to \mathcal{A}_j$ *or* $\exists \mathcal{A}_k : \mathcal{A}_i \to \mathcal{A}_k \wedge \mathcal{A}_k \rightsquigarrow \mathcal{A}_j$.

2.2. PageRank

We rewrite the PageRank into a 2-linear mutual ranking system as follows:

$\mathcal{W} = \mathcal{A}_1$ is the class representing for the set of webpages. $\mathcal{L} = \mathcal{A}_2$ is the class representing for hyperlinks. For each hyperlink $l \in \mathcal{L}$ from web $u \in \mathcal{W}$ to web $v \in \mathcal{W}$, we denote $u = in(l)$ and $v = out(l)$. For each $v \in \mathcal{W}$, we denote: $IN(v) = \{l \in \mathcal{L} | v = out(l)\}$ and $N_{out}(v) = |\{l \in \mathcal{L} | v = in(l)\}|$.

PageRank[14] determined the ranking system of webpages by the following formula: $\forall v \in \mathcal{W}$,

$$R_w(v) = d \sum_{l \in IN(v), u = in(l)} \frac{R_w(u)}{N_{out}(u)} + \frac{1-d}{|\mathcal{W}|} \tag{2.1}$$

where $d \in (0,1)$ is a constant.

Suppose $W_{21} = (\delta_{kt})_{|\mathcal{L}| \times |\mathcal{W}|}$ is a matrix in which $\delta_{kt} = \frac{1}{N_{out}(w_t)}$ if $l_k \in \{l : w_t = in(l)\}$, otherwise 0. W_{21} is a nonnegative real number and normalized columns matrix. Suppose $W_{12} = (\gamma_{tk})_{|\mathcal{W}| \times |\mathcal{L}|}$ is a matrix in which $\gamma_{tk} = 1$ if $w_t = out(l_k)$, otherwise 0. We construct a 2-linear mutual ranking system on two classes $\mathcal{W}$ and $\mathcal{L}$ as follows: Let $\bar{R}_w$ and $\bar{R}_l$ be scores on the classes $\mathcal{W}$, $\mathcal{L}$ respectively. They are satisfied:

$$\bar{R}_l = W_{21} \bar{R}_w \qquad and \qquad \bar{R}_w = d W_{12} \bar{R}_l + (1-d) I_{|\mathcal{W}|}, \tag{2.2}$$

where $I_{|\mathcal{W}|}$ denotes the $|\mathcal{W}|$-dimensional vector in which all its elements are $1/|\mathcal{W}|$. It is not difficult to see that (2.2) confirms: for all web $v \in \mathcal{W}$,

$$\bar{R}_w(v) = d \sum_{l \in IN(v), u = in(l)} \frac{\bar{R}_w(u)}{N_{out}(u)} + \frac{1-d}{|\mathcal{W}|}.$$

Since the equation (2.1) has the unique solution which is the PageRank score (see in [14]), $\bar{R}_w$ is the PageRank score R_w. Vice versa, if R_w is a solution of (2.2), R_w should be $\bar{R}_w$. Thus, the PageRank score R_w is totally determined by the equation (2.2), or in other words, PageRank can be presented as the two-linear ranking system described by (2.2).

Note that, since for each link $l \in \mathcal{L}$, let $u = in(l)$ and $v = out(l)$ then web u affects to link l, $(u \to l)$ and link l affects to web v, $(l \to v)$. Therefore, the class $\mathcal{W}$ total affect and reflect directly to the class $\mathcal{L}$, $\mathcal{W} \rightsquigarrow \mathcal{L}$.

3. N-star Ranking Model

Definition 3.1. *Let $\Omega = \{(\mathcal{A}_i, R_i)\}_{i=1}^{N}$ be a N-linear mutual ranking system. Ω is called a N-*star ranking *if*

1. $\exists i: \ \mathcal{A}_i : (\beta_i > 0) \wedge (I_i$ *is positive* $) \wedge (\forall \mathcal{A}_j (j \neq i) : \mathcal{A}_i \rightsquigarrow \mathcal{A}_j)$.
2. $\forall j \neq i: \ \alpha_{j1} = 1$.

$\mathcal{A}_i$ *is called a* core *of the system* Ω.

If $\Omega = \{(\mathcal{A}_i, R_i)\}_{i=1}^{N}$ is a N-star ranking system described by a linear constraint system $\{\alpha_{ij}, \beta_i, I_i, W_{ij}\}$, $\{\alpha_{ij}, \beta_i, I_i, W_{ij}\}$ is called N-*star constraint system* of Ω.

Proposition 3.1. *PageRank is a 2-star ranking system.*

Proof. Because $1 - d > 0$ and $\mathcal{W} \rightsquigarrow \mathcal{L}$, $\mathcal{W}$ is the core of PageRank. The second condition is clear since $\alpha_{21} = 1$. Hence, PageRank is a 2-star ranking system.

The ranking scores are determined by the N-star constraint system and the classes.

Proposition 3.2. *Suppose the classes $\{\mathcal{A}_i\}_{i=1}^{N}$ and the N-star constraint system $\{\alpha_{ij}, \beta_i, I_i, W_{ij}\}$ are given. There exists a unique $\{R_i\}_{i=1}^{N}$ in which R_i is a score on $\mathcal{A}_i$, such that $\Omega = \{(\mathcal{A}_i, R_i)\}_{i=1}^{N}$ is a N-star ranking described by $\{\alpha_{ij}, \beta_i, I_i, W_{ij}\}$ and for all i, $\sum_{a \in \mathcal{A}_i} R_i(a) = 1$.*

Proof. Assuming without loss of generality that $\mathcal{A}_1$ is a core of a N-star ranking system described by $\{\alpha_{ij}, \beta_i, I_i, W_{ij}\}$, the sequence of scores $R_1, \ldots, R_N$ satisfy the following equations:

$$R_1 = WR_1 \qquad \textit{and} \qquad R_i = W_{i1} R_1 \qquad \forall i = 2, \ldots, N, \tag{3.1}$$

where

$$W = \alpha_{11} W_{11} + \alpha_{12} W_{12} W_{21} + \cdots + \alpha_{1N} W_{1N} W_{N1} + \beta_1 \mathbf{I}_1 \tag{3.2}$$

and $\mathbf{I}_1$ is the $(n_1 \times n_1)$-matrix which its columns are I_1. It is not difficult to infer that because W_{1i} and W_{i1} are transition matrices (i.e. nonnegative and normalized columns matrices), the new square matrix $W_{1i} W_{i1}$ is a stochastic matrix (i.e. a transition and square matrix). The matrices W_{11} and $\mathbf{I}_1$ are also stochastic matrices. Since $\mathbf{I}_1$ has positive entries, $\beta_1 > 0$ (because $\mathcal{A}_1$ is the core), and $\sum_j \alpha_{1j} + \beta_1 = 1$, the matrix W is also a stochastic matrix with positive entries. The Perron-Frobenius theorem (see in [7, 9]) confirms that there exists a unique score R_1 with $\sum_{a\in\mathcal{A}_1} R_1(a) = 1$ such that

$$R_1 = WR_1.$$

From (3.1), the unique existence of R_1 infers the unique existences of $R_2, \ldots, R_N$. Moreover, since $W_{21}, \ldots, W_{N1}$ are normalized columns and $\sum_{a\in\mathcal{A}_1} R_1(a) = 1$, we have $\sum_{a\in\mathcal{A}_i} R_i(a) = 1$ for all $i = 2, \ldots, N$. The proposition is proven.

The ranking scores are computed by following algorithm:

Algorithm : Finding the sequence scores $\{R_i\}_{i=1}^N$

Input : $\alpha_{ij},\ \beta_i,\ W_{ij}, I_i$

Output : $\{R_i\}_{i=1}^N$

begin

Check the N-star ranking model with the core $\mathcal{A}_1$

Let

$$W \leftarrow \alpha_{11}W_{11} + \sum_{i=2}^{N} \alpha_{1i}W_{1i}W_{i1} + \beta_1\mathbf{I}_1$$

Initialize $R_1^{(0)}$: uniform distribution, $k = 0$

repeat

$k = k + 1$

Update $R_1^{(k)} \leftarrow WR_1^{(k-1)}$

until $\|R_1^{(k)} - R_1^{(k-1)}\| \le$ a stopping criterion

Let $R_1 = R_1^{(k)}$ and $R_i = W_{i1}R_1,\quad i = 2, \ldots, N$

end

4. Ranking Authors, Publications and Conferences

In this section, we apply the N-star ranking model for constructing a model to evaluate authors, publications and conferences (journals) in the world of science. Concretely, we consider a four-star ranking model corresponding with four ranking systems: $(\mathcal{A}, R_a)$ - $\mathcal{A}$ is a set of all scientists which has publications, $(\mathcal{P}, R_p)$ - $\mathcal{P}$ is a set of all publications, $(\mathcal{C}, R_c)$ - $\mathcal{C}$ is a set of all sciential conferences and sciential journals, and $(\mathcal{L}, R_l)$ - $\mathcal{L}$ is a set of all citations between publications. $R_a,\ R_p,\ R_c$ and R_l are the scores for each classes $\mathcal{A},\ \mathcal{P},\ \mathcal{C}$ and $\mathcal{L}$, respectively.

For each citation $l \in \mathcal{L}$, $u = in(l)$ and $v = out(l)$ if l is from $u \in \mathcal{P}$ to $v \in \mathcal{P}$. For each publication $v \in \mathcal{P}$, we denote: $IN(v) = \{l \in \mathcal{L} | v = out(l)\}$; $OUT(v) = \{l \in \mathcal{L} | v = in(l)\}$ $N_{out}(v) = |OUT(v)|$. If a publication cites no where, we assume that it cites to all publications; $C(v) \in \mathcal{C}$ is the conference of v; $A(v) \subseteq \mathcal{A}$ is the set of authors of v. For each author $a \in \mathcal{A}$, $P(a) = \{v \in \mathcal{P} | a \in A(v)\}$ is a set of publications of a. For each conference $c \in \mathcal{C}$, $P_c(c) = \{v \in \mathcal{P} | c = C(v)\}$ is a set of publications published in c.

The 4-star ranking system model for ranking authors, publications, conferences and citations is constructed based on some following ideas:

1. The score of an author depends only on his publications, and each publications affects to all of its authors:
   $\forall a \in \mathcal{A}, p \in \mathcal{P}:$

   $$R_a(a) = \sum_{p' \in P(a)} \mathcal{C}^*(a, p') R_p(p') \quad and \quad \sum_{a' \in A(p)} \mathcal{C}^*(a', p) = 1 \tag{4.1}$$

   If *a publication affects equally to its authors* (a), (4.1) is rewritten as follows:
   $\forall a \in \mathcal{A}, p \in \mathcal{P}, a' \in A(p):$

   $$\mathcal{C}^*(a', p) = \frac{1}{|A(p)|} \quad and \quad R_a(a) = \sum_{p' \in P(a)} \frac{R_p(p')}{|A(p')|} \tag{4.2}$$

2. The score of a conference depends only on its publications:

   $$\forall c \in \mathcal{C} : R_c(c) = \sum_{p' \in P_c(c)} R_p(p'). \tag{4.3}$$

3. The score of a citation depends on the citing publication, and each publications affects to all of its citations:
   $\forall l \in \mathcal{L}, p \in \mathcal{P}, p' = in(l):$

   $$R_l(l) = \mathcal{C}^*(l, p') R_p(p') \quad and \quad \sum_{l' \in OUT(p)} \mathcal{C}^*(l', p) = 1 \tag{4.4}$$

   If *a publication affects equally to its citations* (b), (4.4) is rewritten as follows:
   $\forall l \in \mathcal{L}, p \in \mathcal{P}, p' = in(l), l' \in OUT(p):$

   $$\mathcal{C}^*(l', p) = \frac{1}{N_{out}(p)} \quad and \quad R_l(l) = \frac{R_p(p')}{|N_{out}(p')|} \tag{4.5}$$

4. The score of a publication depends on its citations, its authors and its conference and randomly finding by some reader. Each conference affects to all of its publications. Each author affects to all of its publications. Hence :

   $\forall p \in \mathcal{P}, c \in \mathcal{C}, a \in \mathcal{A}, c' = C(p)$:

$$R_p(p) = \alpha_1 \sum_{l' \in IN(p)} R_l(l') + \alpha_2 \sum_{a' \in A(p)} \frac{\mathcal{C}^*(p, a')}{\alpha_2} R_a(a') + \alpha_3 \frac{\mathcal{C}^*(p, c')}{\alpha_3} R_c(c') + \beta_p I_p, \tag{4.6}$$

$$\sum_{p' \in P(a)} \mathcal{C}^*(p', a) = \alpha_2 \quad and \quad \sum_{p' \in P_c(c)} \mathcal{C}^*(p', c) = \alpha_3 \tag{4.7}$$

   where $\alpha_1, \alpha_2, \alpha_3, \beta_p > 0$ and $\alpha_1 + \alpha_2 + \alpha_3 + \beta_p = 1$, I_p is a $|\mathcal{P}|$-dimensional normalized uniform random vector.

   If *a conference affects equally to its publications* (c) and *an author affects equally to its publications* (d), the equation (4.6) and (4.7) are rewritten as follows:

   $\forall p \in \mathcal{P}, c \in \mathcal{C}, a \in \mathcal{A}, c' = C(p), p' \in P(a), p'' \in P_c(c)$:

$$\mathcal{C}^*(p', a) = \frac{\alpha_2}{|P(a)|} \quad and \quad \mathcal{C}^*(p'', c) = \frac{\alpha_3}{|P_c(c)|} \tag{4.8}$$

$$R_p(p) = \alpha_1 \sum_{l' \in IN(p)} R_l(l') + \alpha_2 \sum_{a' \in A(p)} \frac{R_a(a')}{|P(a')|} + \alpha_3 \frac{R_c(c')}{|P_c(c')|} + \beta_p I_p, \tag{4.9}$$

Definition 4.1. *The model which is described by equations (4.1), (4.3), (4.4), (4.6) and (4.7) is called the general 4-star ranking model for the ranking authors, publications and conferences problem.*

Definition 4.2. *The model which is described by equations (4.2), (4.3), (4.5), (4.8) and (4.9) is called the simple 4-star ranking model for the ranking authors, publications and conferences problem.*

Both of the general and simple 4-star ranking models are really N-star ranking in which the publication class is the central.

5. Experiments

5.1. Computing enviroments and Datasets

5.1.1. Computing environments

We aim to setup experiments to test the intuition of N-linear system. So, we decide to start with a less tiers system, from a small dataset, lacking information. Then, we escalate gradually to bigger datasets, with more tiers. We implement a C# program to compute the N-star ranking scores algorithms introduced in Sect. 3 from the DBLP data sets*, on fields of computer science. We has built program to parse DBLP dataset in XML format to extract the authors, title, and publication venue information from the guides [11, 12]. For our first preliminary work [18], we just use a limit computing resource, a desktop with Intel core duo E7500 (2.8GHz) CPU and 2GB RAM to implement our algorithm.

5.1.2. Datasets

Starting with a simple, less tier ranking model in Sect. 4, we evaluate authors, publications and only conferences in DBLP data. We mention our readers that DBLP has no information about the citations between publications. So we narrow down the 4-layer in Sect. 4 to 3-tier in case of lacking citation information. We generatedataset D_1 contains all publications for all conferences collected in DBLP for this 3-layer ranking. We also note that a DBLP publication can be both in a conference and another journal. So, we keep only the publications related to conferences and ignore the journals's ones.

Because we found too many conferences with very small publications in DBLP, we intend to study the distorting effect of small-published conferences. So we setup a no-citation dataset D_2 contain publications in conferences which have at least 300 papers over years. Of course, we must sustain the loss of many young, hot conferences which have too small number of publications in this case.

We need a citation information dataset to evaluate the 4-tier ranking system in Sect. 4. So we amend DBLP datasets with citations collected from Academic Microsoft †. Because of our limit computing and data resource, we create a small citation information D_c dataset, containing publications in conferences belonging to database field only.

*http://dblp.uni-trier.de accessed on May 2013
†http://http://academic.research.microsoft.com accessed on March 2014

Table 1. Experiment datasets.

Datasets	nPubs	nAuthors.	nConfs.	nCites
D_1	1253997	845295	3351	-
D_2	1046030	746560	949	-
D_c	77361	87354	642	156429

The statistical figures of our three experimental datasets are shown in Tab. 1.

5.2. Ranking models with/without citations

5.2.1. Models without citations

We measure two following models to evaluate the ranking results of Sect. 4 in world of science when lacking of citation information:

NPC Model. Without information about the citations between publications, we adopt a *naive model based on publication counting (NPC)* for ranking author and conference. In NPC model, for each author $a \in \mathcal{A}$ $R_a^{NPC}(a) ::= |P(a)|$ and for each conference $c \in \mathcal{C}$ $R_c^{NPC}(c) ::= |P_c(c)|$.

SD3R Model. We implemented a *Simple DBLP 3-star Ranking (SD3R) model* for datasets without citation information with central publication class follow adapted versions of equations (4.2), (4.3), (4.5), (4.8) and (4.9). Because we do not have any information about the citations and for the simplicity, we omitted equations (4.5), (4.8) and set $\alpha_1 = 0, \alpha_2 = \alpha_3$ and $\beta_p = 1 - 2\alpha_2$. The equation (4.9) are rewritten as follows :$\forall p \in \mathcal{P}, c' = C(p)$

$$R_p^{SD3R}(p) = \alpha_2 \sum_{a' \in A(p)} \frac{R_a(a')}{|P(a')|} + \alpha_2 \frac{R_c(c')}{|P_c(c')|} + (1 - 2\alpha_2) I_p, \tag{5.1}$$

5.2.2. Models with citations

We evaluate and compare four citation-based methods when the bibliometric system have full 4-tiers: author, publication, conference and citation.

NCC Model. With citation information between publications, we employ a *naive model based on citation counting (NCC)* for ranking author and conference. In NCC model, for each author $a \in \mathcal{A}$ $R_a^{NCC}(a) ::= |C(a)|$ and for each conference $c \in \mathcal{C}$ $R_c^{NCC}(c) ::= |C_c(c)|$.

H-index Model.* We adopt a well known *H-index* **??** model to rank authors and conferences for citation dataset D_c. In H-index* model, for each author $a \in \mathcal{A}$ $R_a^{H-Index*}(a) ::= |P(a)?????????|$ and for each conference $c \in \mathcal{C}$ $R_c^{H-Index*}(c) ::= |P_c(c)???????????|anhVugiupemnhe.$

IF Model.* For ranking authors and conferences in citation dataset D_c, we adopt *Impact Factor* model* **??** and calculate the average number of citations over number of publications. In *IF* model*, for each author $a \in \mathcal{A}$ $R_a^{IF*}(a) ::= |C(a)|/|P(a)|$ and for each conference $c \in \mathcal{C}$ $R_c^{IF*}(c) ::= |C_c(c)|/|P_c(c)|$.

SC4R Model. We implemented a *Simple Citation 4-star Ranking (SC4R) model* with citation information follow adapted versions of equations (4.2), (4.3), (4.5), (4.8) and (4.9). For the simplicity, we set $\alpha_1 = \alpha_2 = \alpha_3$ and $\beta_p = 1 - 3\alpha_1$. The equation (4.9) are rewritten as follows: $\forall p \in \mathcal{P}, c' = C(p)$ (5.2)

$$R_p^{SC4R}(p) = \alpha_1 \sum_{l' \in IN(p)} R_l(l') + \alpha_1 \sum_{a' \in A(p)} \frac{R_a(a')}{|P(a')|} + \alpha_1 \frac{R_c(c')}{|P_c(c')|} + (1 - 3\alpha_1) I_p,$$

We also would like to compare the results of three citation-based methods with that of *NPC* and *SD3R* methods mentioned above over dataset D_c.

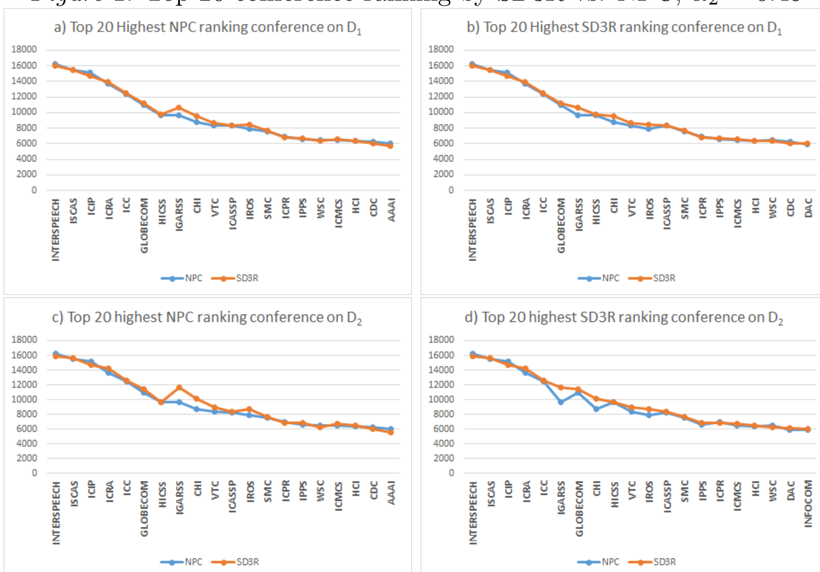
5.2.3. Ranking Evaluation

To compare and evaluate the rank scores of authors and conferences between each ranking models, we scale each rank score values in all models such that their sum all are the same, equal to a constant. We have:
$\sum_{a\in\mathcal{A}} R_a^{NPC}(a) = \sum_{a\in\mathcal{A}} R_a^{SD3R}(a) = \sum_{a\in\mathcal{A}} R_a^{NCC}(a) = \sum_{a\in\mathcal{A}} R_a^{H-index*}(a) = \sum_{a\in\mathcal{A}} R_a^{IF*}(a) = \sum_{a\in\mathcal{A}} R_a^{SC4R}(a)$, and
$\sum_{c\in\mathcal{C}} R_c^{NPC}(c) = \sum_{c\in\mathcal{C}} R_c^{SD3R}(c) = \sum_{c\in\mathcal{C}} R_c^{NCC}(c) = \sum_{c\in\mathcal{C}} R_c^{H-index*}(c) = \sum_{c\in\mathcal{C}} R_c^{IF*}(c) = \sum_{c\in\mathcal{C}} R_c^{SC4R}(c)$.

For each author $a \in \mathcal{A}$ and conference $c \in \mathcal{C}$, we measure the difference between ranking scores of two models i, j :
$\Delta^{i,j}(a) = R_a^i(a) - R_a^j(a)$, $\%\Delta^{i,j}(a) = \frac{\Delta^{ij}(a)}{R_a^j(a)}$, and
$\Delta^{i,j}(c) = R_c^i(c) - R_c^j(c)$, $\%\Delta^{i,j}(c) = \frac{\Delta^{i,j}(c)}{R_c^j(c)}$.

We also do examine the Spearman's rank correlation coefficient for each pair of rank scores between methods. We denote $\rho_a^{i,j}$ and $\rho_c^{i,j}$ as the author and conference rank correlation coefficient between method i and j. By examining the $\Delta^{i,j}$, $\%\Delta^{i,j}$ and $\rho^{i,j}$ functions, we found following interesting results.

Figure 1. Top 20 conference ranking by SD3R vs. NPC, $\alpha_2 = 0.45$



5.3. Experimental results and discussions

5.3.1. Without citation information

Remark 5.1. *SD3R and NPC model rank conferences similarly, but SD3R can reflect how hot the conferences are.*

According to Fig. 1 , all most of top 20 conference ranking values are the same for both methods, in both dataset D_1, D_2. We can see a slightly different values of top 20 conferences between D_1, big dataset with all conferences extracted from DBLP versus D_2, only conferences having more than 300 publications. Let's examine the three most different ranking conferences in these top 20 lists, IGARSS, CHI and IROS. They all increase their ranking values when cumputing by SD3R vs. NPC, for both D_1 and D_2. And the bigger of increasing values in D_2 than D_1 can be explained by the absence of small-published conferences in D_2.

Moreover, all of these three conferences stands in the top 5 increasing ranking conferences list when measuring by SD3R vs. NPC over D_1 and D_2, seeing the Fig. 2-a. Interestingly, all *increasing* conferences are young, annual events and get hot topics, such as remote sensing (IGARSS), computer human interaction (CHI), medical image computing (MICCAI), solid-state circuits (ISSCC), intelligent robot (IROS)...

Figure 2. Top 5 most different ranking scores of conferences by SD3R vs. NPC over D_1 and D_2, $\alpha_2 = 0.45$

a) most increasing value

CONF	FULL NAME	YEAR	NPC	SD3R(D_1)	$\Delta^{SD3R(D1),NPC}$	SD3R(D_2)	$\Delta^{SD3R(D2),NPC}$
IGARSS	IEEE International Geoscience and Remote Sensing Symposium	2005	9689	10640.91	951.91	11636.60	1947.60
CHI	Computer Human Interaction	1990	8737	9571.49	834.49	10074.77	1337.77
MICCAI	Medical Image Computing and Computer - Assisted Intervention	1998	3778	4552.62	774.62	5120.85	1342.85
ISSCC	International Solid-State Circuits Conference	2009	1185	1794.49	609.49	2233.59	1048.59
IROS	International Conference on Intelligent RObots and Systems	1992	7904	8410.25	506.25	8747.37	843.37

b) most decreasing value

CONF	FULL NAME	YEAR	NPC	SD3R(D_1)	$\Delta^{SD3R(D1),NPC}$	SD3R(D_2)	$\Delta^{SD3R(D2),NPC}$
IJCAI	International Joint Conference on Artificial Intelligence (biennial)	1969	5635	5158.60	-476.40	4861.19	-773.81
IFIP	International Federation for Information Processing (biennial)	1959	2796	2383.08	-412.92	2100.05	-695.95
GI	GI-Jahrestagung (language spoken is Germany)	1972	4349	4021.90	-327.10	3794.84	-554.16
AAAI	Conference on Artificial Intelligence (USA)	1980	5990	5689.97	-300.03	5473.98	-516.02
MFCS	Mathematical Foundations of Computer Science (Czechoslovakia area)	1972	2115	1833.59	-281.41	1621.16	-493.84

Figure 3. Measure the different ranking scores values by SD3R vs. NPC, $\alpha_2 = 0.45$

DATASET	avg($\%\Delta_a^{SD3R,NPC}$)	avg($\%\Delta_c^{SD3R,NPC}$)	$\rho_a^{SD3R,NPC}$	$\rho_c^{SD3R,NPC}$
D_1	34.4	5.6	0.97	0.99
D_2	29.6	7.6	0.98	0.99

On the other hand, according to Fig. 2-b, the top five *decreasing* conferences are held for over a long time or biennial events, in local community (GI, MFCS) or long exploited topics such as artificial intelligence (IJCAI, AAAI), image processing(IFIP)...
We also see all the decrease ranking values by SD3R vs.NPC on D_2 are more bigger than that values of D_1. This can be explained by the publishing activity of authors in small-published conferences has been remove in D_2. We remind that D_2 contains only conferences having not less than 300 publications whereas D_1 contains all venues from DBLP.

Fig. 3 shows the average of $\%\Delta_c^{SD3R,NPC}$ of conferences is small less than 10% for both D_1 and D_2. It also emphasizes that SD3R and NPC methods give very similar results when ranking conferences. The Spearman correlation coefficient of ranking conferences $\rho^{SD3R,NPC}$ equals to 0.99 indicates that the conference ranking scores of both methods are perfect monotone.

Remark 5.2. *SD3R ranks authors differently from NPC and can reflect the contribution of the author betters.*

According to Fig. 3, the average value of $\%\Delta_a^{SD3R,NPC}$ of authors both dataset D_1 and D_2 are 31.4% and 29.6% respectively. We conclude that *SD3R* and *NPC* methods result in significantly different scores when ranking authors. The Spearman correlation coefficient $\rho^{SD3R,NPC}$ is near by 1.0 implies that the rank results of *SD3R* and *NPC* are perfect monotone.
We can also find SD3R quite different from NPC method in Fig. 4, showing the top 20 most different ranking scores of authors. We realize that most authors in the top *decreasing* ranking values have a large number of publications in the

Figure 4. Top 20 most different ranking scores of authors by SD3R vs. NPC, $\alpha_2 = 0.45$

first glance, and vice versa. Let's see more detail in Fig. 5, about five authors in Fig. 4-b, Fig. 4-d whose publications are the biggest. Their publications have a large number of co-authors and belong to many conferences. We can also see the decrease ranking values by SD3R vs.NPC on D_1, a bigger dataset, are slightly bigger than that values of D_2, dataset with only conferences having at least 300 publications.

Looking for details in Fig. 4-a, Fig. 4-c, we observe that most of top *increasing* authors have a small number of publications in datasets. Most of their papers have no co-authors and are published in some specialized conferences. Additionally, in Fig. 6, we find out they are really key-person in their research topics

Figure 5. Top 5 most decreasing ranking scores of authors having biggest NPC

NAME	NPC (D_1)	SD3R(D_1)	$\Delta_a^{SD3R,NPC}(D_1)$	nConfs.(D_1)	NPC (D_2)	SD3R(D_2)	$\Delta_a^{SD3R,NPC}(D_2)$	nConfs.(D_2)
Wen Gao	528	398.23	-129.77	95	488	373.12	-114.88	81
Wei Wang	510	441.53	-68.47	236	448	394.00	-54.00	186
Wei Liu	467	393.73	-73.27	219	426	362.59	-63.41	183
Mahmut T. Kandemir	427	362.44	-64.56	78	405	350.31	-54.69	66
Mario Piattini Velthius	417	346.01	-70.99	100	315	259.88	-55.12	65

Figure 6. Top 5 most increasing ranking scores of authors having smallest NPC

NAME	NPC (D_1)	SD3R(D_1)	$\triangle_a^{SD3R,NPC}(D_1)$	nConfs.(D_1)	NPC (D_2)	SD3R(D_2)	$\triangle_a^{SD3R,NPC}(D_2)$	nConfs.(D_2)
Lotfi A. Zadeh	53	130.50	77.50	29	45	103.998	58.99811962	21
Ryotaro Kamimura	62	151.46	89.46	14	55	134.444	79.44385132	11
Ellen M. Voorhees	62	133.97	71.97	11	61	139.214	78.21359319	10
Toshihiko Yamakami	70	195.17	125.17	29	55	159.163	104.1627818	21
Keqin Li	74	151.74	77.74	14	70	141.899	71.89933631	12

in real life, such as Professor Emeritus Lotfi A. Zadeh ‡ from the University of California, Berkeley, who set out the theory of fuzzy sets; Professor Ryotaro Kamimura at Tokai University , specialized on Machine Learning and Pattern Recognition; Computer Scientist Ellen M. Voorhees, at NIST, famous for international workshops series: the Text REtrieval Conference (TREC), TREC Video (TRECVid), and the Text Analysis Conference (TAC); Senior Specialist Toshihiko Yamakami , at ACCESS, Japan Advanced Institute of Science and Technology, professional on Mobile Social Application; Professor Keqin Li § from State University of New York at New Paltz, notable for parallel and distributed computing...
There are some special cases, authors have big NPC ranking scores but still increase their scores when ranking by SD3R. For instances, Edwin Hancock is a very well known scientist on computer vision ¶, Norman C. Beaulieu, an Canadian engineer and professor in the ECE department of the University of Alberta is very famous in broadband digital and communications systems, Professor Irith Pomeranz, affiliated at School of Electrical and Computer Engineering, Purdue University ‖ is noble for Computer Engineering VLSI and Circuit Design...

5.3.2. With citation information

Remark 5.3. *SC4R is more fine-grained than NPC and NCC when ranking of conferences.*

Because of the naive counting feature, both NPC and NCC methods have to suffer from the "too many equal rank problem" for conferences having the same publications or citations. The problem of "equal rank" still happens in SC4R ranking method but not frequently because of the real calculation. The different ranking measurement figures are shown in Fig. 7. The Kendall' tau coefficients points out that ranking by SC4R is more concordant with NCC than NPC method, and each pair of them is not quite the same concordant. The Spearman $\rho^S C4R, NCC$, near by one (0.91), indicates that SC4R and

‡http://en.wikipedia.org/wiki/Lotfi_A._Zadeh
§http://www.cs.newpaltz.edu/∼lik/
¶http://scholar.google.com/citations?user=EjDU2ncAAAAJ
‖https://engineering.purdue.edu/∼pomeranz/

Figure 7. Different ranking measurement of SC4R with $\alpha_1 = 0.3$ vs. NPC, NCC, H-index on D_c

a) Different ranking values of conferences

R_1,R_2	ρ	κ	Avg_Δ
SC4R,NCC	0.910	0.833	1.087
SC4R,NPC	0.877	0.825	0.787
NCC,NPC	0.688	0.676	4.901

b) Different ranking values of authors

R_1,R_2	ρ	κ	Avg_Δ
SD4R,NCC	0.610		0.987
SD4R,NPC	0.715		1.054
SD4R,H-index	0.602		1.631

NCC are perfect monotonic functions. The Avg_Δ values is nearly zero. It states that ranking values of SC4R in total does not quite different from others methods.

Remark 5.4. *SC4R seems to be the mean of the popularity and the prestige ranking of conferences.*

Fig. 8 shows the trend of SC4R lines in top ranking value from three methods. In each graph, the SC4R line always fluctuate between NPC (popularity) line and NCC (prestige) line, nearly coincides to the average of them. The first impression point is the pair of VLDB and HICSS. The number of articles published in HICSS is fivefold of VLDB's but HICSS′s cited papers are just a fourth of VLDB. In our SC4R method, VLDB stands before HICSS. Their ranking values are 8403.14 and 7037.91 respectively, totally not far different.

Let's look at Fig. 9 which shows top 10 rank position by SC4R methods. We can easily point out two prestige database conferences, PODS and ISWC, which has low rank positions in popularity, being 18 and 16 respectively. Further interesting thing is that our ranking list of top ten conferences are very noble in database field following the assessment of experts.

Remark 5.5. *SC4R reflects the contribution of the author betters than others.*

Fig 10-a shows the top 20 highest NPC ranking author value. In this popularity ranking graph, we can see H-index line is the lowest line and separated with other lines. It can be explain by many authors have published a big number of papers but their papers get a few citations. We remind that H-index is proposed for the combination of popularity and prestige value of an author. In other three graphs, showing ranking values by NCC (quality method), H-index (a method combining quantity and quality) and our SC4R method, we can see H-index nearly conincides with NPC line. We suppose that H-index extremely falls into quantity pole in these case.

Meanwhile, the NCC line, standing for the prestige of author via high citation, is always the highest line in Fig. 10-b, Fig. 10-c, Fig. 10-d. It can be concluded that our SC4R is the best method for reflecting the contribution

Figure 8. Top highest ranking value of conferences by NPC, NCC and SC4R on D_c with $\alpha_1 = 0.3$

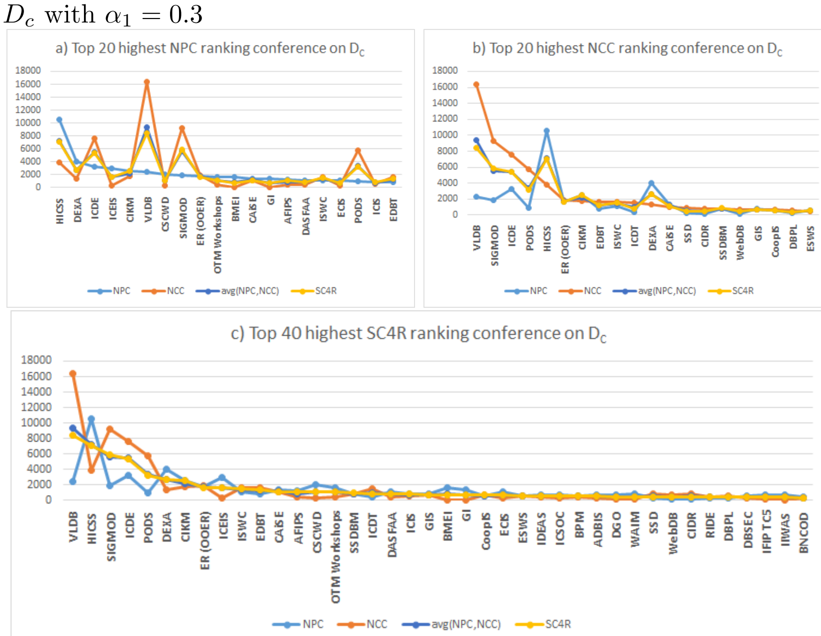
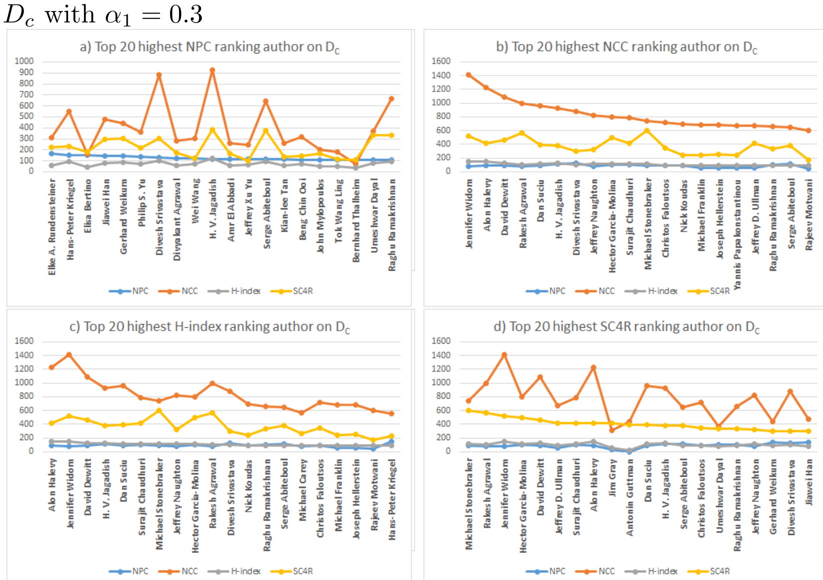


Figure 9. Top 10 SC4R rank position of conferences on D_c with $\alpha_1 = 0.3$

CONF	NAME	SC4R Rank	NPC Rank	NCC Rank	SC4R Value	NPC Value	NCC Value
VLDB	Very Large Data Bases	1	6	1	8403.14	2348	33043
HICSS	Hawaii International Conference on System Sciences	2	1	5	7037.91	10547	7705
SIGMOD	International Conference on Management of Data	3	8	2	5836.41	1881	18707
ICDE	International Conference on Data Engineering	4	3	3	5392.72	3227	15372
PODS	Symposium on Principles of Database Systems	5	18	4	3155.55	941	11605
DEXA	Database and Expert Systems Applications	6	2	11	2650.22	3982	2763
CIKM	International Conference on Information and Knowledge Management	7	5	7	2499.09	2526	3590
ER (OOER)	International Conference on Conceptual Modeling	8	9	6	1640.43	1697	3770
ICEIS	International Conference on Enterprise Information Systems	9	4	32	1554.69	2993	612
ISWC	International Semantic Web Conference	10	16	9	1504.96	1137	3332

Figure 10. Top 20 ranking value of author by NPC, NCC, H-index* and SC4R on D_c with $\alpha_1 = 0.3$



of author. It does not fall extremely into quantity (NPC, H-index) or quality (NCC) poles.

Let's look detail for some special occurences in Fig. 11. The citation number of Professor Jennifer Widomis nearly double than that of Michael Stonebraker, a pioneer of data base research and technology . She also has H-index higher than Michael Stonebraker ** but her number of publications is lower than his one. We can see her SD4R value is a bit lower than Michael Stonebraker's SD4R value. We assume the result is because many publications citing to her papers do not get high score in SC4R ranking, namely not good quality. Alon Halevy is another similar example for high citation, high H-index, high publications but lower SC4R.

We want to note three other opposited cases in which all authors have low NPC, NCC and H-index value, but high SD3R score. They are Jim Gray, Antonin Guttman and Umeshwar Dayal. This situation can be explained by their publications are cited by almost prestige people and high quality publications and are published in noble conferences.

We review the top 20 highest SC4R value people and find out that they all are very famous in database field with many valuable scientific works.

**https://www.csail.mit.edu/user/1547

Figure 11. Top 20 SC4R ranking value on D_c with $\alpha_1 = 0.3$

Author	NPC	NCC	H-index	SC4R	NPC Rank	NCC rank	H-index Rank	SC4R Rank
Michael Stonebraker	96	1653	24	603	27	11	7	1
Rakesh Agrawal	79	2237	22	564	50	4	10	2
Jennifer Widom	86	3150	30	522	40	1	2	3
Hector Garcia-Molina	100	1792	23	501	23	9	8	4
David Dewitt	87	2446	26	459	37	3	4	5
Jeffrey D. Ullman	59	1496	18	419	124	17	25	6
Surajit Chaudhuri	103	1771	25	416	21	10	5	7
Alon Halevy	91	2742	32	415	33	2	1	8
Jim Gray	30	695	11	414	503	62	111	9
Antonin Guttman	5	986	4	398	6568	32	1455	10
Dan Suciu	96	2147	25	394	26	5	6	11
H. V. Jagadish	118	2072	26	380	10	6	3	12
Serge Abiteboul	112	1440	20	380	13	19	12	13
Christos Faloutsos	90	1610	19	345	34	12	17	14
Umeshwar Dayal	107	832	17	332	19	41	28	15
Raghu Ramakrishnan	105	1487	20	332	20	18	13	16
Jeffrey Naughton	84	1834	23	320	44	8	9	17
Gerhard Weikum	143	995	18	303	5	31	23	18
Divesh Srivastava	128	1969	21	300	7	7	11	19
Jiawei Han	143	1061	16	299	4	30	32	20

6. Related work

Our prior work. Our work is the extend version of the prior work [18]. In this version, we has extend the experiments and represent everything more in detail. Concretely, we do the experiments for the case the citations are considered. We also compare our ranking with the H-index, which is the most famous ranking scores for authors recently. Moreover, we improve the results by giving more discussions.

The ranking problem occurs and develop quickly with the era of the Internet and big data. One of the most famous ranking problem is ranking web-pages. A brief overview of this problem can be found at Dilip Kumar Sharma et al. [16]. PageRank were utilized by Google search engine [14]. Since the hyperlink structure among the web-pages are easily represented as a web graph, the PageRank of each web-page can be measured (see Sect. 2.2 for more detail). Hyperlink-Induced Topic Search (HITS) is a link analysis algorithm that rates Web pages, developed by Jon Kleinberg [6]. It was a precursor to PageRank. The idea behind HITS algorithm classify the webs into two classes: (i) hubs, served as large directories point to (ii) authoritative pages. A good hub represented a page that pointed to many other pages, and a good authority represented a page that was linked by many different hubs. The model can be rewritten into 2-linear ranking model.

Named entities: In Natural Language Processing (NLP) communities, named entity recognition is an important problem. Ranking scheme has been applied to solve the problem. Collins [10] proposes a ranking method based on a maximum-entropy tagger. Also, Vercoustre et al. [1] has presented how to apply the ranking method to Wikipedia.

Scientific articles In bibliometrics, the scientific articles (e.g., research papers, technical reports, and so on) are evaluated with respect to the quality (e.g., novelty and originality) as well as academic influence to the communities (e.g., impact) by relaying on the citations (e.g., references and quotation) [3].

Researchers Also, Researchers has been ranked by citation analysis (e.g., how many papers has he/she published, how many times have his/her papers cited, and so on). More interestingly, H-index (Hirsch index) has been designed to measure both the productivity and impact of the published work of the researchers.

Complex system The complex system ranking has been already explored using a different formalism for ranking or classification in heterogeneous networks [19, 5]. The poprank model [19] introduce the Popularity Propagation Factor to express the relationship between classes. Their model is based on the markov chain model which can be applied in the N-linear mutual ranking systems. The quantium ranking [5] is based on quantium navigation. Their formula is come from the quantium theory and quite different to ours.

So far, conferences have been ranked by subjective opinions and consensus among well known experts in a domain. Such lists in computer science area are compiled here[††]. In this work, we have proposed a novel conference ranking framework to integrate all possible evidence.

7. Conclusion and future works

We have introduced and studied N-star ranking for mutual ranking systems. The mutual relationships between ranking objects are described by a system of linear equations, N-mutual ranking system. A N-linear mutual ranking system is a N-star ranking systems if it has a core class which affects and reflects all other classes in the system. The rank scores of the N-star ranking system are unique and computed by a Markov chain. We have pointed out that PageRank

[††]http://intelligent.pe.kr/ConfRank.txt

is a 2-star ranking. It has two classes: the web-pages (a core class) and links.

We have introduced and studied a general and a simple 4-star ranking models for ranking authors, publications, conferences. A general model is a generic one. In a simple model, we consider each publication, author, conference, citation is equally. We have conducted the experiments for the models in which the citations are not considered. The experimental results are based on the DBLP dataset. By comparing the difference between the SD4R vs. NPC models, we have shown that our ranking system can reflect how hot the conference are and record the contribution of the authors better than the naive ranking system. Moreover, our ranking system makes a big change on author ranking.

As future work, we are planning to *i*) get the citations between the publication to upgrade the quality of our ranking system, *ii*) study how to combine a N-star ranking systems with a given ranking systems, *iii*) investigate the time series in N-star ranking and the trend prediction problem, and *iv*) apply N-star ranking systems in various ranking problems, e.g., business ranking, event ranking, and so on.

References

[1] **Anne-Marie Vercoustre, James A. Thom, Jovan Pehcevski.** Entity ranking in Wikipedia. *Proceedings of the 2008 ACM symposium on Applied computing (SAC'08)*, (2008) 1101-1106.

[2] **A. Sidiropoulos, D. Katsaros, Y. Manolopoulos.** Generalized Hirsch h-index for disclosing latent facts in citation networks. *Scientometrics,* **Vol. 72 (2)**, (2006) 253–280.

[3] **Blaise Cronin.** Bibliometrics and beyond: some thoughts on web-based citation analysis. *Journal of Information Science*, **Vol. 27 (1)**, (2001), 1–7 .

[4] **C. Freudenthaler, S. Rendle and L. S. Thieme.** Factorizing Markov Models for Categorical Time Series Prediction. *In Proceedings of IC-NAAM,* (2011) 405-409.

[5] **Eduardo Snchez-Burillo, Jordi Duch, Jess Gmez-Gardenes, David Zueco.** Quantum Navigation and Ranking in Complex Networks *Nature online journal* (2012).

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

[7] **J. Keener.** The Perron-Frobenius theorem and the ranking of football teams. *SIAM Review*, **Vol. 35(1)**, (1993), 80–93.

[8] **Jon Kleinberg** Authoritative Sources in a Hyperlinked Environment. *Journal of the ACM*, **Vol. 46 (5)**, (1999), 604-632.

[9] **L. T. Kien, L. T. Hieu, T. L. Hung, L. A. Vu.** MpageRank: The Stability of Web Graph. *Vietnam Journal of Mathematics*, **Vol. 37**, (2009), 475–489

[10] **Michael Collins.** Ranking algorithms for named-entity extraction: boosting and the voted perceptron. *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics,* (2002), 489–496.

[11] **Michael Ley.** DBLP - Some Lessons Learned. *PVLDB* 2 (2), (2009) 1493–1500.

[12] **Michael Ley, Patrick Reuther.** Maintaining an Online Bibliographical Database: The Problem of Data Quality. *EGC 2006*, (2006) 5–10.

[13] **Microsoft Corporation.** Microsoft Academic Search. *http://academic.research.microsoft.com/* (June -26 -2013).

[14] **S. Brin and L. Page.** The anatomy of a large-scale hypertextual web search engine. *In Proceedings of the 7th International World Wide Web Conference*, (1998), 107–117

[15] **S. Rendle, C. Freudenthaler, L. S. Thieme.** Factorizing personalized Markov chains for next-basket recommendation. *Proceedings of the 19th international conference on World wide web*, (2010) 811–820.

[16] **Sharma, Dilip Kumar and Sharma A. K** A comparative analysis of web page ranking algorithms. *International Journal on Computer Science and Engineering* . (2010) 2670–2676.

[17] **T. Furukawa, S. Okamoto, Y. Matsuo, M. Ishizuka.** Prediction of social bookmarking based on a behavior transition model. *Proceedings of the 2010 ACM Symposium on Applied Computing*, (2010) 1741–1747.

[18] **Vu, L. A., Hai, V. H., Hieu, L. T., Kien, L. T. and Jason, J. J.** A General Model for Mutual Ranking Systems, *Intelligent Information and Database Systems Lecture Notes in Computer Science*, **8397** (2014), 211–200.

[19] **Z. Nie and Y. Zhang and J. Wen and W. Ma.** Object-Level Ranking: Bringing Order to Web Objects, Study of the eXplicit Control Protocol (XCP). *IEEE Infocom* (2005).

Vu Le Anh
Nguyen Tat Thanh University

Ho Chi Minh City
Vietnam
e-mail lavu@ntt.edu.vn

Vo Hoang Hai
Information Technology College
Ho Chi Minh city
Vietnam
e-mail vohoanghai2@gmail.com

Hieu Le Trung
Duy Tan University
Da Nang
Vietnam
e-mail hieukien82@gmail.com

Kien Le Trung
University of Science
Thua Thien Hue
Vietnam
e-mail hieukien@hotmail.com

Jason J. Jung
Yeungnam University
Gyeongsan
Korea
e-mail