

## A General Model for Mutual Ranking Systems

**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 mutual ranking, which ranks the webpages 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 the models in which the citations are not considered. The experimental results are based on the DBLP dataset, which contains more than one million papers, authors and thousands of conferences and journals in computer science. Finally, N-star ranking is a very strong ranking algorithm can be applied in many real-world problems.

---

*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.

However, there are some problems on the ranking schemes of traditional systems (i.e., web searching and impact factor). These systems are based on simple links among only the target entities (e.g., web pages and papers). We assume that this type of referencing process is not enough for measuring the evaluation indicators more precisely.

In this work, we formulate a mathematical model of N-star ranking scheme, which can represent a generalized referencing process. The N-star ranking is based on N-linear mutual ranking system in which  $N$  classes are ranked at once. Especially, as a real-world problem, conference ranking has been regarded as an interesting issue on academic communities. The conference publication (e.g., proceedings) is related with various factors such as authors (and their affiliation), papers (and their citations), and conferences.

The remainder of this paper is as follows. In Sect. ??, we describe the backgrounds of the N-linear mutual ranking (e.g., the rank scoring and conceptual modeling on relationships between classes). Sect. ?? defines the N-star ranking as a N-linear mutual ranking system in which there exists a core class. We study two N-star ranking models for the author, publication and conference ranking problem in different contexts in Sect. ?. Also, Sect. ?? shows the experiments for the simple N-star ranking model of authors, publications and conference ranking. In Sect. ??, we exhibit the related work for comparison with the proposed N-star ranking model. Finally, Sect. ?? draws a conclusion of this work.

## 2. Backgrounds

### 2.1. N-linear mutual ranking system

The couple  $(\mathcal{A}, R)$  is called a *ranking system* if (i)  $\mathcal{A} = \{a_1, \dots, a_n\}$  is a finite set, and (ii)  $R : \mathcal{A} \rightarrow [0; +\infty)$ .  $\mathcal{A}$  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  $R(a) > 0 \quad \forall a \in \mathcal{A}$ .  $n = |\mathcal{A}|$  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  $\alpha_{ij}, \beta_i \in [0; +\infty)$ ,  $I_i = (t_u^{(i)})_{n_i}$  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

$$\sum_j \alpha_{ij} + \beta_i = 1 \quad \text{and} \quad R_i = \sum_{j=1}^N \alpha_{ij} W_{ij} R_j + \beta_i I_i, \quad i = 1, \dots, N.$$

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

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  $a_{jv} \rightarrow a_{iu}$ ) 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  $\mathcal{A}_i \rightarrow \mathcal{A}_j$ ) if  $\forall a_{jv} \in \mathcal{A}_j: \exists a_{iu_1}, a_{iu_2} \in \mathcal{A}_i: a_{jv} \rightarrow a_{iu_2} \wedge a_{iu_1} \rightarrow a_{jv}$ .

**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  $\mathcal{A}_i \rightarrow \mathcal{A}_j$  or  $\exists \mathcal{A}_k : \mathcal{A}_i \rightarrow \mathcal{A}_k \wedge \mathcal{A}_k \rightsquigarrow \mathcal{A}_j$ .

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

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

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  $l_k \in \mathcal{L} \wedge w_t = in(l_k) \in \mathcal{W}$ , 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  $l_k \in \mathcal{L} \wedge w_t = out(l_k) \in \mathcal{W}$ , 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:

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

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

Note that, since for each link  $l \in \mathcal{L}$ , let  $u = in(l)$  and  $v = out(l)$  then web  $u$  affects to link  $l$ , ( $u \rightarrow l$ ) and link  $l$  affects to web  $v$ , ( $l \rightarrow v$ ). 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.  $\Omega$  is called a  $N$ -star ranking if

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

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

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  $\Omega$ .

**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  $R_1, \dots, R_N$  satisfy the following equations:

$$(3.1) \quad R_1 = W R_1 \quad \text{and} \quad R_i = W_{i1} R_1 \quad \forall i = 2, \dots, N,$$

where

$$(3.2) \quad W = \alpha_{11} W_{11} + \alpha_{12} W_{12} W_{21} + \dots + \alpha_{1N} W_{1N} W_{N1} + \beta_1 \mathbf{I}_1$$

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 [?, ?]) confirms that there exists a unique score  $R_1$  with  $\sum_{a \in \mathcal{A}_1} R_1(a) = 1$  such that

$$R_1 = W R_1.$$

From (??), the unique existence of  $R_1$  infers the unique existences of  $R_2, \dots, R_N$ . Moreover, since  $W_{21}, \dots, W_{N1}$  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  $i = 2, \dots, N$ . 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 :**  $\{R_i\}_{i=1}^N$

1. **begin**
  2.     Check the  $N$ -star ranking model with the core  $\mathcal{A}_1$
  3.     Let 
$$W \leftarrow \alpha_{11}W_{11} + \sum_{i=2}^N \alpha_{1i}W_{1i}W_{i1} + \beta_1\mathbf{I}_1$$
  4.     Initialize  $R_1^{(0)}$  : uniform distribution,  $k = 0$
  5.     **repeat**
  6.          $k = k + 1$
  7.         Update  $R_1^{(k)} \leftarrow WR_1^{(k-1)}$
  8.     **until**  $\|R_1^{(k)} - R_1^{(k-1)}\| \leq$  a stopping criterion
  9.     Let  $R_1 = R_1^{(k)}$  and  $R_i = W_{i1}R_1, i = 2, \dots, N$
  10. **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 pub-

lications affects to all of its authors:

$\forall a \in \mathcal{A}, p \in \mathcal{P} :$

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

If a publication affects equally to its authors (a), (??) is rewritten as follows:

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

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

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

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

3. The score of a citation depends on the citing publication, and each publications affects to all of its citations:

$\forall l \in \mathcal{L}, p \in \mathcal{P}, p' = in(l) :$

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

If a publication affects equally to its citations (b), (??) is rewritten as follows:

$\forall l \in \mathcal{L}, p \in \mathcal{P}, p' = in(l), l' \in OUT(p) :$

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

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 :

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

(4.6)

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

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

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 (??) and (??) are rewritten as follows:

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

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

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

The model which is described by equations (??), (??), (??), (??) and (??) is called *the general 4-star ranking model for the ranking authors, publications and conferences problem*. And the model which is described by equations (??), (??), (??), (??) and (??) 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

Code Environment Limit computing resources D1, D2, D3 why we have?

### 5.1. Dataset

*Datasets.* We exploited the DBLP data set\*, which provides bibliographic information on major computer science journals and proceedings, to conduct experiment for the simple 4-star ranking model for the ranking authors, publications and conferences problem on fields of computer science. A publication can be both in a conference and another journal, too. To avoid this gap, we keep only the publications related to conferences and ignore the journals's ones. We has built program to parse DBLP dataset in XML format to extract the authors, title, and publication venue information from the guides [?, ?]. We mention our readers that DBLP has no information about the citations between

---

\*<http://dblp.uni-trier.de> accessed on May 2013



Table 1. Experiment datasets.

Datasets	nPubs	nAuthors.	nConfs.	nCites
$D_1$	1253997	1048576	3351	-
$D_2$	1046030	746560	949	-
$D_3$	77361	87354	644	156429

publications, so we amend it with citations collected from Academic Microsoft<sup>†</sup>. D1, D2, D3 why we have? We generate three datasets for the experiments:  $D_1$  and  $D_2$  with no citation and  $D_3$  with citation information.  $D_1$  is no-citation dataset for all conferences collected in DBLP.  $D_2$  is no-citation dataset with conferences having at least 300 paper.  $D_3$  is citation dataset with conferences belonging to database field.

## 5.2. Without citation models

*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  $c \in \mathcal{C}$   $R_c^{NPC}(c) ::= |P_c(c)|$ .

*SD4R Model.* We implemented a *Simple DBLP 4-star Ranking (SD4R) model* for datasets without citation information with central publication class follow adapted versions of equations (??), (??), (??), (??) and (??). Because we do not have any information about the citations and for the simplicity, we omitted equations (??), (??) and set  $\alpha_1 = 0, \alpha_2 = \alpha_3$  and  $\beta_p = 1 - 2\alpha_2$ . The equation (??) are rewritten as follows:  $\forall p \in \mathcal{P}, c' = C(p)$

$$(5.1) \quad R_p^{SD4R}(p) = \alpha_2 \sum_{a' \in A(p)} \frac{R_a(a')}{|P(a')|} + \alpha_2 \frac{R_c(c')}{|P_c(c')|} + (1 - 2\alpha_2)I_p,$$

## 5.3. With citation information

Besides using models for no-citation data, we calculate and compare with two more models:

*H-index.* For citation dataset  $D_3$ , we adopt H-index ?? model for ranking authors and conferences. In H-index model, for each author  $a \in \mathcal{A}$   $R_a^{H-Index}(a) ::= |P(a)|^{1/2}$  and for each conference  $c \in \mathcal{C}$   $R_c^{H-Index}(c) ::= |P_c(c)|^{1/2}$ .

<sup>†</sup><http://academic.research.microsoft.com> accessed on March 2014

*Citation counting.* For citation dataset  $D_3$ , we adopt "cite-count" model for ranking authors and conferences by naive counting reference to each author and each conference. In "cite-count" model, for each author  $a \in \mathcal{A}$   $R_a^{cite-count}(a) ::= |P(a)|$  and for each conference  $c \in \mathcal{C}$   $R_c^{cite-count}(c) ::= |P_c(c)|$ .

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

(5.2)

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

#### 5.4. Normalized for comparing results

To evaluating the rank scores of the three ranking models, we scale the rank scores of NPC, SD4R and SC4R model such that:

$$\begin{aligned} \sum_{a \in \mathcal{A}} R_a^{NPC}(a) &= \sum_{a \in \mathcal{A}} R_a^{SD4R}(a) = \sum_{a \in \mathcal{A}} R_a^{H-index}(a) = \\ \sum_{a \in \mathcal{A}} R_a^{cite-count}(a) &= \sum_{a \in \mathcal{A}} R_a^{SC4R}(a), \text{ and} \\ \sum_{c \in \mathcal{C}} R_c^{NPC}(c) &= \sum_{c \in \mathcal{C}} R_c^{SD4R}(c) = \sum_{c \in \mathcal{C}} R_c^{H-index}(c) = \\ \sum_{c \in \mathcal{C}} R_c^{cite-count}(c) &= \sum_{c \in \mathcal{C}} R_c^{SC4R}(c). \end{aligned}$$

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

We do examine the Spearman's rank correlation coefficient for each pair of three rank scores. We denote  $\rho_a^{i,j}$  and  $\rho_c^{i,j}$  as the author and conference rank correlation coefficient between method i and j.

#### 5.5. Experimental results

By examining the  $\rho^{i,j}$ ,  $\Delta^{i,j}$  and  $\% \Delta^{i,j}$  functions, we found following interesting results.

##### 5.5.1. Without citation information

- *SD4R ranking scores reflect how hot the conferences are.* Fig ?? has show the top 20 conferences most increasing and decreasing the ranking score over  $D_2$ . All *increasing* conferences are young, annual events and get hot topics now. For example, the top five topics are about remote sensing (IGARSS - IEEE International Geoscience and Remote Sensing

Figure 1. Top 20 conference most differences of ranking value of conferences using SD4R vs. NPC over  $D_2$ ,  $\alpha_2 = 0.45$

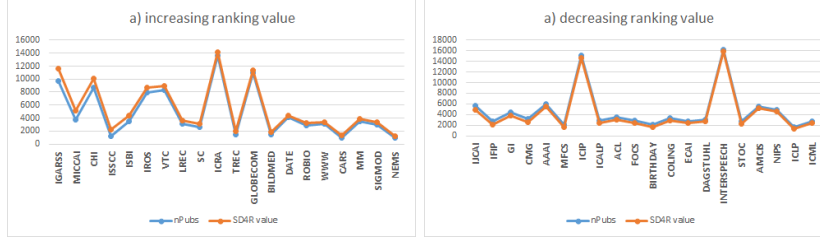


Figure 2. Top 20 authors most different ranking value of NPC vs. SD4R over  $D_2$ ,  $\alpha_2 = 0.45$

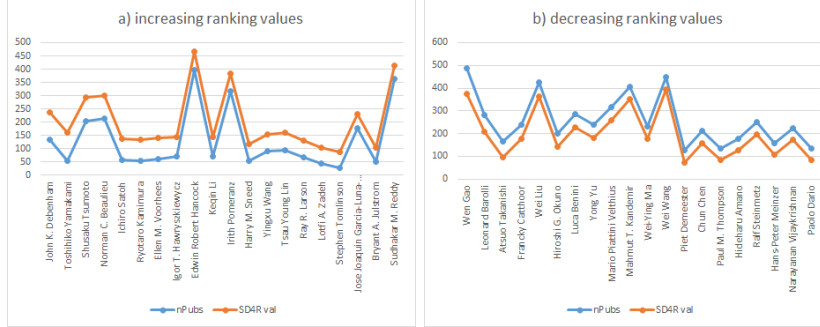


Table 2. The average  $\% \Delta^{NPC, SD4R}$  and the Spearman's rank correlation coefficient of author and conference using NPC, SD4R models on  $D_1, D_2$

Dataset	$avg(\% \Delta_a^{NPC, SD4R})$	$avg(\% \Delta_c^{NPC, SD4R})$	$\rho_a^{NPC, SD4R}$	$\rho_c^{NPC, SD4R}$
$D_1$	34.4%	5.6%	0.97	0.99
$D_2$	29.6%	7.6%	0.98	0.99

Symposium, from 2005), computer human interaction (CHI - Computer Human Interaction, from 1990), medical image computing (MICCAI - Medical Image Computing and Computer - Assisted Intervention, from 1998), (ISBI - IEEE International Symposium on Biomedical Imaging, from 2002), solid-state circuits (ISSCC - International Solid-State Circuits Conference, from 2009)... On the other hand, the top five *decreasing* conferences are held for over a long time or biennial events, in local community (GI-Jahrestagung - language spoken is Germany, from 1972) or with less interesting topics such as performance evaluation (CMG - Computer Measurement Group Conference, from 1976), artificial intelligence (IJCAI - International Joint Conference on Artificial Intelligence, biennial from 1969, AAAI - Conference on Artificial Intelligence, from 1980), image processing (IFIP - International Federation for Information Processing, biennial from 1959),...

- *SD4R ranks authors differently and is perfect monotonic with NPC.* According to Tab. ??, the average of  $\% \Delta_a^{NPC, SD4R}$  of conferences is small less than 10%, whereas that number of authors is much higher, around 30%. It implies that SD4R and NPC methods have similar results in ranking conferences but significantly different results in ranking authors. We can clearly see the similarity and difference of ranking results of conferences and authors in Fig. ?? and Fig. ?? respectively. The Spearman correlation coefficients in Tab. ?? indicate that the rank results of both methods are perfect monotone. Fig. ?? shows the significantly different in top 20 authors ranking for both methods with many interesting points. For instances, in  $D_1$  dataset, Wen Gao's number of publications is over seven hundreds papers, the third highest value in NPC. But he has just a bit over 500 points in ranking results of SD4R and does not appear in top 20 highest SD4R. It's also similar in  $D_2$  dataset, the absolute decreasing value of ranking results by NPC vs. SD4R of Wen Gao is around 120, equivalent to about 23.5% in relative measure. In contrast, we can see Robert L. Glass, Diane Crawford, Ronald R. Yager and Thomas D. Wilson, whose numbers of publication only between 250 and 310,
- *SD4R reflects the contribution of the author better than the NPC.* Fig. ?? has show the top 20 authors most increasing and decreasing the ranking score over  $D_2$ . We observe that all top *decreasing* authors are have a large number of publications. Their publications have a large number of co-authors. All top *increasing* authors are really key-person in their research topics.

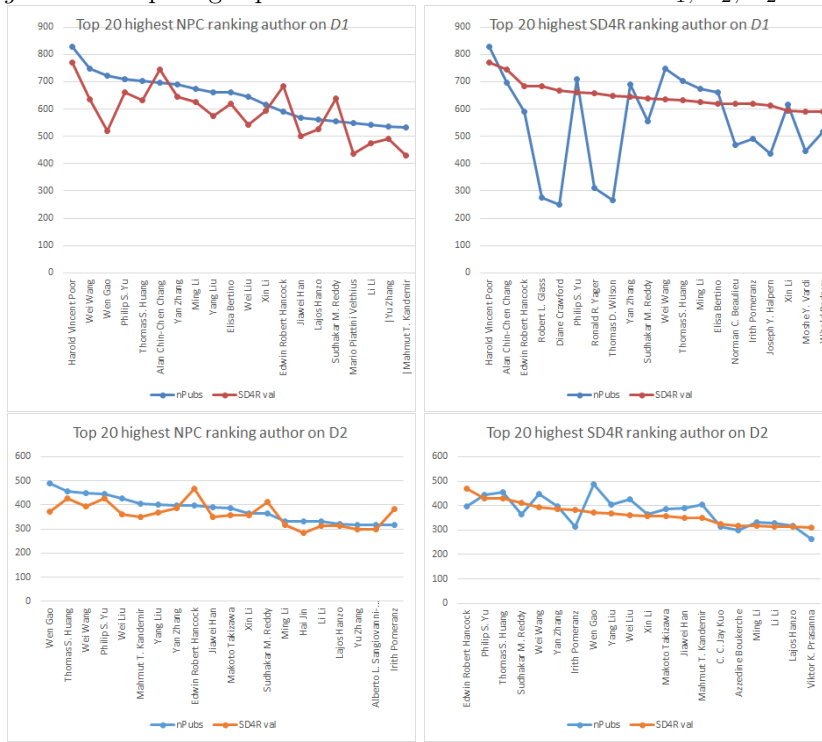
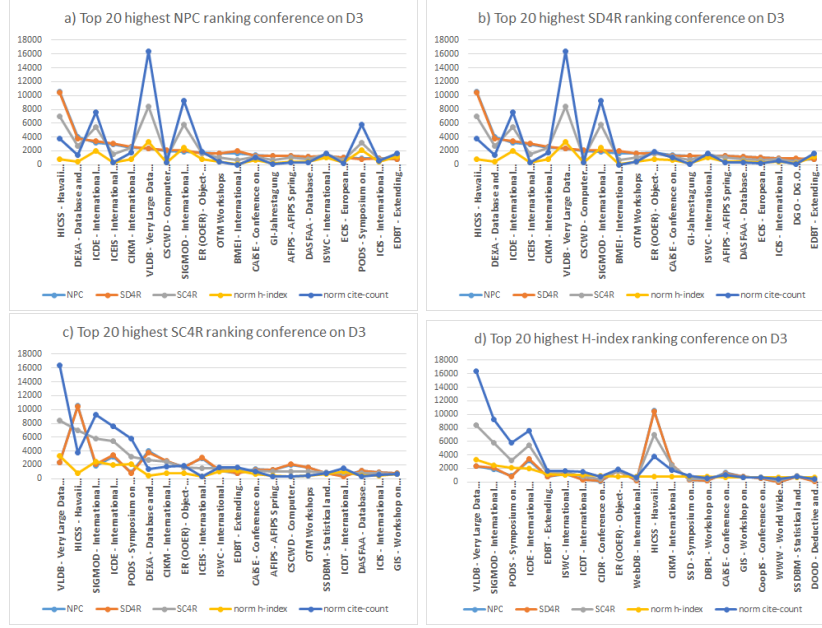
Figure 3. Comparing top 20 authors NPC vs. SD4R over  $D_1, D_2$ ,  $\alpha_2 = 0.45$ 

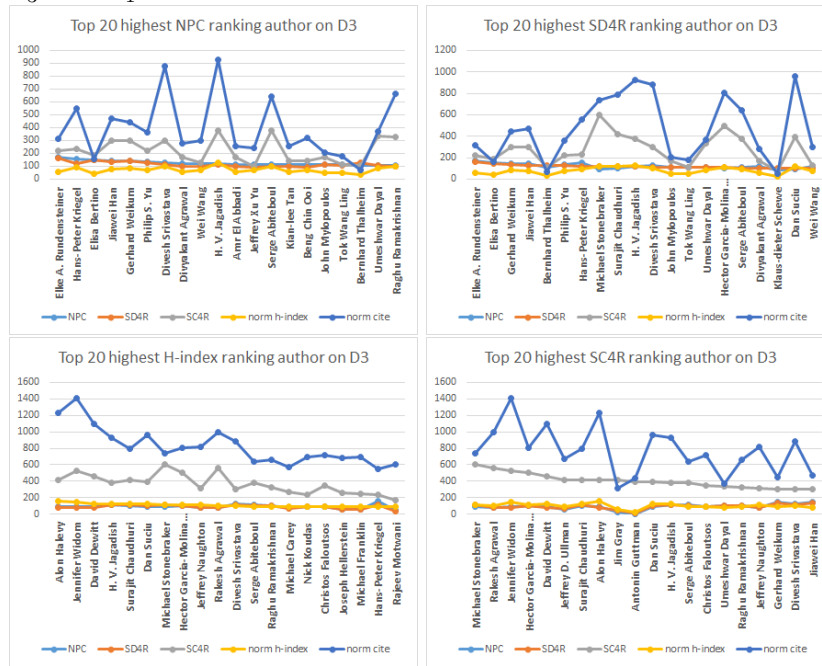
Figure 4. Top 20 ranking value of conference by NPC, SD4R, H-index and SC4R on  $D_3$  with  $\alpha_1 = 0.3$



### 5.5.2. With citation information

- *SC4R rank conferences prestigiously.* Fig. ?? shows the significantly different of SC4R model when comparing to H-index, SD4R and NPC models. While NPC and SD4R rank conferences similarly, SC4R can reflect the prestige of conferences better. Except HICSS - Hawaii International Conference on System Sciences having more than ten thousand papers, VLDB - Very Large Data Bases, SIGMOD - International Conference on Management of Data, ICDE - International Conference on Data Engineering and PODS - Symposium on Principles of Database Systems, listing in the top five of SC4R ranking conferences, are very noble in database field following the assessment of experts.
- *SC4R reflects the contribution of the author better than others.* Fig ?? has show the top 20 highest rank value of authors for each method over  $D_3$ .

Figure 5. Top 20 ranking value of author by NPC, SD4R, H-index and SC4R on  $D_3$  with  $\alpha_1 = 0.3$



## 6. Related work

Link-based ranking has been applied in many application in various domains. In this section, we want to describe such applications and compare them.

- Webpages: One of the most well-known example is PageRank utilized by Google search engine. Since the hyperlink structure among the webpages are easily represented as a web graph, the PageRank of each webpage can be measured (see Sect. ?? for more detail).
- 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 [?] proposes a ranking method based on a maximum-entropy tagger. Also, Vercoustre et al. [?] 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) [?].
- 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 [?, ?]. The poprank model [?] 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 quantum ranking [?] is based on quantum navigation. Their formula is come from the quantum 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<sup>‡</sup>. In this work, we have proposed a novel conference ranking framework to integrate all possible evidence.

---

<sup>‡</sup><http://intelligent.pe.kr/ConfRank.txt>



## 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 is a 2-star ranking. It has two classes: the webpages (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. Keener.** The Perron-Frobenius theorem and the ranking of football teams. *SIAM Review*, **Vol. 35(1)**, (1993), 80–93.
- [7] **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
- [8] **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.
- [9] **Michael Ley.** DBLP - Some Lessons Learned. *PVLDB* 2 (2), (2009) 1493–1500.
- [10] **Michael Ley, Patrick Reuther.** Maintaining an Online Bibliographical Database: The Problem of Data Quality. *EGC 2006*, (2006) 5–10.
- [11] **Microsoft Corporation.** Microsoft Academic Search. <http://academic.research.microsoft.com/> (June -26 -2013).
- [12] **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
- [13] **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.
- [14] **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.
- [15] **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.
- [16] **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).

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 hiuekien@hoatmail.com

**Jason J. Jung**  
Yeungnam University  
Gyeongsan  
Korea  
e-mail adress