Journal of Informetrics 6 (2012) 370–388

Contents lists available at SciVerse ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Time-aware PageRank for bibliographic networks

Dalibor Fiala*

University of West Bohemia, Univerzitní 8, 30614 Plzeň, Czech Republic

ARTICLE INFO

Article history:
Received 8 November 2011
Received in revised form 10 January 2012
Accepted 8 February 2012

Keywords:
PageRank
Citations
Collaboration
Time
Salient researchers
Computer science

ABSTRACT

In the past, recursive algorithms, such as PageRank originally conceived for the Web, have been successfully used to rank nodes in the citation networks of papers, authors, or journals. They have proved to determine prestige and not popularity, unlike citation counts. However, bibliographic networks, in contrast to the Web, have some specific features that enable the assigning of different weights to citations, thus adding more information to the process of finding prominence. For example, a citation between two authors may be weighed according to whether and when those two authors collaborated with each other, which is information that can be found in the co-authorship network. In this study, we define a couple of PageRank modifications that weigh citations between authors differently based on the information from the co-authorship graph. In addition, we put emphasis on the time of publications and citations. We test our algorithms on the Web of Science data of computer science journal articles and determine the most prominent computer scientists in the 10-year period of 1996–2005. Besides a correlation analysis, we also compare our rankings to the lists of ACM A. M. Turing Award and ACM SIGMOD E. F. Codd Innovations Award winners and find the new time-aware methods to outperform standard PageRank and its time-unaware weighted variants.

1. Introduction

When Brin and Page made public their PageRank algorithm in 1998 (Brin & Page, 1998), they would probably hardly have imagined what an enormous impact on computer science this would have in the decade to come. They presented a straightforward method of computing the importance of Web pages using the link structure of the (then still relatively new) World Wide Web. The same concept of "authoritativeness" of Web pages was, at approximately the same time, invented independently by Kleinberg (1999). The idea was surprisingly simple: if a link from one Web page to another one can be viewed as a vote then popular pages will have many in-links. In addition, if those in-links come from pages that themselves have many in-links, popularity becomes prestige. It was soon discovered that this recursive technique (applied successfully by Google) could be used to evaluate (rank) nodes in any (directed) graph. Bibliographic citation networks of papers, authors, journals, institutions, or even countries are good examples of such graphs and some studies making use of PageRank or related methods to find prominent players in them are touched upon in Section 2. However, researchers also felt the need to weigh the edges in bibliographic citation networks (unlike the original PageRank which was unweighted) based on the differences between the Web graph and the citation networks. First, bibliographic networks often contain information that can add value to citations, e.g. citation counts or co-authorship information. Second, unlike the Web graph, bibliometric networks include time, e.g. publication (citation) years, which could also help weigh citations more discriminately. And third, bibliometric networks are static in that citations always point from newer to older publications and they can never be removed. Fiala,

* Tel.: +420 377 63 24 29; fax: +420 377 63 24 01.
E-mail address: dalfia@kiv.zcu.cz

1751-1577/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.joi.2012.02.002

Rousselot, and Ježek (2008) addressed the first problem. They assigned more or less weight to the edges in a citation graph of authors based on the information from a co-authorship graph. The principal assumption was that a citation from a colleague should contribute less to the prestige of the cited author than a citation from a "foreign" researcher. On the other hand, this penalization could be mitigated in some circumstances, for instance, if the number of common publications of those two authors is relatively small compared to the total number of publications by the authors. Time in terms of publication (and thus citation or collaboration) years was ignored in this scenario. However, it is clear that if a citation is made before any common papers are published, it should not be considered a "friendly" citation from a colleague. This problem is addressed in this article. As for the third issue (static citations), some proposals to solve this problem are mentioned in the section on related work.

The principal objectives of the research reported in this paper are as follows:

- Define "time-aware" modifications of the "bibliographic PageRank" formula based on the work by Fiala et al. (2008) that take into account the time (year) when articles are published and citations are made.
- Apply the time-aware as well as the original (time-unaware) bibliographic PageRank variants to a large citation network of computer science researchers to find out the most prominent computer scientists.
- Compare the rankings of researchers generated by the new methods with each other as well as with other established bibliometric techniques in terms of a correlation analysis and a confrontation with the lists of ACM A. M. Turing Award (Turing Award) and ACM SIGMOD E. F. Codd Innovations Award (Codd Award) winners.

This article is organized in the following way: after introducing PageRank and our research goals in Section 1, related work on measuring computer science and various modifications of PageRank is reviewed in Section 2. Afterwards, in Section 3, we describe in detail an extension to the standard PageRank that is suitable for bibliographic networks and that can exploit the time information present in them. Section 4 is concerned with the data to which we applied the novel methods and then we discuss the experimental results in Section 5. Finally, we draw the main conclusions and outline our future work in Section 6.

2. Related work

This section on related work consists of three main paragraphs. The first one is concerned with previous bibliometric work on computer science, which has, somewhat surprisingly, been relatively little explored in the past. The second paragraph enumerates the principal studies known to the author that have sought to add weights to the basic PageRank formula and, finally, research into time-based weighting of PageRank is presented in the third paragraph.

2.1. Computer science

Bar-Ilan (2010) studied how publication and citation counts of some highly cited computer science researchers changed after conference proceedings papers had been added to the Web of Science (WoS). Franceschet (2010) investigates prestige, popularity, and productivity of computer science researchers with regard to journal versus conference papers. He defines a prestigious computer scientist as an ACM A. M. Turing Award winner. Wainer, Goldenstein, and Billa (2011) studied how many publications by computer science researchers are not indexed by established bibliographic databases compared to other scientific fields and concluded that, on average, 66% of a computer scientist's work is unknown to the Web of Science. Bibliometric studies on computer science based on the data from the CiteSeer digital library are presented by Fiala (2011, in press).

2.2. PageRank and weighted PageRank

Bollen, Rodriguez, and Van De Sompel (2006) assigned weights based on the number of citations to the edges in the citation network of journals and computed weighted PageRanks for the journals. Chen, Xie, Maslov, and Redner (2007) calculate PageRank of papers from a set of physics journals. Different weighting and normalization schemes were applied to PageRank by Bergstrom (2007) and González-Pereira, Guerrero-Bote, and Moya-Anegón (2010) to compute journal prestige. The corresponding scores are called Eigenfactor (or article influence when related to papers) and SCImago Journal Rank (SJR), respectively. Ding (2011) computes weighted PageRank for authors in the information retrieval field. She assigns weights based on the number of publications or citations to nodes rather than edges, and experiments with various damping factors in the PageRank formula. A similar study for author co-citation networks is conducted by Ding, Yan, Frazho, and Caverlee (2009). Yan and Ding (2011) explore co-authorship networks in the informetrics field. They calculate PageRank for authors with different damping factors and draw the conclusion that the damping factor does not have much influence on ranking in this type of network. They also define a weighted PageRank in which more weight is assigned to authors with more citations. Ma, Guan, and Zhao (2008) computed PageRank for papers in the field of biochemistry and molecular biology. Xing and Ghorbani (2004) defined the "weighted PageRank" by multiplying the rank of each in-linking node by two factors: the in-degree of the current node divided by the sum of in-degrees of the nodes linked to by the in-linking node, and the out-degree of the current node divided by the sum of out-degrees of the nodes linked to by the in-linking node. This enabled

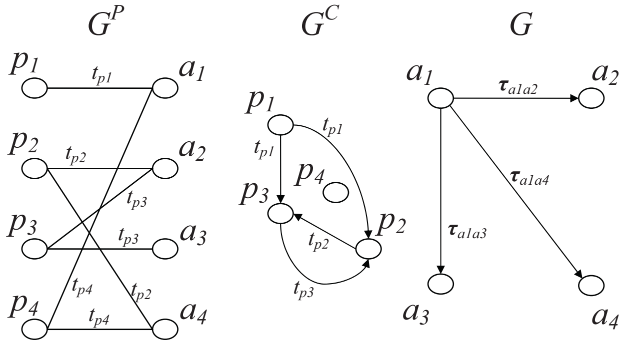

Fig. 1. Example of a co-authorship (G^P), publication citation (G^C), and author citation (G) graph.

more rank to be transferred to more "popular" nodes, i.e. to those that had relatively numerous in-links and/or out-links. The authors reported some success compared to the standard PageRank in obtaining more relevant results from a (very) small set of Web pages. Their approach does not seem reasonable in the case of citation networks of papers or authors because it is not clear why a paper (author) should be rewarded for citing many other papers (authors), i.e. the out-degree factor is doubtful. If just the in-degree factor was retained, their method would somewhat resemble the work by Ding (2011). Sidiropoulos and Manolopoulos (2005) adapted PageRank for publication citation networks in that they gave less weight to the citations from more distant publications (in terms of graph path). They were also the first to compare new ranking methods with established awards such as the ACM SIGMOD E. F. Codd Innovations Award.

2.3. Weighted PageRank considering time

Walker, Xie, Yan, and Maslov (2007) ranked publications in two distinct citation networks of physics papers. They included the age of publications in the PageRank algorithm by favouring citations from more recent articles. They also experimentally verified the until then theoretical concept that the average path length of a random surfer following citations between research publications is only around two. Yan and Ding (2010) also bring time into play when they give more weight to more recent citations (i.e. to the citations from publications that appear shortly after the cited papers). In addition, they more heavily weight citations from prestigious articles, but their prestige (article influence score) is not computed recursively in a self-contained way (like PageRank) but rather taken from a citation database. In their "TimedPageRank", Yu, Li, and Liu (2004) simply decrease the weight of a citation exponentially with the citation age using a base (decay rate) of 0.5. For the prediction of popularity a paper will enjoy in future years, they apply an "ageing factor" as well that linearly declines a paper's TimedPageRank with the paper's age.

In summary, all the authors of the above studies on (time-)weighted PageRank report its superiority to the standard PageRank but, at the same time, find a high correlation of various PageRank variants and other bibliometric measures such as citation counts. None of the studies, however, has combined time information from both the citation and collaboration graphs to rank computer science researchers via the "time-aware" PageRank described in this paper.

3. Methods

The methods of time-aware PageRank described in this paper are based on the techniques used by Fiala et al. (2008) by including the time factor in their PageRank modifications that take into account not only citations between authors but also other information such as the number of common publications between two authors linked by a citation. The key concept was that citations between authors should not be weighted the same but should rather be based on a number of factors reflecting the behaviour of authors. For instance, a citation between two authors who often collaborate with each other is considered less valuable than that between two authors who have never co-authored a single publication. We invite the reader to get more explanations and see examples in Fiala et al. (2008). In the following paragraphs, we will re-define "the bibliographic PageRank" from our previous work and expand it with time aspects so that it allows for the computation of "time-aware bibliographic PageRank".

3.1. Definitions

To understand Fig. 1, let $G^P = (P \cup A, E^P, \boldsymbol{T}^P)$ be an undirected, edge-weighted, bipartite graph (co-authorship graph), $P \cup A$ a set of vertices ($P = \{p_1, \ldots, p_n\}$ a set of publications, $A = \{a_1, \ldots, a_m\}$ a set of authors), $E^P \subseteq P \times A$ a set of edges, and $\boldsymbol{T}^P$ an $n \times m$ matrix of non-negative weights – publication years. Each edge $\{p, a\} \in E^P$, $p \in P$, $a \in A$ means that author a has (co-)authored publication p that appeared in year $\boldsymbol{T}^P_{pa}$. (If $\boldsymbol{T}^P_{ij} = 0$ then there is no such edge $\{i, j\}$ in E^P.) Let $G^C = (P, E^C, \boldsymbol{T}^C)$ be a directed edge-weighted graph (publication citation graph), $P = \{p_1, \ldots, p_n\}$ a set of vertices (the same set of publications), $E^C \subseteq P \times P$

a set of edges (citations between publications), and $\boldsymbol{T}^C$ an $n \times n$ matrix of non-negative weights – citation years. Each edge $\{p_1, p_2\} \in E^C$, $p_1 \in P$, $p_2 \in P$ means that publication p_1 from year $\boldsymbol{T}^C_{p_1 p_2}$ cites publication p_2. (If $\boldsymbol{T}^C_{ij} = 0$ then there is no such edge $\{i, j\}$ in E^C.) Now, we will combine the two graphs G^P and G^C into one more graph we will further work with. Let $G = (A, E)$ be a directed edge-weighted graph (author citation graph), $A = \{a_1, \ldots, a_m\}$ a set of vertices (the same set of authors) and $E \subseteq A \times A$ a set of edges (citations between authors). For every $p \in P$ let $A_p = \{a \in A{:}\ \exists\{p, a\} \in E^P\}$ be the set of authors of publication p. For each (a_1, a_2), $a_1 \in A$, $a_2 \in A$, $a_1 \neq a_2$ where there exists $(p_1, p_2) \in E^C$ such that $\{p_1, a_1\} \in E^P$ and $\{p_2, a_2\} \in E^P$ and $A_{p1} \cap A_{p2} = \varnothing$ (i.e. no common authors in citing and cited publications are allowed) there is an edge $(a_1, a_2) \in E$ (no parallel edges are admitted). Thus, $(a_1, a_2) \in E$ if and only if $\exists(p_1, p_2) \in E^C \wedge \exists\{p_1, a_1\} \in E^P \wedge \exists\{p_2, a_2\} \in E^P \wedge A_{p_1} \cap A_{p_2} = \varnothing \wedge a_1 \neq a_2$.

Before assigning weights to the edges in E, we further define:

- $w_{u,v} = |C|$ where $C = \{p_1 \in P{:}\ \exists\{p_1, u\} \in E^P \wedge \exists\{p_2, v\} \in E^P \wedge \exists\{p_1, p_2\} \in E^C \wedge p_1 \neq p_2\}$, as the number of citations from u to v;
- $f^t_{u,v} = |P^t_u| + |P^t_v|$ where $P^t_i = \{p \in P : \exists\{p, i\} \in E^P \wedge \boldsymbol{T}^P_{pi} < t\}$, as the number of publications by u appearing before year t plus the number of publications by v appearing before year t (called *publicationsT*); if $t = \infty$ (i.e. time is not taken into account), $f^t_{u,v}$ becomes $f_{u,v}$ (time-unaware, called *publications*);
- $c^t_{u,v} = |CP^t|$ where $CP^t = \{p \in P : \exists\{p, u\} \in E^P \wedge \exists\{p, v\} \in E^P \wedge \boldsymbol{T}^P_{pu} < t \wedge \boldsymbol{T}^P_{pv} < t\}$, as the number of common publications by u and v published before year t (called *collaborationT*); if $t = \infty$, $c^t_{u,v}$ becomes $c_{u,v}$ (called *collaboration*);
- $hd^t_{u,v} = |ADC^t_u| + |ADC^t_v|$ where $ADC^t_i = a \in A : \exists p \in P$ such that $\{p, a\} \in E^P \wedge \{p, i\} \in E^P \wedge \boldsymbol{T}^P_{pa} < t \wedge \boldsymbol{T}^P_{pi} < t$, as the number of all distinct co-authors of u in the papers published before year t plus the number of all distinct co-authors of v in the papers published before year t (called *allDistCoauthorsT*)); if $t = \infty$, $hd^t_{u,v}$ becomes $hd_{u,v}$ (called *allDistCoauthors*);
- $h^t_{u,v} = |ADC^t_u| + |ADC^t_v|$ where ADC^t_i is defined as above, but is a multiset, as the number of all co-authors of u in the papers published before year t plus the number of all co-authors of v in the papers published before year t (called *allCoauthorsT*); if $t = \infty$, $h^t_{u,v}$ becomes $h_{u,v}$ (called *allCoauthors*);
- $td^t_{u,v} = |DCA^t|$ where $DCA^t = \{a \in A{:}\ \exists p \in P$ such that $\{p, a\} \in E^P \wedge \{p, u\} \in E^P \wedge \{p, v\} \in E^P \wedge \boldsymbol{T}^P_{pu} < t \wedge \boldsymbol{T}^P_{pv} < t$, as the number of distinct co-authors in the common publications by u and v appearing before year t (called *distCoauthorsT)*; if $t = \infty$, $td^t_{u,v}$ becomes $td_{u,v}$ (called *distCoauthors*);
- $t^t_{u,v} = |DCA^t|$ where DCA^t is defined as above, but is a multiset, as the number of co-authors in the common publications by u and v appearing before year t (called *allDistCoauthorsT)*; if $t = \infty$, $t^t_{u,v}$ becomes $t_{u,v}$ (called *allDistCoauthors*);
- $g^t_{u,v} = f^t_{u,v} - |SP^t_u| - |SP^t_v|$ where $SP^t_i = \{p \in P : \{p, i\} \in E^P \wedge d_{G^P}(p) = 1 \wedge \boldsymbol{T}^P_{p_i} < t$, as the number of publications by u that appeared before year t, where u is not the only author, plus the number of publications by v that appeared before year t, where v is not the only author (called *allCollaborationsT*); if $t = \infty$, $g^t_{u,v}$ becomes $g_{u,v}$ (called *allCollaborations*).

3.2. Time-aware PageRank

Now, we associate a vector of weight pairs $\boldsymbol{\tau}_{uv} = ((c^{t1}_{u,v}, b^{t1}_{u,v}), (c^{t2}_{u,v}, b^{t2}_{u,v}), \ldots (c^{tk}_{u,v}, b^{tk}_{u,v})^k)$ with each edge $(u, v) \in E$, where $k = w_{u,v}$ (the number of citations from author u to author v) and $t1, \ldots, tk$ are the citation years selected as all those non-zero elements $\boldsymbol{T}^C_{ij}$, where $i \in P_u$ and $j \in P_v$, and we denote $P_a = \{p \in P{:}\ \exists\{p, a\} \in E^P\}$ as the set of publications of every author $a \in A$. $w_{u,v}$ and $c^t_{u,v}$ are described above, and $b^t_{u,v}$ can be equal to one of the seven following values according to the semantics of edge weights we want to stress: (a) zero, (b) $f^t_{u,v}$, (c) $h^t_{u,v}$, (d) $hd^t_{u,v}$, (e) $g^t_{u,v}$, (f) $t^t_{u,v}$,)g) $td^t_{u,v}$. We then define the rank $R(u)$ for author u as follows, bearing in mind that the superscript i means an index in vector $\boldsymbol{\tau}$ and not a year:

$$R(u) = \frac{1-d}{|A|} + d \sum_{(v,u)\in E} R(v) \frac{\sum_{i=1}^{w_{v,u}} 1/((c^i_{v,u}+1)/(b^i_{v,u}+1)\sum_{(v,j)\in E} 1)}{\sum_{(v,k)\in E}\sum_{i=1}^{w_{v,k}} 1/((c^i_{v,k}+1)/(b^i_{v,k}+1)\sum_{(v,j)\in E} 1)} \tag{1}$$

If we wish to ignore time (i.e. publication and citation years) and set all the coefficients $t1, \ldots, tk$ to infinity, vector $\boldsymbol{\tau}_{uv}$ takes the form $((c_{u,v}, b_{u,v})^1, (c_{u,v}, b_{u,v})^2, \ldots, (c_{u,v}, b_{u,v})^k)$ and Eq. (1) can be re-written as

$$R(u) = \frac{1-d}{|A|} + d \sum_{(v,u)\in E} R(v) \frac{w_{v,u}/((c_{v,u}+1)/(b_{v,u}+1)\sum_{(v,j)\in E} w_{v,j})}{\sum\limits_{(v,k)\in E} w_{v,k}/((c_{v,k}+1)/(b_{v,k}+1)\sum_{(v,j)\in E} w_{v,j})}, \tag{2}$$

which is exactly how the time-unaware modifications of PageRank were defined by Fiala et al. (2008). These modifications penalized citations by colleagues (influence of c) but relaxed the penalty in some circumstances such as a great number of co-authors (influence of b). Now we can easily show how Eq. (2) can be further reduced to the standard PageRank formula. First, we set all b's to zero and take into account only the collaboration coefficients c:

$$R(u) = \frac{1-d}{|A|} + d \sum_{(v,u)\in E} R(v) \frac{w_{v,u}/\left((c_{v,u}+1)\sum_{(v,j)\in E} w_{v,j}\right)}{\sum_{(v,k)\in E} w_{v,k}/\left((c_{v,k}+1)\sum_{(v,j)\in E} w_{v,j}\right)}. \tag{3}$$

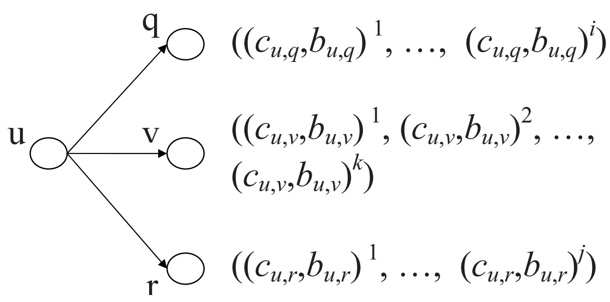

Fig. 2. Example of an author citation graph E with weight vectors $\boldsymbol{\tau}$.

Second, we disregard the co-authorship information by setting all c's to zero and obtain the weighted PageRank formula, in which the edges in the author citation graph G are weighted with w's:

$$R(u) = \frac{1-d}{|A|} + d \sum_{(v,u)\in E} R(v) \frac{w_{v,u} / \left(\sum_{(v,j)\in E} w_{v,j}\right)}{\sum_{(v,k)\in E} w_{v,k} / \left(\sum_{(v,j)\in E} w_{v,j}\right)} \tag{4}$$

And third, we set all the edge weights w in G to 1 and receive a standard PageRank formula which is equivalent to that by Brin and Page (1998):

$$R(u) = \frac{1-d}{|A|} + d \sum_{(v,u)\in E} R(v) \frac{1/(D_{out}(v))}{\sum_{(v,k)\in E} 1/(D_{out}(v))}, \tag{5}$$

where d is the damping factor (set to 0.9 in our experiments) and $D_{out}(v)$ is the out-degree of vertex v. The damping factor represents the probability of following a link from the current node in the graph. Brin and Page (1998) set it to 0.85, Walker et al. (2007) recommend it to be 0.5 for publication citation networks. However, there is no consensus yet on what the damping factor should be in author citation graphs.

The edge weights are pre-computed so the convergence of the PageRank modifications above does not differ from the standard PageRank (depending on d). In our experiments (see Section 5), the rankings became relatively stable after 20–30 iterations and we always continued to 50 iterations at most.

3.3. Example

In Fig. 2 there is a simplified example of an author citation graph E with four nodes u, q, v, r, three edges (u, q), (u, v), and (u, r) and three weight vectors $\boldsymbol{\tau}_{uq}$, $\boldsymbol{\tau}_{uv}$, and $\boldsymbol{\tau}_{ur}$ assigned to them.

Now, let us suppose that u cites v three times, in 1980, 1990, and 2000. For the sake of simplicity, we will assume that all the coefficients b are equal to zero, i.e. we will solely rely on the citations and collaborations between authors. We will consider two cases. In the first case, c is 0 in 1980 (i.e., the number of common publications of u and v before 1980 was 0), 1 in 1990, and 2 in 2000. In the second case, c is 2 in 1980, 1990, and 2000 (see Fig. 3). The interpretation of the scenarios might be the following: when author u first cited author v in 1980, they did not know each other yet (scenario 1 on the left-hand side of Fig. 3). When u cited v for the second time ten years later, they were colleagues already and had written one common publication in the meantime. At the time of the third citation in 2000, their co-authorship relation was even stronger because they already had two common publications (still scenario 1). This is quite different from scenario 2 on the right-hand side of Fig. 3, in which u and v probably know each other well in 1980 when u cites v for the first time as they already had two common publications at that time. However, they did not write any more articles together and their

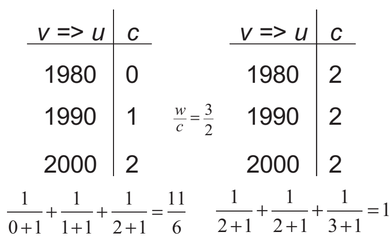

Fig. 3. Example of time-unaware (left) and time-aware (right) citation weighting.

collaboration count c remains unchanged in 1990 and 2000 when u repeatedly cites v. If we ignore the citation years, the contribution (or weight) of the citations from u to v is 3/2 in both scenarios, which is the nominator in Eq. (2) if all b's are 0. But, somehow, we feel that it is unjust and that the citation in 1980 should be weighted more if the authors do not know each other (left) than if they had already published together (right). Similarly, but perhaps less strictly, it happens in 1990 if the co-authorship relation between the authors is weaker (left) and stronger (right). Therefore, the co-authorship and other information entering the PageRank computation should always reflect the time of citation. This is exactly what we do in our time-aware PageRank modifications and formalize it in Eq. (1).

At the bottom of Fig. 3, we can see the time-aware contributions of the individual citations. In 1980 the contribution is 1 (left) and 1/3 (right), in 1990 it is 1/2 (left) and 1/3 (right), and, finally, in 2000 it is 1/3 in both cases. Thus, the total weight of citations in scenario 1 is 11/6, almost twice as much as that in scenario 2. Therefore, we may feel that the time-aware weighting has brought more justice to the prestige computation.

4. Data

To conduct practical experiments with the new evaluation method (time-aware PageRank), we needed to acquire some real-world data. For this purpose, we decided to download publication data from the Web of Science database, which is a well established data source for bibliometric studies. As we were only interested in the field of computer science, it was first necessary to determine the field boundaries. Since WoS does not enable the science domain to be specified in a straightforward way, we were forced to limit ourselves to publications appearing in journals classified as computer science sources. To compile such a list of relevant journals, in March 2011, we consulted the Journal Citation Reports® Science Edition 2009 (the most recent JCR at that time) in the following seven computer science subject categories: artificial intelligence, cybernetics, hardware and architecture, information systems, interdisciplinary applications, software engineering, and theory and methods. The list contained 426 journals whose names we could use in the search queries submitted programmatically to the WoS web services via their API. The time period we were interested in was the decade at the turn of the century: 1996–2005. Name changes of journals in that period were not taken into account. Unfortunately, the "lite" version of WoS web services does not allow the specifying of what document types are to be retrieved, nor is the document type information available in the documents retrieved. Therefore, we simply downloaded meta data from the Science Citation Index on all documents (of any type) published in those 426 journals in the years 1996–2005. In this way, we obtained 205 780 "core" documents (strictly stated, their meta data such as title, authors, source, year, etc.). The next step was to find citations to these core documents from documents published up to December 31, 2010. To this end, we submitted further queries to WoS web services to determine citing documents for each individual core document. We found 1 569 057 citations from a total of 643 302 citing documents. Of the citing documents, only 91 728 were core documents for which all meta data were available. As a result, we were concerned with the analysis of 276 957 citations between core documents. In the core documents themselves, there were 187 016 different authors (disambiguated just by their surnames and given names' initials) with 1 471 312 citations between them (without self-citations). The results discussed in Section 5 are based on the author citation graph. Detailed statistics of the data retrieved from WoS will be given in a separate article.

The data collection we have chosen has an obvious limitation: it is biased towards computer scientists who prefer publishing their research in journals, although it is well known that computer science research is presented at conferences to a greater extent than other fields of science (Bar-Ilan, 2010; Franceschet, 2010; Wainer et al., 2011). On the other hand, computer science journal articles receive more citations on average than conference papers (Franceschet, 2010) and we can expect that with a growing pressure on the visibility of papers and a faster journal editorial process, both of which we have been witnessing in recent years, the need for publishing computer science research in journals will increase.

5. Results and discussion

Table 1 shows the standings of the top 50 researchers as calculated by the "basic" methods – citation counts, in-degree, HITS, standard PageRank and weighted PageRank. By definition, citation counts are always greater or equal to in-degree. Since authors are not disambiguated, some names evidently represent more people with the same name as we can easily convince ourselves using a bibliographic database, e.g. in the case of "Jain, AK" or "Tanaka, K". On the other hand, some other names are apparently unique, e.g. "Kanade, T". The top authors by citations, in-degree, and HITS are very much the same with "Jain, AK", "Pentland, A", "Duin, RPW", and "Kanade, T" always appearing at the top. The interpretation of "Sapiro, G" being more highly ranked than "Kanade, T" in citations but more lowly ranked in in-degree is that "Sapiro, G" received more citations than "Kanade, T" but from fewer authors than "Kanade, T" did. Top-ranked authors by PageRank and by weighted PageRank are different from the first three rankings but similar to each other, with "Srinivasan, GR" and "Murley, PC" being at the very top.

5.1. Time-aware versus time-unaware rankings

As far as the rankings by the "advanced" methods (both time-aware and time-unaware) are concerned, the top 50 researchers in each ranking are shown in Tables A.1–A.3 in Appendix A. There are 14 rankings in total, with seven pairs

Table 1
Top 50 researchers by citations, in-degree, HITS, and (weighted) PageRank.

	Citations		In-degree		HITS	PageRank	Weighted PR
1	Jain, AK	3103	Jain, AK	1912	Jain, AK	Srinivasan, GR	Srinivasan, GR
2	Pentland, A	1140	Pentland, A	851	Pentland, A	Murley, PC	Murley, PC
3	Duin, RPW	1103	Duin, RPW	769	Belhumeur, PN	Tang, HHK	Ziegler, JF
4	Sapiro, G	1036	Kanade, T	757	Duin, RPW	Freeman, LB	Freeman, LB
5	Kanade, T	1026	Gupta, A	681	Kriegman, DJ	Ziegler, JF	Tang, HHK
6	Tanaka, K	1018	Breiman, L	636	Kanade, T	Leinen, P	Leinen, P
7	Belhumeur, PN	971	Sapiro, G	634	Kikinis, R	Bey, J	Bey, J
8	Kriegman, DJ	964	Jain, R	631	Ayache, N	Juang, JG	Juang, JG
9	Scholkopf, B	959	Ayache, N	624	Jain, R	Juang, HG	Juang, HG
10	Breiman, L	952	Picard, RW	624	Smeulders, AWM	Korec, I	Curtis, HW
11	Viergever, MA	937	Belhumeur, PN	623	Kittler, J	Cegielski, P	Montrose, CJ
12	Kikinis, R	933	Viergever, MA	602	Maes, F	Wiener, N	Muhlfeld, HP
13	Wang, HO	933	Kittler, J	598	Vandermeulen, D	Muses, C	OGorman, TJ
14	Osher, S	920	Kriegman, DJ	596	Sapiro, G	Litkowski, KC	Ross, JM
15	Bates, DW	917	Kikinis, R	585	Hespanha, JP	McTavish, DG	Korec, I
16	Hyvarinen, A	896	Scholkopf, B	569	Suetens, P	Gazarik, MJ	Wiener, N
17	Jain, R	868	Hyvarinen, A	564	Duncan, JS	Kamen, EW	Cegielski, P
18	Muller, KR	868	Cox, IJ	562	Wells, WM	Prou, JM	Taber, AH
19	Calderbank, AR	866	Yu, PS	560	Viergever, MA	Wagneur, E	Walsh, JL
20	Tse, DNC	866	Lee, J	558	Picard, RW	Fidelman, U	Muses, C
21	Picard, RW	864	Muller, KR	544	Gupta, A	Ristow, GH	Litkowski, KC
22	Ayache, N	855	Huang, TS	542	Santini, S	Myers, JS	McTavish, DG
23	Gupta, A	852	Black, MJ	530	Huang, TS	Sampson, G	Gazarik, MJ
24	Kittler, J	838	Burges, CJC	515	Hawkes, DJ	Thomason, A	Kamen, EW
25	Yu, PS	802	Smeulders, AWM	508	Hill, DLG	Yngve, VH	Prou, JM
26	Hill, DLG	800	Osher, S	498	Poggio, T	Behbehani, J	Wagneur, E
27	Bezdek, JC	798	Szeliski, R	495	Moghaddam, B	Robinson, DL	Renegar, J
28	Tarokh, V	781	Bates, DW	489	Worring, M	Schwarzer, S	Fidelman, U
29	Hawkes, DJ	766	Oja, E	485	Sclaroff, S	Wachmann, B	Ristow, GH
30	Bro, R	764	Duncan, JS	482	Marchal, G	Wang, WY	Simon, DR
31	Black, MJ	757	Manjunath, BS	481	Manjunath, BS	Curtis, HW	Robinson, DL
32	Cox, IJ	748	Foster, I	480	Black, MJ	Montrose, CJ	Myers, JS
33	Duncan, JS	733	Zhu, SC	479	Zhu, SC	Muhlfeld, HP	Sampson, G
34	Shortliffe, EH	729	Santini, S	476	Collignon, A	OGorman, TJ	Thomason, A
35	Cimino, JJ	717	Suetens, P	475	Scholkopf, B	Ross, JM	Yngve, VH
36	Shahar, Y	713	Jain, A	471	Baluja, S	Taber, AH	Vazirani, U
37	Yager, RR	712	Flynn, PJ	470	Rowley, HA	Walsh, JL	Bernstein, E
38	Amari, S	711	Bezdek, JC	465	Mao, JC	Russell, CA	Schwarzer, S
39	Huang, TS	710	Thrun, S	461	Grimson, WEL	Chin, B	Wachmann, B
40	Suetens, P	710	Wells, WM	459	Prince, JL	Enger, TA	Wang, WY
41	Oja, E	709	Kim, J	457	Taylor, CJ	Hosier, P	Russell, CA
42	Musen, MA	692	Shortliffe, EH	455	Muller, KR	Klein, WA	Chin, B
43	Maes, F	689	Poggio, T	453	Jain, A	LaFave, LE	Enger, TA
44	Vandermeulen, D	689	Malik, J	452	Kimmel, R	Messina, B	Hosier, P
45	Wells, WM	685	Schapire, RE	452	Cox, IJ	Nicewicz, M	Klein, WA
46	Kimmel, R	683	Sejnowski, TJ	450	Jolesz, FA	Orro, JM	LaFave, LE
47	Zhu, SC	681	Maes, F	449	Matas, J	Scott, TS	Messina, B
48	Lee, J	668	Vandermeulen, D	449	Vailaya, A	Sullivan, TD	Nicewicz, M
49	Jain, A	666	Kumar, V	448	Rueckert, D	Sussman, RJ	Orro, JM
50	Schapire, RE	664	Jennings, NR	447	Ma, WY	Sykes, AJ	Scott, TS

of rankings, one of which is always the time-aware variant of the other: collaboration, publications, co-authors, distinct co-authors, all collaborations, all co-authors, and all distinct co-authors. The top-ranked authors by all methods are very much the same, e.g. with “Srinivasan, GR”, “Murley, PC”, and “Ziegler, JF” in high positions in each ranking. In fact, how similar are the individual rankings as a whole? Tables 2–4 examine this aspect. In Table 2 we can see how the time-aware methods

Table 2
Spearman’s rank correlation coefficients of time-aware rankings.

	collaborationT	*publicationsT*	*allCo-authorsT*	*allDistCo-authorsT*	*allCollaborationsT*	*co-authorsT*	*distCo-authorsT*
collaborationT	**1**	0.975977	0.968838	0.973377	0.975813	0.999031	0.999095
publicationsT	0.975977	**1**	0.990443	0.990434	0.996298	0.976097	0.976147
allCoauthorsT	0.968838	0.990443	**1**	0.995957	0.992759	0.969081	0.969119
allDistCoauthorsT	0.973377	0.990434	0.995957	**1**	0.992303	0.973480	0.973527
allCollaborationsT	0.975813	0.996298	0.992759	0.992303	**1**	0.975932	0.975982
coauthorsT	0.999031	0.976097	0.969081	0.973480	0.975932	**1**	0.999905
distCoauthorsT	0.999095	0.976147	0.969119	0.973527	0.975982	0.999905	**1**

Table 3
Spearman's rank correlation coefficients of both kinds of rankings.

	collaborationT	*publicationsT*	*allCo-authorsT*	*allDistCo-authorsT*	*allCollaborationsT*	*co-authorsT*	*distCo-authorsT*
collaboration	**0.999923**	0.975964	0.968850	0.973384	0.975817	0.999022	0.999074
publications	0.993022	**0.971791**	0.964273	0.968614	0.971584	0.994887	0.994677
allCoauthors	0.985647	0.963150	**0.955729**	0.960214	0.963047	0.987923	0.987665
allDistCoauthors	0.990937	0.969841	0.962263	**0.966590**	0.969608	0.993019	0.992810
allCollaborations	0.993664	0.972378	0.964848	0.969202	**0.972109**	0.995404	0.995199
coauthors	0.998322	0.975931	0.968821	0.973308	0.975756	**0.999304**	0.999151
distCoauthors	0.998737	0.975948	0.968946	0.973337	0.975784	0.999652	**0.999572**

are correlated with each other. The table is symmetric and presents Spearman's rank correlation coefficients for each pair of time-aware rankings. The coefficients, which are all significant at the 0.01 level two-tailed, vary between 0.97 and 1 and suggest a very high correlation of all time-aware rankings. Similarly, very high Spearman's rank correlation coefficients can be observed in Table 3, which is non-symmetric and shows the correlation between time-aware and time-unaware PageRank variants. The most interesting figures are on the diagonal, where we can see how much new information is added if we use a time-aware variant instead of a standard PageRank modification. Provided that the lower the correlation achieved, the more new information is added using a time-aware method, the method taking into account all co-authors of an author prior to a citation (*allCoauthorsT*) instead of without regard to the citation time (*allCoauthors*) seems to work best. Table 4 is symmetric again. This time it shows how all the time-unaware rankings correlate with one another. The highest correlation can be observed with citations versus in-degree (0.997), the lowest correlation with HITS versus *allCoauthors* (0.730). All in all, HITS is relatively less correlated (0.74) with all the PageRank-based methods, but it is very highly positively correlated (0.93) with both of the first-order methods – citations and in-degree. As for the PageRank variants, their correlation coefficients with citations and in-degree are all around 0.83 and they are quite close to each other with correlations about 0.99. All the Spearman's rank correlation coefficients are significant at the 0.01 level two-tailed. The correlation between citations, in-degree, HITS, and (weighted) PageRank on the one side and the time-aware PageRanks on the other is not shown in Table 4, but the coefficients would be quite similar to those for the time-aware PageRanks regarding the high correlation between the time-aware and time-unaware rankings in Table 3.

5.2. ACM A. M. Turing Award winners

In a further experiment, we wanted to compare the rankings obtained by the various methods with a "true" human-made baseline ranking of some kind. In the computer science domain, such a "ranking" can be made of the list of ACM A. M. Turing Award laureates. Even though the list of award winners is actually not a ranking, it enables one to compare computer-generated lists of authoritative researchers with the scientists considered prestigious by their peers and has been successfully used in several comparative studies in the past (e.g. Fiala et al., 2008; Sidiropoulos & Manolopoulos, 2005). Table 5 shows the ranks of Turing Award winners from the years 1991 to 2010 (the past 20 years) produced by all of the 19 ranking methods described above. "Hartmanis, J" (1993), "Dahl, O" and "Nygaard, K" (2001), "Naur, P" (2005), and "Thacker, C" (2009) do not appear anywhere in the rankings and their rows are blank. The first two columns in Table 5 comprise citations and in-degree (the most frequently used research evaluation method) followed by HITS, PageRank, and weighted PageRank. Then there is a block of seven time-unaware PageRank modifications and a set of their seven time-aware counterparts. The ranks generated by the recursive techniques (from HITS onwards) were computed after fifty iterations (with the Spearman's rho between the rankings of two consecutive iterations being very close to 1 after just a few iterations) and are less important than the summary figures at the bottom of the table.

These numbers are the best rank, worst rank, average rank, medium rank, and standard deviation. Obviously, the lower the numbers the "better" the ranking in that it places the Turing Award winners higher (low ranks mean high positions). Therefore, an optimum ranking (with respect to the Turing Award) would place the awardees at ranks 1–23 (without those five researchers omitted) thus having a best rank of 1, worst rank of 23, average and median rank of 12 and a standard deviation of 6.63. Standard PageRank (in a darker column) achieves better indicators (except for the worst rank and standard deviation) than both citations and in-degree and much better than HITS, but its weighted variant does not seem to perform more efficiently. As far as the time-unaware PageRank modifications are concerned, their best ranks are better than citations or in-degree have but similar or worse than those of (weighted) PageRank. The same holds for the average rank and standard deviation. On the other hand, worst ranks and median ranks are almost always better than PageRank has. Approximately the same conclusions may be drawn for the time-aware modifications of PageRank with two exceptions: notably better average ranks were yielded by *allCoauthorsT* and especially by *allDistCoauthorsT*, i.e. by the methods that take into account the number of all co-authors of both the citing and cited author prior to a citation. All in all, there are many better indicators than PageRank achieved and these are highlighted. There are more of them in the time-aware methods than in the time-unaware ones. (The ratio is 21 to 14.) A graphical representation of the results in Table 5 is displayed in Fig. 4 (the award winners without ranks do not appear there).

In Fig. 4 we can see a general slight shift towards better (lower) ranks when moving from left to right, i.e. from citations and in-degree across recursive methods and PageRank modifications to the time-aware variants of PageRank. This would

Table 4
Spearman's rank correlation coefficients of time-unaware rankings.

	Cites	InDeg	HITS	PR	PR weighted	*collaboration*	*publications*	*allCoauthors*	*allDistCoauthors*	*allCollaborations*	*coauthors*	*distCoauthors*
Cites	**1**	0.9973	0.9269	0.8353	0.8322	0.8318	0.8295	0.8235	0.8277	0.8301	0.8322	0.8324
InDeg	0.9973	**1**	0.9284	0.8364	0.8311	0.8308	0.8283	0.8225	0.8266	0.8289	0.8311	0.8313
HITS	0.9269	0.9284	**1**	0.7538	0.7448	0.7445	0.7405	0.7301	0.7378	0.7415	0.7449	0.7450
PR	0.8353	0.8364	0.7538	**1**	0.9956	0.9956	0.9900	0.9831	0.9880	0.9906	0.9945	0.9950
PR weighted	0.8322	0.8311	0.7448	0.9956	**1**	0.9998	0.9936	0.9864	0.9916	0.9943	0.9987	0.9990
collaboration	0.8318	0.8308	0.7445	0.9956	0.9998	**1**	0.9928	0.9853	0.9906	0.9934	0.9982	0.9986
publications	0.8295	0.8283	0.7405	0.9900	0.9936	0.9928	**1**	0.9958	0.9989	0.9997	0.9959	0.9956
allCoauthors	0.8235	0.8225	0.7301	0.9831	0.9864	0.9853	0.9958	**1**	0.9972	0.9953	0.9894	0.9890
allDistCoauthors	0.8277	0.8266	0.7378	0.9880	0.9916	0.9906	0.9989	0.9972	**1**	0.9986	0.9943	0.9939
allCollaborations	0.8301	0.8289	0.7415	0.9906	0.9943	0.9934	0.9997	0.9953	0.9986	**1**	0.9964	0.9961
coauthors	0.8322	0.8311	0.7449	0.9945	0.9987	0.9982	0.9959	0.9894	0.9943	0.9964	**1**	0.9996
distCoauthors	0.8324	0.8313	0.7450	0.9950	0.9990	0.9986	0.9956	0.9890	0.9939	0.9961	0.9996	**1**

Table 5
ACM A. M. Turing Award winners (1991–2010) and their ranks.

Year	Winner	Citations	In-degree	HITS	PageRank	Weighted PageRank	*collaboration*	*publications*	*allCoauthors*	*allDistCoauthors*	*allCollaborations*	*coauthors*	*distCoauthors*
1991	Milner, R	16 590	12 612	50 753	12 234	11 814	11 428	17 335	22 964	22 025	17 123	14 437	14 152
1992	Lampson, B	6842	7698	18 864	4295	5504	5407	6744	7556	6901	6701	5839	5736
1993	Hartmanis, J												
1993	Stearns, RE	24 668	26 864	42 424	12 148	13 400	12 562	17 878	22 796	22 549	17 635	11 883	13 196
1994	Feigenbaum, EA	35 599	31 932	29 363	7737	7089	7431	3278	2268	2 357	3 261	5 464	5 436
1994	Reddy, R	14 257	14 998	11 659	7786	8412	8309	10 095	11 248	10 713	9 961	8 805	8 744
1995	Blum, M	8934	7618	33 665	6456	5745	5682	8227	9954	9 004	8 086	6 058	5 956
1996	Pnueli, A	7631	8704	20 785	6062	7066	7324	4522	3879	3 738	4 448	4 515	4 178
1997	Engelbart, D	73 807	74 009	58 367	72 511	72 129	72 108	72 588	72 651	72 505	72 569	72 139	72 133
1998	Gray, J	1139	1514	3469	814	1103	1112	674	462	406	668	639	602
1999	Brooks, FP	2609	3459	13 669	2937	3429	3376	4194	4435	4294	4177	3747	3646
2000	Yao, AC	21 696	23 911	36 410	13 812	11 711	11 487	22 671	28 340	25 165	17 860	12 916	12 558
2001	Dahl, O												
2001	Nygaard, K												
2002	Rivest, RL	17 719	19 912	27 857	18 337	24 745	24 682	26 376	27 077	26 454	26 254	25 228	25 038
2002	Shamir, A	12 309	11 345	16 354	978	971	1027	876	1448	1 043	873	926	914
2002	Adleman, LM	2975	3 204	23 617	90	75	73	240	545	297	222	101	96
2003	Kay, A	74 842	74 991	75 210	77 939	77 442	77 428	76 884	75 607	76 689	76 985	77 341	77 357
2004	Cerf, VG	18 173	20 372	40 052	24 344	24 738	24 354	29 368	30 313	29 937	29 143	25 971	25 792
2004	Kahn, RE	16 450	18 672	37 702	19 355	23 906	23 521	27 978	29 155	28 679	27 675	24 920	24 776
2005	Naur, P												
2006	Allen, F	6308	7178	8569	28 841	29 024	29 057	28 691	29 631	28 773	28 540	27 920	28 315
2007	Clarke, EM	683	869	5869	1080	1205	1182	1594	1828	1 622	1 562	1 346	1 291
2007	Emerson, EA	4859	3974	14 513	6082	4627	4604	5282	5932	5 451	5 226	4 666	4 659
2007	Sifakis, J	9186	10 953	18 865	16 937	18 788	24 930	7965	5824	5 504	7 855	7 581	6 832
2008	Liskov, B	11 662	9802	30 434	9165	10 230	9902	12 936	13 732	13 007	12 686	10 942	10 790
2009	Thacker, C												
2010	Valiant, LG	15 980	16 523	32 513	10 186	9711	9935	8420	8501	8 281	8 239	8 839	9 105
	Best rank	**683**	**869**	**3469**	**90**	**75**	**73**	**240**	**462**	**297**	**222**	**101**	**96**
	Worst rank	**74 842**	**74 991**	**75 210**	**77 939**	**77 442**	**77 428**	**76 884**	**75 607**	**76 689**	**76 985**	**77 341**	**77 357**
	Average rank	**17 605**	**17 875**	**28 304**	**15 658**	**16 211**	**16 388**	**17 166**	**18 093**	**17 626**	**16 859**	**15 749**	**15 709**
	Median rank	**12 309**	**11 345**	**27 857**	**9165**	**9711**	**9902**	**8420**	**9954**	**9 004**	**8 239**	**8 805**	**8 744**
	Rank std. dev.	**19 284**	**19 239**	**17 121**	**19 844**	**19 861**	**19 926**	**20 107**	**20 166**	**20 269**	**20 084**	**20 002**	**20 030**

Table 5 (*Continued*)

Year	Winner	*collaborationT*	*publicationsT*	*allCoauthorsT*	*allDistCoauthorsT*	*allCollaborationsT*	*coauthorsT*	*distCoauthorsT*
1991	Milner, R	11 488	14 080	13 682	12 330	16 113	13 527	13 274
1992	Lampson, B	5438	5294	6299	5628	5946	5551	5517
1993	Hartmanis, J							
1993	Stearns, RE	13 058	13 711	13 877	14 027	13 718	13 102	13 862
1994	Feigenbaum, EA	7108	8909	8223	7628	8332	7182	7179
1994	Reddy, R	8335	9652	8526	7990	11 168	8659	8621
1995	Blum, M	5656	9367	9769	9120	9255	5913	5830
1996	Pnueli, A	7395	6269	4127	3748	5803	3770	3689
1997	Engelbart, D	72 115	62 265	68 327	67 279	68 912	72 117	72 117
1998	Gray, J	1083	914	626	573	878	1180	1165
1999	Brooks, FP	3386	3956	3294	2997	4479	3553	3519
2000	Yao, AC	11 604	15 398	17 259	15 557	14 827	12 085	12 018
2001	Dahl, O							
2001	Nygaard, K							
2002	Rivest, RL	24 697	16 265	17 081	17 190	15 847	24 926	24 824
2002	Shamir, A	1010	1916	2222	2032	1725	941	939
2002	Adleman, LM	73	844	937	811	751	85	81
2003	Kay, A	77 415	74 852	74 667	75 163	74 126	77 433	77 437
2004	Cerf, VG	24 510	25 235	18 882	16 927	26 393	25 091	24 993
2004	Kahn, RE	23 676	26 755	15 643	13 908	25 437	24 197	24 130
2005	Naur, P							
2006	Allen, F	29 084	25 829	14 864	13 726	25 413	28 568	28 544
2007	Clarke, EM	1177	1025	835	809	941	1291	1254
2007	Emerson, EA	4625	3582	4083	4133	3480	4595	4577
2007	Sifakis, J	24 880	23 630	13 580	12 455	19 228	5985	5 844
2008	Liskov, B	9914	8631	6424	5632	8381	10 725	10 653
2009	Thacker, C							
2010	Valiant, LG	9994	8444	13 876	13 002	12 312	9172	9354
	Best rank	**73**	**844**	**626**	**573**	**751**	**85**	**81**
	Worst rank	**77 415**	**74 852**	**74 667**	**75 163**	**74 126**	**77 433**	**77 437**
	Average rank	**16 423**	**15 949**	**14 657**	**14 029**	**16 238**	**15 637**	**15 627**
	Median rank	**9914**	**9367**	**9769**	**9120**	**11 168**	**8659**	**8621**
	Rank std. dev.	**19 929**	**18 182**	**18 462**	**18 466**	**18 726**	**20 006**	**20 008**

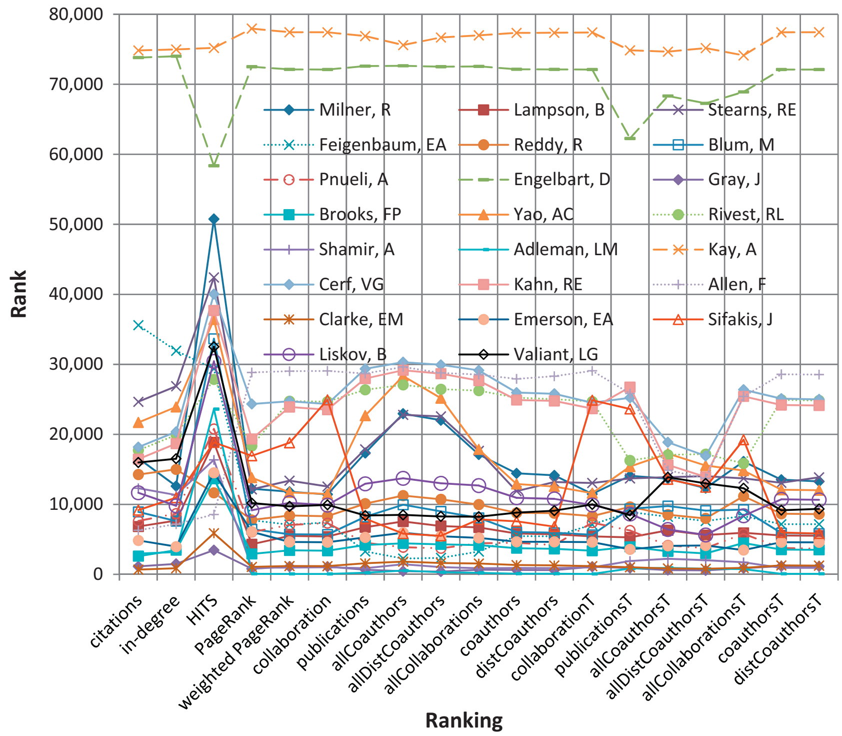

Fig. 4. ACM A. M. Turing Award winners and their ranks in different rankings.

suggest that the time-aware PageRank does reflect prestige perceived by humans (expressed by awards) better than common indicators such as citation counts or the standard PageRank and its weighted variations. Of course, there are some outliers in contradiction with this trend such as “Sifakis, J” and the sudden worsening of his rank with *collaborationT* and *publicationsT* or the overall bad performance of HITS for almost all of the authors, but this may also be interpreted as a feature of that particular ranking. For instance, the relatively bad ranks of “Sifakis, J” reveal that he has relatively frequently collaborated with the researchers citing him (both *collaboration* and *collaborationT*) and that he has written a great number of publications but rather after he was cited, thus having a good rank in *publications* and a bad rank in *publicationsT*. Some other authors, such as “Kay, A” or “Engelbart, D” are very badly ranked by almost all of the methods. This may be caused by the fact that they did not publish in journals in the time period under investigation. And indeed, they both have only three publications in our data set. But as we pointed out earlier, the individual ranks are less important and not discussed here than the overall trend, in which time-aware PageRanks seem to be closer to the “true” ranking than the other indicators.

5.3. ACM SIGMOD E. F. Codd Innovations Award winners

To bring additional evidence that would document the superiority of the time-aware methods over the time-unaware ones, we take advantage of yet another award – ACM SIGMOD E. F. Codd Innovations Award. The award winners from the years 1992 to 2011 are shown in Table 6 along with the ranks achieved in various rankings. (“Bayer, R” was not present in our data and, therefore, was not ranked.) Again, the ranks generated by the standard PageRank are in a darker column and the aggregate indicators yielded by both the time-unaware and time-aware rankings outperforming PageRank are highlighted. For instance, all seven worst ranks by time-aware methods outperform PageRank, but only one time-unaware worst rank does. In total, 24 time-aware indicators are better than PageRank compared to only 8 time-unaware ones. Also in Fig. 5 we can see that *allCoauthorsT* and *allDistCoauthorsT* generally produce better ranks for the award winners. The worst ranked researchers, “Kitsuregawa, M” and “Selinger, P”, published relatively few journal articles in the time period under study (14 and 3, respectively), but there is no such gap between them and the other laureates as in Fig. 4.

The better performance of the time-aware methods over their time-unaware counterparts is further documented in Figs. 6 and 7. In Fig. 6, the solid blue lines represent best ranks (MIN), worst ranks (MAX), average ranks (AVG), median ranks (MED), and standard rank deviations (DEV) of the time-unaware (standard) PageRank modifications and the dashed red lines represent the time-aware PageRank variants. As for the Turing Award, three dashed lines are below their solid counterparts – MAX, DEV, and AVG. This means that from the point of view of these three indicators the time-aware methods outperform the time-unaware ones by generating lower (i.e. better) ranks for the awardees. As far as the Codd Award is concerned, even

Table 6
ACM SIGMOD E. F. Codd Innovations Award winners and their ranks.

Year	Winner	Citations	In-degree	HITS	PageRank	Weighted PageRank	*collaboration*	*publications*	*allCoauthors*	*allDistCoauthors*	*allCollaborations*	*coauthors*	*distCoauthors*
1992	Stonebraker, M	13 816	11 821	15 821	18 112	23 700	23 559	26 392	28 690	27 276	26 242	25 528	25 366
1993	Gray, J	1514	1139	3469	814	1103	1112	674	462	406	668	639	602
1994	Bernstein, PA	560	429	2301	758	808	828	904	990	895	912	877	818
1995	DeWitt, DJ	14 670	15 354	20 509	25 963	28 031	27 983	26 237	24 510	25 481	26 211	28 510	28 296
1996	Mohan, C	12 167	14 863	10 582	9724	8948	8832	10 235	11 758	11 237	10 188	9415	9371
1997	Maier, D	7954	6859	10 832	8760	8655	9595	2077	1350	1290	2032	1936	1847
1998	Abiteboul, S	3054	3348	14 794	3934	3986	4361	795	563	722	820	2271	2817
1999	Garcia-Molina, H	1007	936	2003	2442	2720	2654	3842	4395	4010	3790	3024	2929
2000	Agrawal, R	533	395	1592	458	569	551	807	955	819	789	633	610
2001	Bayer, R												
2002	Selinger, P	70 765	70 420	60 364	58 514	56 330	55 759	63 652	71 922	69 288	63 339	59 105	58 678
2003	Chamberlin, D	44 810	43 091	41 095	34 935	38 619	37 837	44 033	50 969	49 257	44 038	43 336	42 732
2004	Fagin, R	1413	1033	1251	327	423	392	1352	2504	1795	1261	632	562
2005	Carey, MJ	5285	4209	6911	5556	6328	6363	6397	5028	5361	6483	6299	6113
2006	Ullman, JD	16 518	16 855	7118	24 271	23 886	23 664	27 892	33 144	28 701	27 544	24 246	24 541
2007	Widom, J	2676	2284	4440	2250	2216	2540	990	872	774	965	776	732
2008	Vardi, MY	1369	1605	16783	1066	1178	1214	556	534	689	553	1184	1203
2009	Kitsuregawa, M	75 050	74 905	49 529	70 797	70 681	70 576	72 726	74 800	73 536	72 636	71 108	70 852
2010	Dayal, U	31 806	29 660	25 204	44 691	45 438	45 368	45 022	43 704	44 106	44 948	45 463	45 436
2011	Chaudhuri, S	408	268	871	637	790	765	1143	1498	1277	1117	897	842
	Best rank	**408**	**268**	**871**	**327**	**423**	**392**	**556**	**462**	**406**	**553**	**632**	**562**
	Worst rank	**75 050**	**74 905**	**60 364**	**70 797**	**70 681**	**70 576**	**72 726**	**74 800**	**73 536**	**72 636**	**71 108**	**70 852**
	Average rank	**16 072**	**15 762**	**15 551**	**16 527**	**17 074**	**17 050**	**17 670**	**18 876**	**18 259**	**17 607**	**17 152**	**17 071**
	Median rank	**5285**	**4209**	**10 582**	**5556**	**6328**	**6363**	**3842**	**4395**	**4010**	**3790**	**3024**	**2929**
	Rank std. dev.	**22 588**	**22 424**	**16 814**	**20 885**	**20 940**	**20 777**	**22 605**	**24 306**	**23 668**	**22 551**	**21 895**	**21 779**

Year	Winner	*collaborationT*	*publicationsT*	*allCoauthorsT*	*allDistCoauthorsT*	*allCollaborationsT*	*coauthorsT*	*distCoauthorsT*
1992	Stonebraker, M	23 671	15 176	11 997	11 690	15 303	23 676	23 661
1993	Gray, J	1083	914	626	573	878	1180	1165
1994	Bernstein, PA	818	849	845	848	885	807	792
1995	DeWitt, DJ	28 058	26 351	25 610	24 888	25 787	27 985	27 888
1996	Mohan, C	8908	12 515	8763	6809	12 227	8926	8 889
1997	Maier, D	8568	5184	4445	4357	4826	8848	8 788
1998	Abiteboul, S	4602	2391	2386	2316	2280	2938	3 322
1999	Garcia-Molina, H	2659	1205	1070	1009	1132	2858	2 829
2000	Agrawal, R	553	419	414	323	391	586	595
2001	Bayer, R							
2002	Selinger, P	55 958	51 998	52 043	51 543	53 608	57 701	57 587
2003	Chamberlin, D	37 616	38 401	34 662	32 568	37 800	42 433	42 013
2004	Fagin, R	383	686	655	641	696	540	500
2005	Carey, MJ	6298	5919	6015	5311	5728	6516	6487
2006	Ullman, JD	23 699	11 647	11 295	11 307	11 580	24 132	23 972
2007	Widom, J	2162	1381	1530	1464	1352	2314	2254
2008	Vardi, MY	1219	727	769	828	713	1185	1168
2009	Kitsuregawa, M	70 590	61 126	59 358	59 795	60 820	70 792	70 707
2010	Dayal, U	45 389	36 150	33 212	35 192	35 105	45 321	45 339
2011	Chaudhuri, S	770	185	236	231	172	813	800
	Best rank	**383**	**185**	**236**	**231**	**172**	540	500
	Worst rank	**70 590**	**61 126**	**59 358**	**59 795**	**60 820**	**70 792**	**70 707**
	Average rank	**17 000**	**14 380**	**13 470**	**13 247**	**14 278**	**17 345**	**17 303**
	Median rank	**6298**	**5184**	**4445**	**4357**	**4826**	**6516**	**6487**
	Rank std. dev.	**20 822**	**18 614**	**18 007**	**18 043**	**18 653**	**21 316**	**21 259**

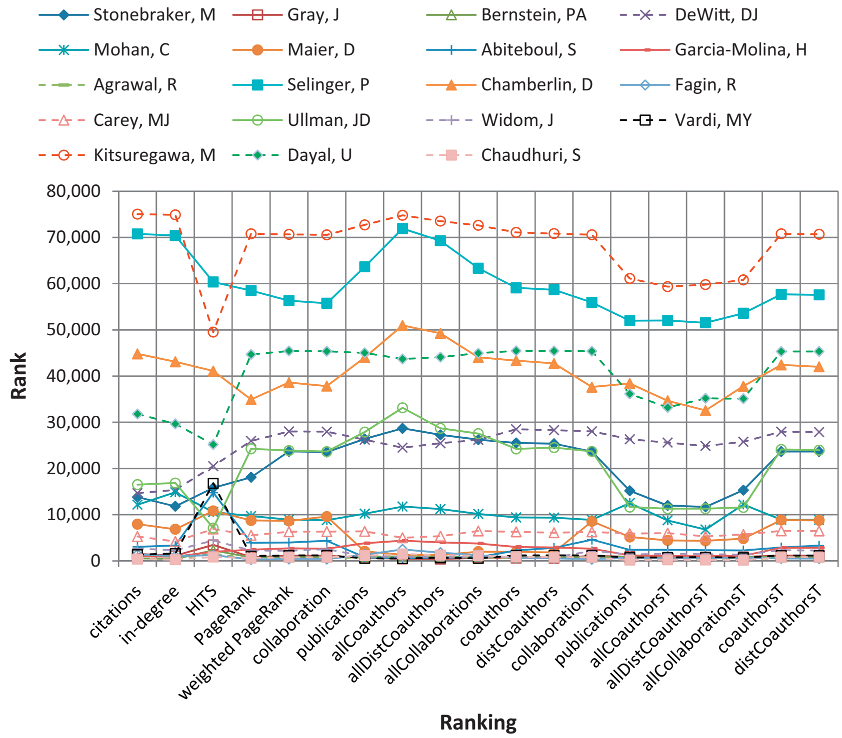

Fig. 5. ACM SIGMOD E. F. Codd Innovations Award winners and their ranks in different rankings.

four indicators speak in favour of the time-aware methods – MAX, DEV, AVG, and MIN. The only indicator that is worse with both method types is the median rank (MED), which is, however, not very distinct as the solid and dashed lines lie close to each other. In Fig. 7 box plots of the time-aware and time-unaware rankings are presented for each pair of rankings. In the case of both awards we can observe that the boxes of the time-aware rankings tend to be placed more towards lower (better) ranks than those of the time-unaware rankings.

6. Conclusions and future work

Algorithms based on the recursive technique called PageRank (Brin & Page, 1998), which was first applied to the Web graph in order to determine the significance of Web pages, have been successfully used in many other situations since then. These methods enable one to evaluate nodes in any directed graphs and rank them according to their importance. In bibliometrics, citation networks of papers or authors, among others, can represent such directed graphs in which the nodes are papers (or authors) and the edges are citations between them. The prominence of researchers has long been detected by first-order methods such as simple citation counts, but it has been shown that popularity, not prestige, is often reflected by citation numbers. On the contrary, higher-order (recursive) methods such as PageRank are able to find prestigious actors that may have fewer citations but from prestigious sources. Also, PageRank-like ranking methods for bibliographic networks can take advantage of the additional information that is not present in a Web graph to weight edges in the network, e.g. co-authorship (Fiala et al., 2008) or time data (Walker et al., 2007; Yan & Ding, 2010; Yu et al., 2004). Fiala et al. (2008) assigned different weights to the edges in a citation network of authors bearing in mind that a citation from a colleague was less valuable than that from a foreign researcher, but they did not distinguish whether the possible collaboration occurred before the citation was made or afterwards. In this article, we have made an attempt to remedy this situation. The main contributions of the research presented in this paper are as follows:

- We extended the model by Fiala et al. (2008) to incorporate the time of publications (and citations) in their “bibliographic PageRank” to create a “time-aware PageRank” for bibliographic networks. In this model, citations between researchers weight differently depending on a number of factors such as the number of common publications and whether or not they were published before a citation was made.
- We applied seven time-aware PageRank variants along with their time-unaware counterparts and five other common ranking methods (citations, in-degree, HITS, PageRank, and weighted PageRank) to the Web of Science data for computer science journal articles from the period 1996–2005 in order to find the most influential computer scientists publishing their work in journals in the decade at the turn of the century.

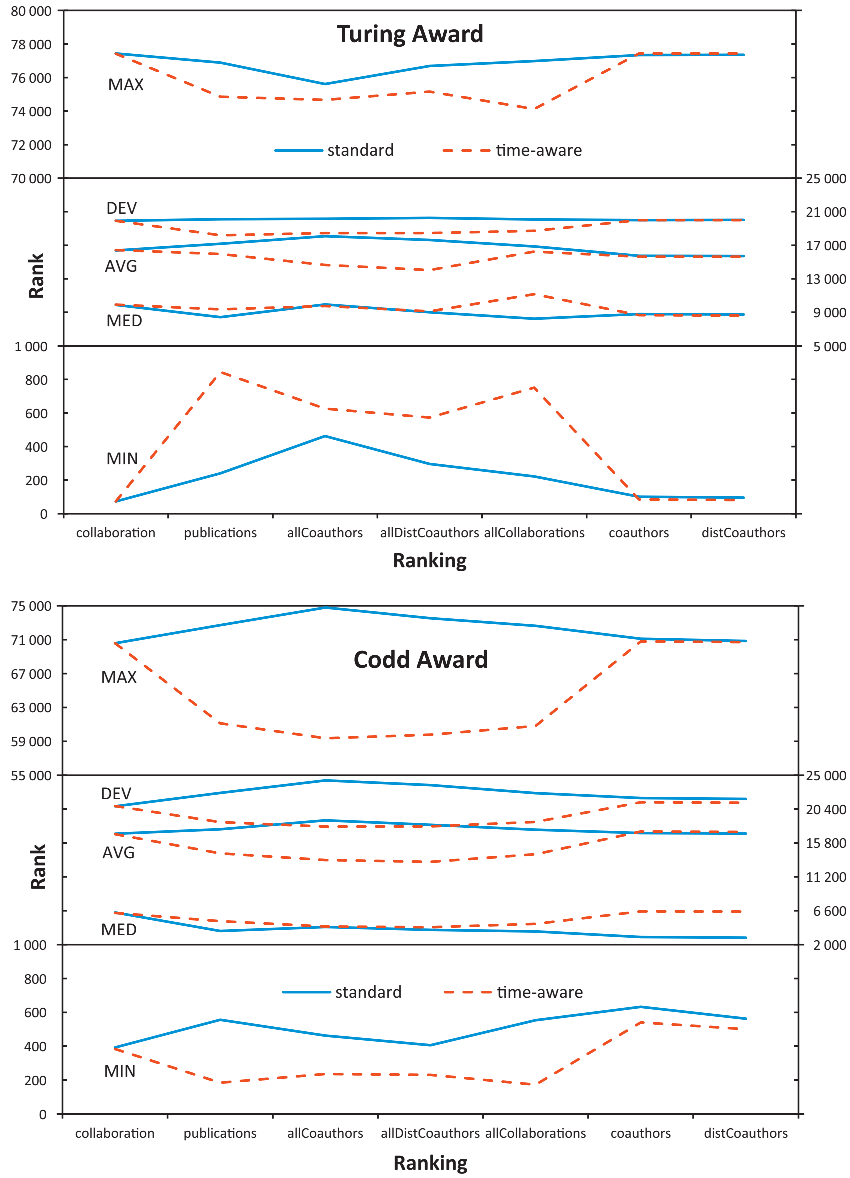

Fig. 6. Aggregate indicators of time-unaware (standard) and time-aware rankings.

- We conducted a thorough correlation analysis of the time-aware rankings themselves as well as of the time-aware and time-unaware rankings and other bibliometrics measures such as citations or in-degree. We also compared all the 19 rankings with the lists of ACM A. M. Turing Award laureates from the years 1991–2010 and ACM SIGMOD E. F. Codd Innovations Award winners from the years 1992–2011.

Based on our experiments, we achieved the following main results:

- All the 19 rankings are significantly highly positively correlated with each other. The very lowest correlation (around 0.74 of Spearman's rho) was found between HITS authorities and the other PageRank modifications. As for the new time-aware PageRanks, the lowest correlation (0.956), and thus the most added information when compared to its time-unaware counterpart, was observed between the variants in which the number of all co-authors in all publications of both the citing and cited authors are considered.
- The most prominent computer scientists contributing to WoS-indexed journals in the decade 1996–2005 detected by citations, in-degree, and HITS are "Jain, AK", "Pentland, A", and "Duin, RPW", whereas those determined by PageRank and all its variants are "Srinivasan, GR", "Murley, PC", and "Ziegler, JF".

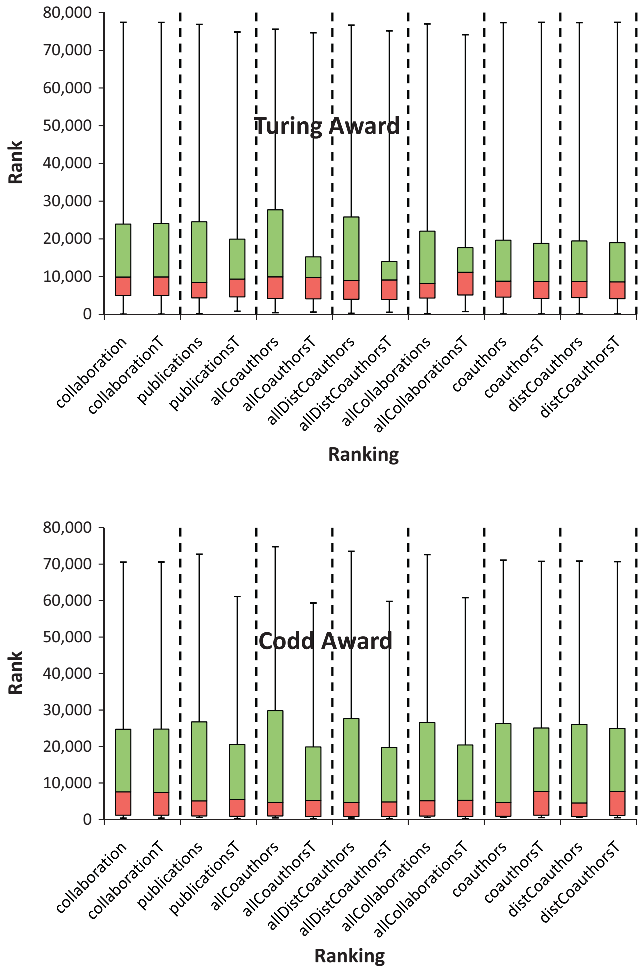

Fig. 7. Box plots of time-unaware and time-aware rankings.

- As far as the award winners are concerned, they generally receive better ranks in the time-aware rankings (as can be seen in Figs. 4 and 5), but it is impossible to proclaim the "best" ranking because each individual ranking brings an improvement in some aspect (see Tables 5 and 6). However, compared to the standard (unweighted) PageRank in terms of several statistical indicators, the time-aware variants outperform the time-unaware ones (see Tables 5 and 6 and Figs. 6 and 7).

For the time-aware PageRank modifications to be more effective, a greater citation window would probably be needed. This would result in a larger number of citations and collaborations of authors in different years. Then, the time-aware and time-unaware rankings should diverge from each other even more than in this study. Therefore, we would like to examine data spanning a greater time period in our future work on this promising topic. Other possibilities of adding more information to citations' weights would include investigating citation loops between authors and assigning less weight to the citations of authors who cite each other.

Acknowledgements

This work was supported by the European Regional Development Fund (ERDF), project "NTIS – New Technologies for Information Society", European Centre of Excellence, CZ.1.05/1.1.00/02.0090. The related software may be found at http://textmining.zcu.cz/downloads/tarank.php. Thanks are due to the anonymous reviewers for their insightful comments.

Appendix A.

Table A.1
Top 50 researchers by both kinds of rankings (part 1).

	collaboration	*collaborationT*	*publications*	*publicationsT*	*allCoauthors*	*allCoauthorsT*
1	Srinivasan, GR	Srinivasan, GR	Srinivasan, GR	Srinivasan, GR	Sudan, M	Srinivasan, GR
2	Murley, PC	Murley, PC	Ziegler, JF	Jain, AK	Ziegler, JF	Murley, PC
3	Tang, HHK	Ziegler, JF	Verdu, S	Murley, PC	Verdu, S	Jain, AK
4	Freeman, LB	Freeman, LB	Sudan, M	Ziegler, JF	Srinivasan, GR	Ziegler, JF
5	Ziegler, JF	Tang, HHK	Shamai, S	Freeman, LB	Sapiro, G	Freeman, LB
6	Leinen, P	Leinen, P	Murley, PC	Tang, HHK	Osher, S	Tang, HHK
7	Bey, J	Bey, J	Freeman, LB	Sudan, M	Shamai, S	Sudan, M
8	Juang, JG	Juang, JG	Tse, DNC	Shamai, S	Jain, AK	Calderbank, AR
9	Juang, HG	Juang, HG	Jain, AK	Tse, DNC	Tse, DNC	Renegar, J
10	Korec, I	Korec, I	Osher, S	Calderbank, AR	Bartlett, PL	Shamai, S
11	Curtis, HW	Curtis, HW	Sapiro, G	Cimino, JJ	Lin, YB	Pentland, A
12	Montrose, CJ	Montrose, CJ	Tarokh, V	Pentland, A	Kschischang, FR	Tse, DNC
13	Muhlfeld, HP	Muhlfeld, HP	Vardy, A	Kanade, T	Cimino, JJ	Osher, S
14	OGorman, TJ	OGorman, TJ	Kschischang, FR	Breiman, L	Vardy, A	Sapiro, G
15	Ross, JM	Ross, JM	Tang, HHK	Tarokh, V	Shortliffe, EH	Cimino, JJ
16	Wiener, N	Wiener, N	McEliece, RJ	Sapiro, G	Scholkopf, B	Gupta, A
17	Cegielski, P	Cegielski, P	Cimino, JJ	Sejnowski, TJ	Bates, DW	Sejnowski, TJ
18	Taber, AH	Taber, AH	Leinen, P	Gupta, A	McEliece, RJ	Viergever, MA
19	Walsh, JL	Walsh, JL	Lapidoth, A	Verdu, S	Tarokh, V	Kikinis, R
20	Muses, C	Muses, C	Arora, S	Lee, J	Arora, S	Kanade, T
21	Litkowski, KC	Litkowski, KC	Oja, E	Osher, S	Bro, R	Tarokh, V
22	McTavish, DG	McTavish, DG	Schapire, RE	Leinen, P	Duin, RPW	Lee, J
23	Gazarik, MJ	Gazarik, MJ	Bartlett, PL	Jain, R	Oja, E	Schapire, RE
24	Kamen, EW	Kamen, EW	Mesiar, R	Jordan, MI	Zuckerman, D	Yu, PS
25	Prou, JM	Prou, JM	Yager, RR	Yu, PS	Lapidoth, A	Freund, RM
26	Wagneur, E	Wagneur, E	Bro, R	Viergever, MA	Marzetta, TL	Alon, N
27	Ristow, GH	Ristow, GH	Marzetta, TL	MacKay, DJC	Mesiar, R	Amari, S
28	Fidelman, U	Fidelman, U	Forney, GD	Yager, RR	Overhage, JM	Motwani, R
29	Simon, DR	Simon, DR	Zuckerman, D	Schapire, RE	Freeman, LB	Bates, DW
30	Renegar, J	Renegar, J	Bey, J	Amari, S	Schapire, RE	McDonald, CJ
31	Robinson, DL	Robinson, DL	Warmuth, MK	Alon, N	Forney, GD	Jordan, MI
32	Myers, JS	Myers, JS	Shortliffe, EH	Vardy, A	Shahar, Y	Verdu, S
33	Sampson, G	Sampson, G	Scholkopf, B	Feige, U	Shu, CW	Paxson, V
34	Thomason, A	Thomason, A	Helleseth, T	Richardson, TJ	Kimmel, R	Vardy, A
35	Yngve, VH	Yngve, VH	Lin, YB	Bey, J	Musen, MA	Jain, R
36	Vazirani, U	Vazirani, U	Amari, S	Motwani, R	Yager, RR	Hill, DLG
37	Bernstein, E	Bernstein, E	Sharir, M	Goldreich, O	Smola, AJ	Scholkopf, B
38	Wang, WY	Schwarzer, S	Duin, RPW	Renegar, J	Warmuth, MK	Muller, KR
39	Schwarzer, S	Wachmann, B	Hochwald, BM	Szeliski, R	Amari, S	Ross, JM
40	Wachmann, B	Wang, WY	Kimmel, R	Picard, RW	Williamson, RC	Curtis, HW
41	Russell, CA	Russell, CA	Jordan, MI	Kittler, J	Long, PM	Montrose, CJ
42	Chin, B	Chin, B	Long, PM	Bartlett, PL	Linder, T	Muhlfeld, HP
43	Enger, TA	Enger, TA	Calderbank, AR	Paxson, V	Helleseth, T	OGorman, TJ
44	Hosier, P	Hosier, P	Williamson, DP	Hyvarinen, A	Maass, W	Bartlett, PL
45	Klein, WA	Klein, WA	Shahar, Y	Sharir, M	Campbell, KE	Leinen, P
46	LaFave, LE	LaFave, LE	Freund, Y	Hochwald, BM	Chlamtac, I	Towsley, D
47	Messina, B	Messina, B	Shu, CW	Tanaka, K	Jordan, MI	Kim, J
48	Nicewicz, M	Nicewicz, M	Renegar, J	Black, MJ	Greenes, RA	Willinger, W
49	Orro, JM	Orro, JM	Szegedy, M	Arora, S	Williamson, DP	Chute, CG
50	Scott, TS	Scott, TS	Maass, W	Kim, J	Fang, YG	Breiman, L

Table A.2
Top 50 researchers by both kinds of rankings (part 2).

	allDistCoauthors	*allDistCoauthorsT*	*allCollaborations*	*allCollaborationsT*
1	Ziegler, JF	Srinivasan, GR	Srinivasan, GR	Srinivasan, GR
2	Srinivasan, GR	Murley, PC	Ziegler, JF	Jain, AK
3	Sudan, M	Ziegler, JF	Verdu, S	Murley, PC
4	Freeman, LB	Freeman, LB	Sudan, M	Ziegler, JF
5	Osher, S	Tang, HHK	Murley, PC	Freeman, LB
6	Sapiro, G	Jain, AK	Shamai, S	Tang, HHK

Table A.2 (*Continued*)

	allDistCoauthors	*allDistCoauthorsT*	*allCollaborations*	*allCollaborationsT*
7	Verdu, S	Sudan, M	Freeman, LB	Sudan, M
8	Shamai, S	Renegar, J	Jain, AK	Shamai, S
9	Jain, AK	Calderbank, AR	Tse, DNC	Tse, DNC
10	Tse, DNC	Pentland, A	Osher, S	Renegar, J
11	Kschischang, FR	Shamai, S	Sapiro, G	Calderbank, AR
12	McEliece, RJ	Gupta, A	Tang, HHK	Tarokh, V
13	Tarokh, V	Tse, DNC	Tarokh, V	Kanade, T
14	Arora, S	Freund, RM	Vardy, A	Pentland, A
15	Vardy, A	Sapiro, G	Kschischang, FR	Cimino, JJ
16	Cimino, JJ	Kanade, T	McEliece, RJ	Sapiro, G
17	Zuckerman, D	Alon, N	Leinen, P	Sejnowski, TJ
18	Bates, DW	Sejnowski, TJ	Cimino, JJ	Osher, S
19	Shortliffe, EH	Lee, J	Arora, S	Gupta, A
20	Marzetta, TL	Cimino, JJ	Oja, E	Verdu, S
21	Oja, E	Osher, S	Schapire, RE	Jordan, MI
22	Bartlett, PL	Kikinis, R	Lapidoth, A	Lee, J
23	Bro, R	Leinen, P	Bartlett, PL	Viergever, MA
24	Schapire, RE	Ross, JM	Marzetta, TL	Yu, PS
25	Forney, GD	Curtis, HW	Bro, R	Freund, RM
26	Scholkopf, B	Montrose, CJ	Bey, J	Schapire, RE
27	Murley, PC	Muhlfeld, HP	Forney, GD	Vardy, A
28	Lapidoth, A	OGorman, TJ	Warmuth, MK	Leinen, P
29	Leinen, P	Motwani, R	Zuckerman, D	Amari, S
30	Amari, S	Schapire, RE	Scholkopf, B	Alon, N
31	Overhage, JM	Tarokh, V	Sharir, M	Motwani, R
32	Szegedy, M	Viergever, MA	Helleseth, T	Goldreich, O
33	Shu, CW	Jordan, MI	Shortliffe, EH	Bartlett, PL
34	Williamson, DP	Paxson, V	Duin, RPW	Szeliski, R
35	Warmuth, MK	Kim, J	Amari, S	Kittler, J
36	Duin, RPW	Amari, S	Hochwald, BM	Breiman, L
37	Lin, YB	Vardy, A	Lin, YB	Feige, U
38	Kimmel, R	Bey, J	Kimmel, R	Hochwald, BM
39	Shahar, Y	Lakshman, TV	Jordan, MI	Bey, J
40	Helleseth, T	Feige, U	Calderbank, AR	Sharir, M
41	Jordan, MI	Yu, PS	Mesiar, R	Black, MJ
42	Campbell, KE	Arora, S	Williamson, DP	Jain, R
43	Long, PM	Breiman, L	Shu, CW	Towsley, D
44	Sharir, M	Willinger, W	Long, PM	Lakshman, TV
45	Musen, MA	Vera, JR	Renegar, J	Kim, J
46	Freund, Y	Tanaka, K	Freund, Y	Paxson, V
47	Bey, J	Shor, PW	Shahar, Y	Richardson, TJ
48	Hochwald, BM	Verdu, S	Szegedy, M	Tanaka, K
49	Calderbank, AR	Jain, R	Breiman, L	Kikinis, R
50	Greenes, RA	Hill, DLG	Bates, DW	MacKay, DJC

Table A.3
Top 50 researchers by both kinds of rankings (part 3).

	coauthors	*coauthorsT*	*distCoauthors*	*distCoauthorsT*
1	Srinivasan, GR	Srinivasan, GR	Srinivasan, GR	Srinivasan, GR
2	Ziegler, JF	Murley, PC	Ziegler, JF	Murley, PC
3	Freeman, LB	Ziegler, JF	Freeman, LB	Ziegler, JF
4	Murley, PC	Freeman, LB	Murley, PC	Freeman, LB
5	Tang, HHK	Tang, HHK	Tang, HHK	Tang, HHK
6	Leinen, P	Leinen, P	Leinen, P	Leinen, P
7	Bey, J	Bey, J	Bey, J	Bey, J
8	Juang, JG	Juang, JG	Juang, JG	Juang, JG
9	Juang, HG	Juang, HG	Juang, HG	Juang, HG
10	Wiener, N	Wiener, N	Wiener, N	Wiener, N
11	Korec, I	Curtis, HW	Korec, I	Curtis, HW
12	Cegielski, P	Montrose, CJ	Cegielski, P	Montrose, CJ
13	Curtis, HW	Muhlfeld, HP	Curtis, HW	Muhlfeld, HP
14	Montrose, CJ	OGorman, TJ	Montrose, CJ	OGorman, TJ
15	Muhlfeld, HP	Ross, JM	Muhlfeld, HP	Ross, JM
16	OGorman, TJ	Korec, I	OGorman, TJ	Korec, I
17	Ross, JM	Cegielski, P	Ross, JM	Cegielski, P

Table A.3 (*Continued*)

	coauthors	*coauthorsT*	*distCoauthors*	*distCoauthorsT*
18	Renegar, J	Taber, AH	Renegar, J	Taber, AH
19	Sudan, M	Walsh, JL	Muses, C	Walsh, JL
20	Schapire, RE	Muses, C	Simon, DR	Muses, C
21	Simon, DR	Renegar, J	Litkowski, KC	Renegar, J
22	Muses, C	Litkowski, KC	McTavish, DG	Litkowski, KC
23	Litkowski, KC	Simon, DR	Gazarik, MJ	McTavish, DG
24	McTavish, DG	McTavish, DG	Kamen, EW	Simon, DR
25	Vazirani, U	Gazarik, MJ	Prou, JM	Gazarik, MJ
26	Bernstein, E	Kamen, EW	Wagneur, E	Kamen, EW
27	Taber, AH	Prou, JM	Sudan, M	Prou, JM
28	Walsh, JL	Wagneur, E	Taber, AH	Wagneur, E
29	Ristow, GH	Fidelman, U	Walsh, JL	Fidelman, U
30	Fidelman, U	Ristow, GH	Ristow, GH	Ristow, GH
31	Gazarik, MJ	Vazirani, U	Fidelman, U	Vazirani, U
32	Kamen, EW	Bernstein, E	Vazirani, U	Bernstein, E
33	Prou, JM	Robinson, DL	Bernstein, E	Robinson, DL
34	Wagneur, E	Myers, JS	Schapire, RE	Myers, JS
35	Bennett, CH	Sampson, G	Bennett, CH	Sampson, G
36	Shamai, S	Thomason, A	Robinson, DL	Thomason, A
37	Osher, S	Yngve, VH	Breiman, L	Yngve, VH
38	Breiman, L	Sudan, M	Myers, JS	Sudan, M
39	Jain, AK	Bennett, CH	Sampson, G	Wang, WY
40	Tarokh, V	Wang, WY	Thomason, A	Schwarzer, S
41	Myers, JS	Breiman, L	Yngve, VH	Wachmann, B
42	Sampson, G	Schwarzer, S	Jain, AK	Russell, CA
43	Thomason, A	Wachmann, B	Shamai, S	Chin, B
44	Yngve, VH	Russell, CA	Schwarzer, S	Enger, TA
45	Robinson, DL	Chin, B	Wachmann, B	Hosier, P
46	Calderbank, AR	Enger, TA	Calderbank, AR	Klein, WA
47	Sapiro, G	Hosier, P	Tarokh, V	LaFave, LE
48	Freund, Y	Klein, WA	Freund, RM	Messina, B
49	Schwarzer, S	LaFave, LE	McEliece, RJ	Nicewicz, M
50	Wachmann, B	Messina, B	Behbehani, J	Orro, JM

References

Bar-Ilan, J. (2010). Web of science with the conference proceedings citation indexes: The case of computer science. *Scientometrics*, *83*(3), 809–824.
Bergstrom, C. (2007). Eigenfactor: Measuring the value and prestige of scholarly journals. *College and Research Libraries News*, *68*(5), 314–316.
Bollen, J., Rodriguez, M. A., & Van De Sompel, H. (2006). Journal status. *Scientometrics*, *69*(3), 669–687.
Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. In *Proceedings of the 7th World Wide Web conference* Brisbane, Australia, (pp. 107–117).
Chen, P., Xie, H., Maslov, S., & Redner, S. (2007). Finding scientific gems with Google's PageRank algorithm. *Journal of Informetrics*, *1*(1), 8–15.
Codd Award. http://www.sigmod.org/sigmod-awards/sigmod-awards#innovations.
Ding, Y. (2011). Applying weighted PageRank to author citation networks. *Journal of the American Society for Information Science and Technology*, *62*(2), 236–245.
Ding, Y., Yan, E., Frazho, A., & Caverlee, J. (2009). PageRank for ranking authors in co-citation networks. *Journal of the American Society for Information Science and Technology*, *60*(11), 2229–2243.
Fiala, D. (2011). Mining citation information from CiteSeer data. *Scientometrics*, *86*(3), 553–562.
Fiala, D. Bibliometric analysis of CiteSeer data for countries. *Information Processing and Management*, in press.
Fiala, D., Rousselot, F., & Ježek, K. (2008). PageRank for bibliographic networks. *Scientometrics*, *76*(1), 135–158.
Franceschet, M. (2010). The role of conference publications in CS. *Communications of the ACM*, *53*(12), 129–132.
González-Pereira, B., Guerrero-Bote, V. P., & Moya-Anegón, F. (2010). A new approach to the metric of journals' scientific prestige: The SJR indicator. *Journal of Informetrics*, *4*(3), 379–391.
Kleinberg, J. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM*, *46*(5), 604–632.
Ma, N., Guan, J., & Zhao, Y. (2008). Bringing PageRank to the citation analysis. *Information Processing and Management*, *44*(2), 800–810.
Sidiropoulos, A., & Manolopoulos, Y. (2005). A citation-based system to assist prize awarding. *SIGMOD Record*, *34*(4), 54–60.
Turing Award. http://awards.acm.org/homepage.cfm?srt=all&awd=140.
Wainer, J., Goldenstein, S., & Billa, C. (2011). Invisible work in standard bibliometric evaluation of computer science. *Communications of the ACM*, *54*(5), 141–148.
Walker, D., Xie, H., Yan, K.-K., & Maslov, S. (2007). Ranking scientific publications using a model of network traffic. *Journal of Statistical Mechanics: Theory and Experiment*, *6*
Xing, W., & Ghorbani, A. (2004). Weighted PageRank algorithm. In *Proceedings of the 2nd annual conference on communication networks and services research* Fredericton, Canada, (pp. 305–314).
Yan, E., & Ding, Y. (2010). Weighted citation: An indicator of an article's prestige. *Journal of the American Society for Information Science and Technology*, *61*(8), 1635–1643.
Yan, E., & Ding, Y. (2011). Discovering author impact: A PageRank perspective. *Information Processing and Management*, *47*(1), 125–134.
Yu, P. S., Li, X., & Liu, B. (2004). On the temporal dimension of search. In *Proceedings of the 13th World Wide Web conference* New York, USA, (pp. 1180–1181).