PageRank variants in the evaluation of citation networks

Michal Nykl

Karel Ježek

Dalibor Fiala*

Martin Dostal

University of West Bohemia, Department of Computer Science and Engineering
Univerzitní 8, 30614 Plzeň, Czech Republic
* Corresponding author. Tel.: +420 377 63 24 29.
Email addresses: nyklm@kiv.zcu.cz (M. Nykl), jezek_ka@kiv.zcu.cz (K. Ježek), dalfia@kiv.zcu.cz (D. Fiala), madostal@kiv.zcu.cz (M. Dostal).

Abstract: This paper explores a possible approach to a research evaluation, by calculating the renown of authors of scientific papers. The evaluation is based on the citation analysis and its results should be close to a human viewpoint. The PageRank algorithm and its modifications were used for the evaluation of various types of citation networks. Our main research question was whether better evaluation results were based directly on an author network or on a publication network. Other issues concerned, for example, the determination of weights in the author network and the distribution of publication scores among their authors. The citation networks were extracted from the computer science domain in the ISI Web of Science database. The influence of self-citations was also explored. To find the best network for a research evaluation, the outputs of PageRank were compared with lists of prestigious awards in computer science such as the Turing and Codd award, ISI Highly Cited and ACM Fellows. Our experiments proved that the best ranking of authors was obtained by using a publication citation network from which self-citations were eliminated, and by distributing the same proportional parts of the publications' values to their authors. The ranking can be used as a criterion for the financial support of research teams, for identifying leaders of such teams, etc.

Keywords: PageRank, citation analysis, research evaluation, author ranking, ISI Web of Science.

1. Introduction

The evaluation of universities' prestige usually covers several areas such as research results, education, student satisfaction and others. When evaluating research, publications play an important role. Publications and their citations can best show the top researcher in the selected

Preprint of: Nykl, M., Ježek, K., Fiala, D., & Dostal, M. (2014). PageRank variants in the evaluation of citation networks. *Journal of Informetrics*, 8(3), 683-692.

field of science. This evaluation is usually based on the number of publications indexed in e.g. the ISI Web of Science[1] (hereafter WoS) with regard to the number of citations and Journal Impact Factor (Garfield, 1972). The Impact Factor[2] of journal J in a given year (e.g. 2011) is the number of citations in this year (2011) to all items published in journal J two years before (2010 and 2009) divided by the number of journal J's citable items (i.e. excluding notes, editorials, etc.) published in those two years (2010 and 2009). Note that in the evaluation, the impact factor of citing journals is not taken into account.

Our main method for the evaluation of citation networks is the PageRank algorithm, which uses the impact of citing nodes (articles, authors and so on) for determining the importance of cited nodes. PageRank was introduced by Brin and Page (1998) to rank websites and became part of the Google search engine. From its introduction, PageRank has been examined for convergence, acceleration, rating prediction, etc. For example, Langville and Meyer (2006) is a good starting point for its deeper study.

PageRank has been frequently used for citation analysis. Fiala (2012) worked with the publication citation network and the authorship network to create an author citation network. The determination of edge weights with regard to the publication date and co-authorship is also solved. Other variants of bibliographic network evaluations (comprising e.g. co-citation or co-authorship network) are compared by Yan and Ding (2012). Sidiropoulos and Manolopoulos (2006) used the list of *ACM SIGMOD E. F. Codd Innovation Award* holders to compare the results of human and machine rankings of authors. We used the same approach to determine the quality of author rankings but also used some other human evaluation methods. Yu et al. (2012) explored a network which combines information from citations, reviews, comments, and information on the reputation of social network users who read articles and comment on them. A comparison and combination of PageRank and the journal impact factor are presented by Bollen et al. (2006).

Our main research question was whether better evaluation results were based directly on an author network or on a publication network. We investigate several variants of author or publication citation networks. The influence of self-citations is explored and a further two variants of author ratings are proposed and studied. The author's rating can be obtained either from the weighted author citation networks or as a distribution of publication values among their authors. Other questions, therefore, concern, for example, how to determine the weights

[1] *ISI Web of Science* - http://www.webofknowledge.com
[2] Computation of Journal Impact Factor in the WoS database - http://admin-apps.webofknowledge.com/JCR/help/h_impfact.htm

Preprint of: Nykl, M., Ježek, K., Fiala, D., & Dostal, M. (2014). PageRank variants in the evaluation of citation networks. *Journal of Informetrics*, 8(3), 683-692.

in the author network and how to distribute the publication scores among their authors. The evaluation results are compared with lists of the holders of four prestigious computer science awards. Our main contribution demonstrates that the best ranking of authors is obtained by using a publication citation network from which self-citations are eliminated and by distributing the same proportional parts of the publications' values to their authors.

The following section describes the data from the WoS collection, the used lists of prestigious awards and the construction of citation networks of papers or authors. The *Types of citation networks* section provides information on how to add weights to edges in the author citation network and how to distribute the publication scores among authors. The next section is devoted to our modifications of the PageRank algorithm. The experiments and their results are summarized in the *Experiments* section and discussed in the *Discussion* section. The conclusion and recommendation are presented in the last section.

2. Data used

All of our experiments can be run on an arbitrary bibliographic data collection, but we used the already purchased Thomson Reuters collection employed in our previous studies (Fiala, 2012). This collection consists of all publications classified as *"article"* published in Journal Citation Reports 2009 in the computer science category between 1996 and 2005. This category covers all seven WoS subcategories: Artificial Intelligence, Cybernetics, Hardware & Architecture, Information Systems, Interdisciplinary Applications, Software Engineering and Theory & Methods.

Using this data, we create two citation networks – the publication network and the author network. The networks can consider various types of self-citations (see Figure 1). The first variant, marked *ALL*, takes into account all citations and is, therefore, the most benevolent. The second variant, marked *NOT*, removes citations between publications having at least one common author. For this reason, it is the strictest variant. The last variant is marked *PART*. It is applicable only to author networks and is created from the *ALL* variant by removing all self-loops. Other variants of self-citations are mentioned by Yan, Ding, and Sugimoto (2010), who eventually used self-citations with lower weights.

Our data collection contains 149,347 articles from 386 journals. The average publication was written by 2.5 authors and has 1.3 citations in the *ALL* variant and 1 citation in the *NOT* variant. The average author has 2.4 publications and 6.8 citations in the *ALL* variant, 6.6 citations in the *PART* variant and 5.4 citations in the *NOT* variant.

Insert Figure 1 here.

In the further described networks we use the following notions:

- *nodes* represent publications or authors
- *edges* represent citations
- *dangling nodes* (marked DN) are nodes which lack an outgoing edge
- *uncited* nodes are nodes which lack an incoming edge
- *isolated* are nodes which lack both outgoing and incoming edges

The numbers of nodes, edges, DNs, uncited nodes, and isolated nodes in our networks are shown in Table 1.

Insert Table 1 here.

The comparison of human-made and machine-made author rankings is evaluated with the help of several lists of scientists who are holders of prestigious awards in computer science (mentioned below). From these lists names were removed which, due to their incompleteness, were ambiguous. This approach is similar to that mentioned by Sidiropoulos and Manolopoulos (2006) and Lin et al. (2013). Name disambiguation and unification has not been performed.

We used lists of the following prestigious awards:

- ***ACM A. M. Turing Award***[3] – ACM's most prestigious technical award is given for major contributions of lasting importance to computing. We use 39 well-distinguishable names from the period 1966–2010.
- ***ACM SIGMOD Edgar F. Codd Innovations Award***[4] – This is given for innovative and highly significant contributions of enduring value to the development, understanding, or use of database systems and databases. We use 15 well-distinguishable names from the period 1992–2010.
- ***ACM Fellows***[5] – The ACM Fellows program was established in 1993 to recognize and honor outstanding ACM members for their achievements in computer science and information technology and for their significant contributions to the mission of ACM. We use 576 well-distinguishable names from the period 1994–2011.

[3] *ACM Turing Award* - http://amturing.acm.org
[4] *ACM SIGMOD Codd* Award - http://www.sigmod.org/sigmod-awards/
[5] *ACM Fellows* - http://fellows.acm.org

- ***ISI Highly Cited***[6] – These highly cited researchers were identified by the Thomson Reuters team between 2000 and 2008 based on an analysis of papers covered in WoS from 1981 to 2008. We use 280 well-distinguishable names.

These unified lists contain 805 scientists who were found in WoS.

3. Types of citation networks

This section presents several types of citation networks with regard to the weights of the edges. Let us start with the difference between publication networks and author networks. Whereas publication networks contain information on the time sequences of publications, author networks lack this information. This difference may give an advantage to some authors (e.g. author C in Figure 2). We suppose that the publication network better describes the author's influence and is more useful for the evaluation of scientists than the author network. As shown in our experiments, this assumption was proven.

Insert Figure 2 here.

Other network variants can be obtained by using weights of edges in author networks. Table 2 shows three variants of assigning weights to the author networks from Figure 1. In the first case, marked *N*, weights expressing the numbers of authors' citations in the publication network are assigned to the graph edges. For example, if author A has two publications both citing a publication by author B, then in the author network there exists an edge from A to B with weight 2. In the second case, marked *1/N*, the publications' values are uniformly distributed to outgoing citations. This means that if a publication by author A1 cites a publication by two authors A4 and A5, then in the author network there exist edges from A1 to A4 and A1 to A5 both with weight 1/2. In the last case, marked *1*, weight 1 is assigned to all edges.

Insert Table 2 here.

A rating of authors can also be obtained from the values of their publications. In this case, we evaluated the PageRank scores for all publications. Consequently, we distributed the publication values to their authors. The distribution can be done by assigning sums of publication values to all their authors, either regardless of the number of co-authors (marked *SUM*), or as a proportional part of the publication value depending on the number of authors

[6] *ISI Highly Cited* - http://www.isihighlycited.com

(marked *DIV*). For example, if the author has three co-authors in his first publication P1 and five co-authors in his second publication P2, then in variant *DIV* he receives 1/4 of the P1 score plus 1/6 of the P2 score.

Other variants of network evaluations, not explored in this paper, provide various ways of including the author's position in the list of co-authors below the paper's title. For example, Zhao (2005) investigated the influence of authors' positions in the list of co-authors, but he used only the author network, where only the first authors, first N authors, or all authors of a publication were considered. Also, some score distribution variants, like those shown by Assimakis and Adam (2010), can determine the distribution of publication non-proportional parts with regard to the author's position in the list of co-authors. With respect to the *DIV* variant results, we aim to test this approach in our next experiments.

4. Evaluation of the networks and experiments

The main algorithm that we use for the evaluations of citation networks is the PageRank algorithm (Brin & Page, 1998; Langville & Meyer, 2006). PageRank was designed for the ranking of websites. In this area it was the first recursive algorithm which used in the computation of a website score the scores of the referring (citing) websites. The second algorithm, developed at about the same time, was HITS by Kleinberg (1999). However, unlike HITS, PageRank was widely used and became part of the Google search engine.

We use Formula (1) where $PR_x(A)$ is the PageRank score of node A in iteration x, and d is the damping factor. The damping factor determines to what extent the final score is computed using the citing nodes' scores (in the PageRank definition this is described as how many times a web user follows hyperlinks to other websites when browsing) and to what extent a random teleport is used (in the definition the random teleport is equal to a situation, in which the user types the address of a random website in the web browser address bar and does not follow any hyperlinks). By observing web users, the PageRank authors identified that users used the random teleport once in six steps on average. Therefore, they recommended setting the damping factor to 0.85. The first part of Formula (1), marked $F(A)$, represents the random teleport and assigns the same probability of a random jump to all nodes in the network (see Formula (2)), where $|V|$ is the cardinality of the nodes set V. The random teleport is important for convergence when the PageRank values of all nodes in the set are computed iteratively. The second part of the formula, marked $L_x(A)$, represents following the hyperlinks (i.e. graph edges) and uses the values of the citing nodes to determine the values of

the cited nodes. Our variant contains weights of edges and resolves the problem of dangling nodes (nodes which lack an outgoing edge), connecting them to all nodes in the network (see Formula (3)), where U is the set of nodes which have an edge incoming to node A, $w_{uto A}$ is the weight of the edge from node u to node A, $w_{u\text{out}}$ is the sum of weights of all outgoing edges from node u, and D is the set of all dangling nodes in the network.

$$PR_{x+1}(A) = (1-d) \cdot F(A) + d \cdot L_x(A) \quad (1)$$

$$F(A) = 1 \,/\, |V| \quad (2)$$

$$L_x(A) = \left(\sum_{u \in U} \frac{PR_x(u) \cdot w_{u\text{toA}}}{w_{u\text{out}}} + \frac{1}{|V|} \sum_{s \in D} PR_x(s)\right) \quad (3)$$

In general, the random teleport gives the same teleportation probability to all nodes. A non-uniform teleportation is called *personalization* (Brin & Page, 1998). We experimented with two personalizations. In the first case, we replaced the uniform distribution of the random jumping to all nodes with a distribution favoring authors with more publications (see Formula (4)), where A_{Pub} is the number of author A's publications and V represents the set of all authors. High publication numbers may identify authors who are popular and, as was mentioned by Ding (2011), "being popular" is necessary for "being prestigious". Therefore, we assume that using publication numbers as personalization can provide better results of author evaluations. We can also say that authors are rewarded for their productivity.

$$F(A) = A_{Pub} \Big/ \sum_{a \in V} a_{Pub} \quad (4)$$

As was shown by Glänzel (2001), publications written by more authors coming from more countries are cited more frequently. Our question is whether more authors of a publication can contribute to a higher publication quality and, consequently, to a better evaluation of authors. Therefore, in the second case, we use personalization favoring publications with a higher number of authors (see Formula (5)), where PAut is the number of publication P's authors and V represents the set of all publications. We experimented with Formula (2), (4), and (5) as the first part of Formula (1).

$$F(P) = P_{Aut} \Big/ \sum_{p \in V} p_{Aut} \quad (5)$$

To compare PageRank results with a less ingenious method, we used in-degree ranking calculating node values according to Formula (6), where *InD(A)* is the in-degree score of node A, U is the set of nodes which have an edge incoming to node A, and $w_{uto A}$ is the

weight of the edge leading from node *u* to node *A*. Other ways of automatic author ranking can be found in Sidiropoulos and Manolopoulos (2006).

$$InD(A) = \sum_{u \in U} w_{u\text{toA}} \tag{6}$$

The aim of our experiments was to find a way of automatic scientist ranking which produces a ranking closest to the established rankings used by humans. Therefore, we use all the above described variants of citation network evaluations and compared the obtained resulting lists of authors with the previously mentioned lists of awarded scientists. The results obtained are contained in Table 3, where the rows combine the use of:

- author/publication network
- methods of evaluation (*in-degree* counting or *PageRank* algorithm)
- exclusion/inclusion of self-citations (marked as *NOT* or *ALL* in the case of publication networks and *NOT*, *PART* or *ALL* in the author network evaluation)
- variants of assigned weights in the author networks (*1* or *1/N* or *N*) and distribution of publication values to their authors (*DIV* or *SUM*)
- variant of the first part of the PageRank formula (1) – Formula (2) or Formula (4) in the case of author networks and Formula (2) or Formula (5) in the case of publication networks.

In this table, the value contained in each cell of the column "sum" represents the sum of positions of awarded authors in the respective ranking variant. The authors with the highest ranks occupy the first positions in the ranking; therefore, the lowest sum of positions is the best. If two authors have the same scores and are, for example, on the second and third position in the ranking, then their position is determined as 2.5. The columns "r" show the ranks of all 39 evaluation variants for five different author sets when they are sorted in ascending order by the respective sums. Again, the lower the value of "r", the better the specific variant. For a better illustration of results, the best seven ranking variants in Table 3 are highlighted with a gray background and the worst twelve variants are highlighted in bold text. The author ranking based on a pure citation count (we may call it a baseline ranking) is the in-degree variant working with network *ALL* and weights *N*. In Table 3, this variant is marked with an asterisk (see the row "*N**") and highlighted with a black background.

5. Discussion

Let us start by considering whether a better resulting order of authors can be obtained by evaluating author citation network or publication citation network. A comparison of the results is presented in Table 3. The average values of columns "r" for variants marked *Publication* give better results than the average values in columns "r" for variants marked *Author*. In the column *Unification* (unifying all the four used sets of awarded scientists), the average value of "r" is 12.3 for variants *Publication* and 23.4 for variants *Author*.

In the evaluation of the publication citation networks, the PageRank algorithm outperforms the simpler in-degree evaluation. This result was expected because PageRank is able to take into account the quality of citing publications. In the column *Unification* the average rank of the PageRank variants is 5.8 and of the in-degree variants 25.3.

Insert Table 3 here.

From Table 3 we can see the better results for publication networks without self-citations (marked *NOT*). In the column *Unification*, an average rank of 4.3 is given, while those networks containing self-citations (marked *ALL*) give an average rank of 7.3. Further, we can see that it is better to distribute to the authors the proportional part than the full value of their publications (from the publication citation network without self-citations). Compare the variants marked *DIV* and *SUM*. For these variants the average ranks in the column *Unification* are 1.5 and 7. The only exception is the evaluation based on the *Codd* prize. The reason could be: 1) this prize is awarded only in the area of database systems and our collection covers the complete computer science area; 2) the number of awarded authors is too small (only 15 persons).

Finally, a better order of authors is provided by the PageRank variant which uses Formula (5) rather than Formula (2). Different results are yielded by the evaluation with the *Turing* and *Codd* prizes. However, these results are not substantially worse and are justified by the small number of awarded authors.

Tables 4 and 5 show the differences between the most interesting variants of the evaluation. The variations contain the variants applying the PageRank algorithm to the publication citation network. The pure author citation rank and the best PageRank variant using the author network are added too. Table 4 contains the Spearman rank correlation coefficients. The lower correlation between the results of the publication network and the author network and the pure author citation count is also shown. Table 5 shows how many authors are included by different pairs of evaluation variants on the top 100 positions. We also

tested authors' top 1000 positions, but the results were similar to those shown in Table 5. The numbers presented indicate that the most similar rankings are produced by Formula (2) and (5) and that the pure author citation rank is the most distant from all other variants.

Insert Table 4 here.

Insert Table 5 here.

The top 20 positions from the author rankings obtained by the five best variants are presented in Table 6. The authors who are in the first three positions in at least one ranking are highlighted by a gray background. Those who are not among the top 20 researchers in the relevant column but are among the top 20 in any of the other columns are mentioned in the bottom part of the table.

As we can see from Table 6, the top names do not vary substantially, e.g., the name *Jain, AK* permanently occupies one of the top three places. The reason could be that it represents more persons having the same surname. The other interesting name is *Setiono, R*. It is in first place in the variants using self-citations of authors (variants *ALL*) but in substantially worse positions in the other three variants (variants *NOT*). The most probable explanation is too frequent usage of self-citations by the author and his co-authors. The top places of *Setiono, R* in variant *DIV* compared with variant *SUM* indicates he published his articles with only a small number of co-authors.

Insert Table 6 here.

6. Concluding remarks

In this paper we introduced several variants of citation networks and their usage in the evaluation of authors' prestige. In our created orders of authors, those authors who are holders of prestigious awards granted by scientific societies were found. The goal was to determine those evaluation variants providing the closest ranking to the human opinion. Our main research question was whether better results are provided by an evaluation based directly on an author network or an evaluation based on a publication network. The other problems under study were how to determine the weights in the author network and how to distribute publication scores among their authors.

As the best variant we recognized the one applying the PageRank algorithm to the publication citation network without self-citations and distributing the PageRank values of

publications proportionally to their authors. Briefly put, this is the variant *Publication/PageRank/NOT/DIV/(5)* from Table 3. The results by Formula (5) are of nearly the same quality as those by Formula (2). The pure citation count gives considerably worse results. We should underline that the results obtained are based on the data from the ISI Web of Science database (almost 150,000 computer science journal articles from 1996-2005) and that we also applied the same algorithms to data from other sources (DBLP and CiteSeer), but the results were not so convincing in order to be included in this paper. Our explanation of the results not presented here is the lower quality of the citation networks based on these data sources (only 0.21 citations per publication in DBLP and many indexing errors in CiteSeer (Fiala, 2011).

In the future we would like to apply the PageRank algorithm to journal citation networks to compare our results with author rankings taking into account Journal Impact Factors of journals in which authors published. These values could consequently be integrated into our formulas to obtain author orderings. The distribution of publications' values to authors depending on the authors' order in the paper byline is also worthy of investigation. We consider the evaluation of prestige of workplaces and institutions as another interesting challenge.

Acknowledgments

This work was supported by the UWB grant SGS-2013-029 Advanced Computer and Information Systems and by the European Regional Development Fund (ERDF), project "NTIS - New Technologies for Information Society", European Center of Excellence, CZ.1.05/1.1.00/02.0090. For Dalibor Fiala and Martin Dostal, this work was supported in part by the Ministry of Education of the Czech Republic under grant MSMT MOBILITY 7AMB14SK090

References

Assimakis, N., & Adam, M. (2010). A new author's productivity index: p-index. *Scientometrics*, *85*(2), 415–427.

Bollen, J., Rodriquez, 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. *Computer Networks and ISDN Systems*, *30*(1-7), 107–117.

Ding, Y. (2011). Applying weighted PageRank to author citation networks. *Journal of the American Society for Information Science and Technology*, *62*(2), 236–245.

Fiala, D. (2011). Mining citation information from CiteSeer data. *Scientometrics*, *86*(3), 1–12.

Fiala, D. (2012). Time-aware PageRank for bibliographic networks. *Journal of Informetrics*, *6*(3), 370–388.

Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science*, *178*(60), 471–479.

Glänzel, W. (2001). National characteristics in international scientific co-authorship relations. *Scientometrics*, *51*(1), 69-115.

Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM*, *46*(5), 604-632.

Langville, A. N., & Meyer, C. D. (2006). *Google's PageRank and beyond the science of search engine rankings*. Princeton, NJ, USA: Princeton University Press.

Lin, L., Xu, Z., Ding, Y., & Liu, X. (2013). Finding topic-level experts in scholarly networks. *Scientometrics*, *97*(3), 797–819.

Sidiropoulos, A., & Manolopoulos, Y. (2006). Generalized comparison of graph-based ranking algorithms for publications and authors. *Journal of Systems and Software*, *79*(12), 1679–1700.

Yan, E., & Ding, Y. (2012). Scholarly network similarities: How bibliographic coupling networks, citation networks, cocitation networks, topical networks, coauthorship networks, and coword networks relate to each other. *Journal of the American Society for Information Science and Technology*, *63*(7), 1313–1326.

Yan, E., Ding, Y., & Sugimoto, C. R. (2010). P-Rank: An indicator measuring prestige in heterogeneous scholarly networks. *Journal of the American Society for Information Science and Technology*, *62*(3), 467–477.

Yu, K., Chen, X., & Chen, J. (2012). A multidimensional PageRank algorithm of literatures. *Journal of Theoretical and Applied Information Technology*, *44*(2), 308–315.

Zhao, D. (2005). Going beyond counting first authors in author co-citation analysis. *Proceedings of the American Society for Information Science and Technology*, 42(1).

Preprint of: Nykl, M., Ježek, K., Fiala, D., & Dostal, M. (2014). PageRank variants in the evaluation of citation networks. *Journal of Informetrics*, 8(3), 683-692.

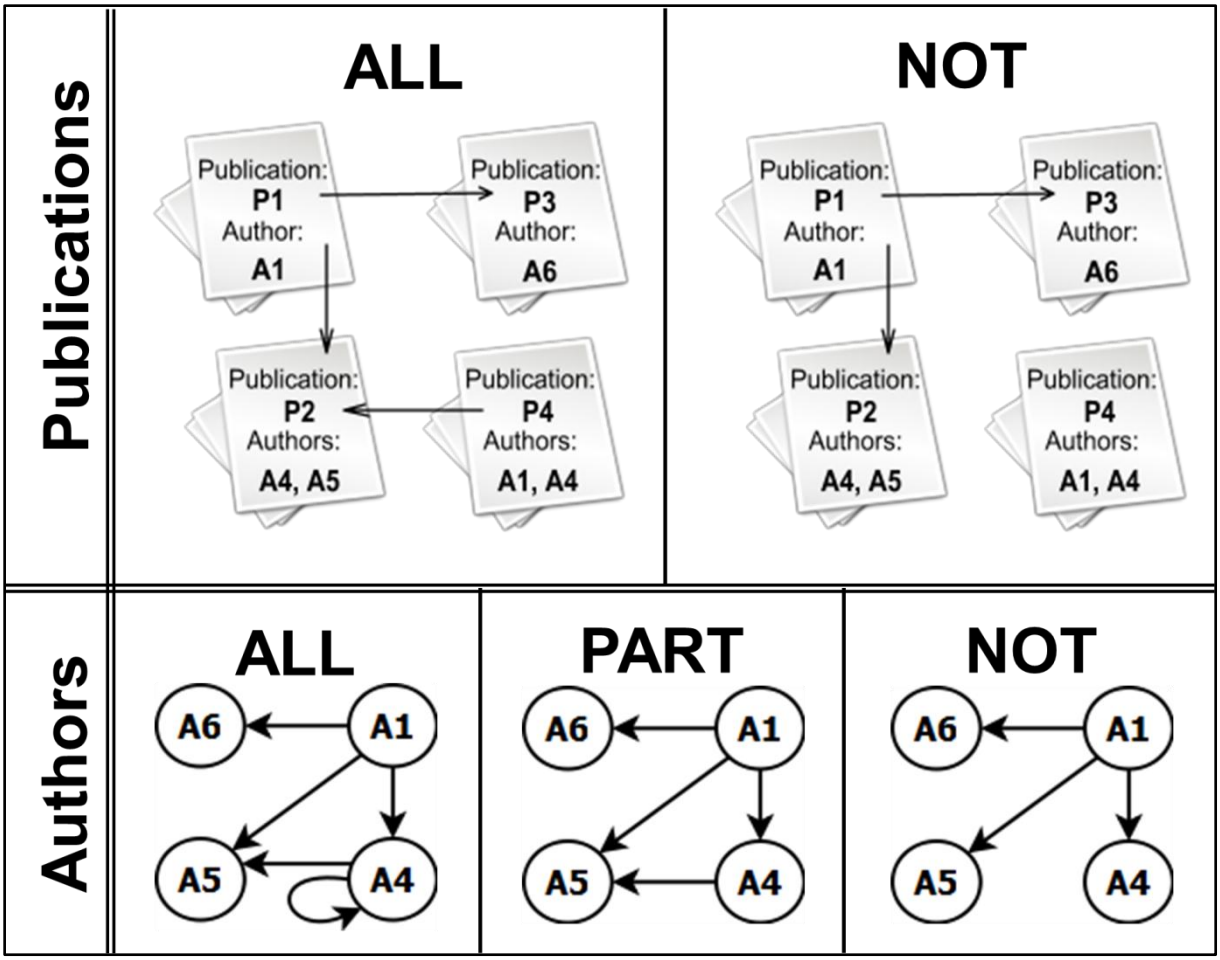

Fig. 1 Types of self-citations variants used

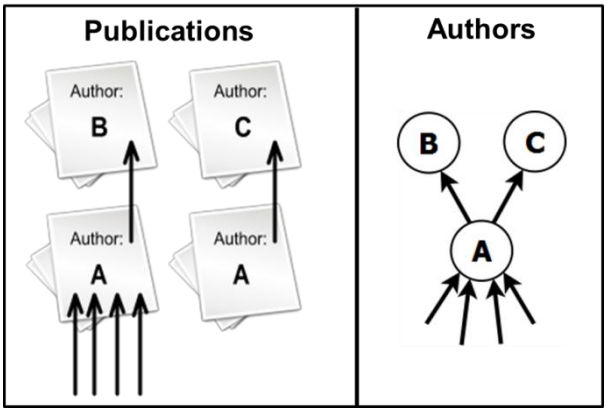

Fig. 2 Difference between the citation networks of publications and of authors

Table 1 Numbers of elements in the citation networks created from WoS

Types of networks	Nodes	Self-cit.	Edges	DN	Uncited	Isolated
Publication	149 347	ALL	191 447	79 571	90 901	49 774
		NOT	145 372	92 694	103 312	64 517
Author	157 440	ALL	1 062 886	71 354	83 146	48 333
		PART	1 039 339	71 843	83 662	48 482
		NOT	852 356	82 170	94 094	56 907

Preprint of: Nykl, M., Ježek, K., Fiala, D., & Dostal, M. (2014). PageRank variants in the evaluation of citation networks. *Journal of Informetrics*, 8(3), 683-692.

Table 2 Variants of assigning weights to the author networks from Figure 1

Edge	ALL			PART			NOT		
	N	**1/N**	**1**	**N**	**1/N**	**1**	**N**	**1/N**	**1**
(A1, A6)	1	1	1	1	1	1	1	1	1
(A1, A5)	2	1	1	2	1	1	1	0.5	1
(A1, A4)	2	1	1	2	1	1	1	0.5	1
(A4, A5)	1	0.5	1	1	0.5	1	---	---	---
(A4, A4)	1	0.5	1	---	---	---	---	---	---

Table 3 Comparison of the resulting author rankings with the lists of prestigious award winners.

(The values in columns "sum" contain the sum of positions of awarded authors in the respective ranking. (The lower the better.) The columns "r" show the rank of each evaluation variant in the list of all 39 variants from the lowest sum of ranks (1) to the highest (39). The row "*N**" represents the ranking by pure citation counts of authors ("baseline" ranking). This row is highlighted in a black background. The best seven evaluation variants for each list of awarded authors are highlighted by a gray background and the worst 12 variants are highlighted in bold text.)

network	method	self-cit.	weight	formula	ACM Turing *(39)*		ACM Codd *(15)*		ACM Fellows *(576)*		ISI Highly Cited *(280)*		Unification *(805)*	
					sum	**r**	**sum**	**r**	**sum**	**r**	**sum**	**r**	**sum**	**r**
Author	In-degree	NOT	1		1,60E+06	16	**5,94E+05**	**34**	1,76E+07	25	**7,30E+06**	**34**	**2,54E+07**	**30**
			1/N		1,51E+06	10	**5,79E+05**	**31**	1,74E+07	22	**7,16E+06**	**31**	2,49E+07	23
			N		1,61E+06	19	**6,02E+05**	**35**	1,76E+07	26	**7,30E+06**	**33**	**2,54E+07**	**31**
		PART	1		**1,73E+06**	**31**	5,17E+05	16	**1,80E+07**	**35**	**7,44E+06**	**36**	**2,59E+07**	**36**
			1/N		1,64E+06	24	5,43E+05	21	**1,78E+07**	**29**	**7,28E+06**	**32**	**2,55E+07**	**32**
			N		**1,76E+06**	**34**	5,36E+05	18	**1,82E+07**	**38**	**7,50E+06**	**38**	**2,62E+07**	**38**
		ALL	1		**1,74E+06**	**33**	5,25E+05	17	**1,81E+07**	**36**	**7,46E+06**	**37**	**2,60E+07**	**37**
			1/N		1,66E+06	27	5,51E+05	24	**1,79E+07**	**32**	**7,32E+06**	**35**	**2,56E+07**	**35**
			N*		**1,77E+06**	**35**	5,45E+05	23	**1,83E+07**	**39**	**7,52E+06**	**39**	**2,62E+07**	**39**
	PageRank	NOT	1	(2)	1,50E+06	7	**6,23E+05**	**38**	1,72E+07	19	6,79E+06	19	2,45E+07	19
				(4)	1,57E+06	11	4,98E+05	12	1,56E+07	7	6,60E+06	10	2,27E+07	8
			1/N	(2)	1,45E+06	5	**6,06E+05**	**37**	1,71E+07	18	6,79E+06	18	2,44E+07	18
				(4)	1,50E+06	8	4,81E+05	11	1,54E+07	6	6,58E+06	8	2,25E+07	6
			N	(2)	1,50E+06	6	**6,28E+05**	**39**	1,72E+07	20	6,78E+06	17	2,45E+07	20
				(4)	1,57E+06	12	5,02E+05	13	1,56E+07	8	6,58E+06	9	2,27E+07	7
		PART	1	(2)	1,61E+06	18	5,45E+05	22	**1,78E+07**	**28**	6,87E+06	21	2,51E+07	26
				(4)	1,64E+06	25	3,91E+05	4	1,62E+07	14	6,69E+06	13	2,34E+07	14
			1/N	(2)	1,57E+06	13	**5,76E+05**	**30**	1,76E+07	27	6,89E+06	22	2,50E+07	24
				(4)	1,59E+06	14	4,48E+05	8	1,60E+07	12	6,69E+06	12	2,32E+07	11
			N	(2)	1,63E+06	23	**5,58E+05**	**28**	**1,79E+07**	**33**	6,92E+06	23	**2,53E+07**	**28**
				(4)	**1,67E+06**	**29**	4,01E+05	6	1,63E+07	15	6,74E+06	14	2,35E+07	15
		ALL	1	(2)	1,63E+06	21	5,56E+05	26	**1,79E+07**	**34**	6,95E+06	25	**2,53E+07**	**29**
				(4)	**1,66E+06**	**28**	4,00E+05	5	1,64E+07	16	6,75E+06	16	2,36E+07	16
			1/N	(2)	1,60E+06	15	**5,88E+05**	**33**	**1,78E+07**	**30**	6,98E+06	26	2,52E+07	27
				(4)	1,61E+06	20	4,54E+05	9	1,62E+07	13	6,74E+06	15	2,34E+07	13
			N	(2)	1,65E+06	26	**5,70E+05**	**29**	**1,81E+07**	**37**	**7,00E+06**	**28**	**2,55E+07**	**34**
				(4)	**1,68E+06**	**30**	4,10E+05	7	1,64E+07	17	6,80E+06	20	2,37E+07	17
Publication	In-degree	NOT	DIV		1,51E+06	9	**5,85E+05**	**32**	1,73E+07	21	6,94E+06	24	2,46E+07	21
			SUM		1,60E+06	17	**6,05E+05**	**36**	1,74E+07	23	6,98E+06	27	2,49E+07	22
		ALL	DIV		1,63E+06	22	5,58E+05	27	1,76E+07	24	**7,06E+06**	**29**	2,51E+07	25
			SUM		**1,73E+06**	**32**	5,54E+05	25	**1,79E+07**	**31**	**7,13E+06**	**30**	**2,55E+07**	**33**
	PageRank	NOT	DIV	(2)	1,24E+06	1	5,12E+05	15	1,47E+07	2	6,38E+06	2	2,14E+07	2
				(5)	1,28E+06	3	5,06E+05	14	1,46E+07	1	6,35E+06	1	2,14E+07	1
			SUM	(2)	**1,83E+06**	**36**	4,56E+05	10	1,52E+07	5	6,42E+06	3	2,24E+07	5
				(5)	**2,04E+06**	**38**	2,94E+05	1	1,56E+07	9	6,47E+06	6	2,27E+07	9
		ALL	DIV	(2)	1,27E+06	2	5,41E+05	20	1,50E+07	4	6,44E+06	5	2,18E+07	4
				(5)	1,33E+06	4	5,37E+05	19	1,49E+07	3	6,44E+06	4	2,18E+07	3
			SUM	(2)	**1,88E+06**	**37**	3,86E+05	3	1,56E+07	10	6,56E+06	7	2,28E+07	10
				(5)	**2,09E+06**	**39**	3,19E+05	2	1,60E+07	11	6,60E+06	11	2,33E+07	12

Preprint of: Nykl, M., Ježek, K., Fiala, D., & Dostal, M. (2014). PageRank variants in the evaluation of citation networks. *Journal of Informetrics*, 8(3), 683-692.

Table 4 Spearman rank correlation coefficients for some evaluation variants.
(The twelve best correlation coefficients of different pairs of evaluation variants are highlighted. The row marked with "*N**" represents the author ranking based on the pure citation count.)

network	method	self-citations	weights	formula	Author		Publication							
					In-D.	PR	PageRank							
					ALL	NOT	NOT				ALL			
					N*	1/N	DIV		SUM		DIV		SUM	
					-	(4)	(2)	(5)	(2)	(5)	(2)	(5)	(2)	(5)
Author	In-D.	ALL	N*	-	1	0,608	0,561	0,562	0,661	0,629	0,555	0,559	0,644	0,615
	PR	NOT	1/N	(4)	0,608	1	0,841	0,860	0,876	0,870	0,829	0,847	0,857	0,844
Publication	PageRank	NOT	DIV	(2)	0,561	0,841	1	**0,997**	0,915	0,877	**0,985**	**0,977**	0,894	0,849
				(5)	0,562	0,860	**0,997**	1	0,932	0,903	**0,985**	**0,984**	0,914	0,877
			SUM	(2)	0,661	0,876	0,915	0,932	1	**0,986**	0,905	0,921	**0,979**	**0,959**
				(5)	0,629	0,870	0,877	0,903	**0,986**	1	0,873	0,898	**0,974**	**0,980**
		ALL	DIV	(2)	0,555	0,829	**0,985**	**0,985**	0,905	0,873	1	**0,996**	0,918	0,876
				(5)	0,559	0,847	**0,977**	**0,984**	0,921	0,898	**0,996**	1	0,938	0,906
			SUM	(2)	0,644	0,857	0,894	0,914	**0,979**	**0,974**	0,918	0,938	1	**0,987**
				(5)	0,615	0,844	0,849	0,877	**0,959**	**0,980**	0,876	0,906	**0,987**	1

Table 5 Numbers of authors included by both evaluation variants of a pair of rankings on the first 100 positions in the ranking.
(The 12 best results are highlighted. The row marked with "*N**" represents the author ranking based on the pure citation count.)

networks	methods	self-citations	weights	formula	Author		Publication							
					In-D.	PR	PageRank							
					ALL	NOT	NOT				ALL			
					N*	1/N	DIV		SUM		DIV		SUM	
					-	(4)	(2)	(5)	(2)	(5)	(2)	(5)	(2)	(5)
Author	In-D.	ALL	N*	-	100	41	47	50	56	55	47	50	57	58
	PR	NOT	1/N	(4)	41	100	51	52	40	39	40	41	37	36
Publication	PageRank	NOT	DIV	(2)	47	51	100	**96**	69	67	**79**	**79**	66	62
				(5)	50	52	**96**	100	72	70	**81**	**81**	70	66
			SUM	(2)	56	40	69	72	100	**96**	65	67	**86**	**85**
				(5)	55	39	67	70	**96**	100	63	65	**84**	**84**
		ALL	DIV	(2)	47	40	**79**	**81**	65	63	100	**97**	69	65
				(5)	50	41	**79**	**81**	67	65	**97**	100	72	68
			SUM	(2)	57	37	66	70	**86**	**84**	69	72	100	**96**
				(5)	58	36	62	66	**85**	**84**	65	68	**96**	100

Table 6 The top 20 positions from the author rankings obtained by the five best variants.

(The authors who are in the top three positions in at least one ranking are highlighted by a gray background. Those who are not among the top 20 researchers in the relevant column but are among the top 20 in any of the other columns are mentioned in the bottom part of the table.)

Position	Publication / PageRank				
	DIV				SUM
	NOT		ALL		NOT
	(2)	(5)	(2)	(5)	(2)
1	Simon, DR	Simon, DR	Setiono, R	Setiono, R	Jain, AK
2	Breiman, L	Breiman, L	Jain, AK	Jain, AK	Vazirani, U
3	Jain, AK	Jain, AK	Yager, RR	Breiman, L	Bernstein, E
4	Moltenbrey, K	Yager, RR	Breiman, L	Yager, RR	Simon, DR
5	Yager, RR	Moltenbrey, K	Simon, DR	Simon, DR	Breiman, L
6	Robertson, B	Vazirani, U	Moltenbrey, K	Vazirani, U	Tanaka, K
7	Vazirani, U	Bernstein, E	Vazirani, U	Moltenbrey, K	Yager, RR
8	Bernstein, E	Zadeh, LA	Bernstein, E	Bernstein, E	Kim, J
9	Zadeh, LA	Robertson, B	Pedrycz, W	Pedrycz, W	Lee, J
10	Pedrycz, W	Pedrycz, W	Robertson, B	Zadeh, LA	Chang, CC
11	Amari, S	Hyvarinen, A	Zadeh, LA	Robertson, B	Lee, S
12	Hyvarinen, A	Amari, S	Amari, S	Amari, S	Pedrycz, W
13	Chang, CC	Oja, E	Wang, J	Wang, J	Wang, J
14	Oja, E	Chang, CC	Hyvarinen, A	Hyvarinen, A	Osher, S
15	Wang, J	Tanaka, K	Oja, E	Oja, E	Kim, JH
16	Tanaka, K	Wang, J	Chang, CC	Lee, J	Wang, Y
17	Lee, J	Burges, CJC	Lee, J	Chang, CC	Moltenbrey, K
18	Burges, CJC	Lee, J	Tanaka, K	Tanaka, K	Bennett, CH
19	Lee, S	Lee, S	Egghe, L	Lee, S	Oja, E
20	Zhang, J	Kim, J	Dannenberg, RB	Picard, RW	Wang, HO
	(21) Kim, J	(21) Zhang, J	(22) Lee, S	(22) Dannen..	(24) Zhang, J
	(26) Kim, JH	(25) Kim, JH	(23) Picard, RW	(26) Kim, JH	(28) Amari, S
	(32) Picard, ..	(31) Wang, Y	(27) Zhang, J	(27) Egghe, L	(34) Picard, ..
	(33) Wang, Y	(33) Picard, ..	(28) Kim, JH	(28) Zhang, J	(41) Hyvarin..
	(41) Egghe, L	(52) Egghe, L	(31) Kim, J	(31) Kim, J	(46) Robert..
	(56) Wang, ..	(53) Wang, ..	(35) Burges, CJC	(32) Burges, ..	(57) Zadeh, ..
	(59) Osher, S	(56) Osher, S	(36) Wang, Y	(36) Wang, Y	(128) Burge..
	(68) Setiono, ..	(75) Setiono, ..	(52) Osher, S	(41) Osher, S	(172) Setion..
	(85) Bennett..	(92) Bennett..	(65) Wang, HO	(64) Wang, ..	(204) Egghe, ..
	(1401) Dann..	(1585) Dann..	(100) Bennett, ..	(106) Bennet..	(3948) Dann..