Noname manuscript No.
(will be inserted by the editor)

Review of the indirect citations paradigm: Theory and practice of the assessment of papers, authors and journals

Eleni Fragkiadaki · Georgios Evangelidis

Abstract The family of indicators presented in this paper includes indices created by taking into account not only the direct but also the indirect impact of citations and references. Three types of citation graphs are presented, namely, the *Paper-Citation graph,* the *Author-Citation graph* and the *Journal-Citation graph*, along with different methods for constructing them. In addition, the concept of generations of citations is examined in detail, again by presenting various methods for defining them found in the literature. Finally, a number of indirect indicators for papers, authors and journals are discussed, which among others, include *PageRank, CiteRank, indirect H-index* and the *EigenFactor score.*

1 Introduction

In an effort to assess the scholarly impact of a research unit, such as paper, author or journal, scientists have defined a number of bibliometric indicators. The indicators utilize knowledge present in the available pool of papers, like the list of references, the list of authors and the year of publication, in order to provide a valid and meaningful way of assessing the unit in question.

We distinguish two families of indicators, namely, the direct and the indirect indicators. Direct indicators utilize the information about the direct citations received by the unit under scrutiny. Such indicators are the *total number of papers* (Hirsch 2005, 2007; Costas and Bordons 2008; Wu 2010), the *total number of citations* (van Eck and Waltman 2008; Wu 2010), *the mean number of citations per paper* (Hirsch 2005, 2007; Costas and Bordons 2008; Wu 2010), *the percentage of highly cited papers* (Costas and Bordons 2008; Waltman and van Eck 2012), the *h-index (Hirsch 2005, 2007) family* of indicators like the $h_I - index$ (Batista et al 2006), the $A - index$ (Jin, Liang, Rousseau, and Egghe 2007), the $R - index$ (Jin, Liang, Rousseau, and Egghe 2007), the $h^2 - index$ (Kosmulski

Department of Applied Informatics, School of Information Sciences, University of Macedonia, 156 Egnatia St, GR-54006, Thessaloniki, Greece
E-mail: eleni.fra@gmail.com · gevan@uom.edu.gr

2006), the $ch-index$ (Kosmulski 2006), the $h_a-index$ (van Eck and Waltman 2008), the $w-index$ (Wu 2010), the $g-index$ (Egghe 2006), the $f-index$ (Tol 2008), the $q^2-index$ (Cabrerizo, Alonso, Herrera-Viedma, and Herrera 2010), the $h^n-index$ (Sidiropoulos et al 2007) and many more, the $h_T-index$ (Anderson, Hankin, and Killworth 2008) and the $IQP-index$ (Antonakis and Lalive 2008).

Indirect indicators consider not only the direct but also the indirect impact of a unit's references and citations. These indicators are the subject of this paper, which aims at providing an overview of the concepts of indirect influence and generations of citations and references. The papers included in our study are the ones that deal with indirect indicators for some or all the scientific units we examine. An initial subset of the papers included were already known to us, while for the discovery of new research papers, we used the reference lists from the initial subset, as well as keyword searches in standard bibliographic databases to identify similar papers. Of course, we can not claim we present a complete overview of the literature for the particular subject, although we do feel that the papers included do constitute a valid subset of the available literature.
The different types of research units are presented in Section 2. As discussed later in this paper, some of the indicators are defined for a particular unit, whereas others may be applied to all units by appropriately modifying the input data. In all cases, the data originates from the *Paper-Citation graph,* a detailed description of which can be found in Section 3. This graph can be transformed to define the *Derived graphs* used in the assessment of authors or journals, and to the best of our knowledge this is the first time such a detailed description of all of the different types of *Derived graphs* has been attempted. The concept of generations of citations and references is examined in Section 4 along with various methods for defining them. The indirect indicators are listed and briefly described in Section 5 and are classified in Section 6. Finally, a summary of all topics discussed in this paper is provided in Section 7.

2 Research Units

We consider a publication to be the elementary entity of scientific assessment. Knowledge is disseminated via publications either from individual researchers or from research groups. The term publication is a general term that can be used to describe any type of peer reviewed document whose purpose is to treat a specific scientific area, and that can be used and referenced by other researchers. Therefore, we consider master theses, PhD theses, technical reports, journal and conference papers, review letters, short communications, books and book chapters to be examples of valid publications. Throughout this paper, we use the term paper to describe any type of the above publications.

Another entity or research unit is the author. Again, we use the term author to describe individual researchers, scientists, professors, authors, analysts that participate in the creation of a paper and appear in the paper's author list.

Finally, the term journal is used to describe any well defined collection of papers, like, for example, printed journals, online journals, open-access journals, archives, repositories, conference proceedings, etc.

For each of these units different types of indicators have been defined that take into account different characteristics and properties of the units. In the list that follows we present some of the criteria used in the evaluation process.

Papers: The most commonly used metric for the impact of a paper is the *number of citations* received.

Authors: A number of different criteria have been defined, some of which are directly connected to the papers an author has co-authored. In brief, these criteria are: (a) the *total number of papers* co-authored, (b) the *total number of citations* received by the aforementioned papers, (c) the *mean number of citations* received, (d) the *age of the individual papers*, (e) the *age of the individual citations* received, (f) the *scientific field* treated by the author, (g) the *number of self-citations*, (h) the *number of co-authors* and, finally, (g) the *importance of the journals* where the papers are published in. Depending on the selection of criteria there are a number of different indicators that can be defined and used in order to assess an individual author.

Journals: For the evaluation of a journal, usually, what counts is the *number of papers* published within a specified period (e.g., a year) and the *number of citations* received by the papers published in the journal in a specific time window (which varies depending on the indicator examined).

3 Citation Graphs

A *Citation graph* is a graph created from the meta-data provided for each paper in a closed set of papers. A *Citation graph* is constructed based on the links that exist between papers and are defined via the References section of each paper.

3.1 Paper-Citation graph

The most common *Citation graph* is the *Paper-Citation graph* where the graph nodes are the papers of the collection and the edges represent the references given by each individual paper. The main property of the *Paper-Citation graph* is that it is a directed graph in which we may encounter cycles despite the time constraint enforced by the nature of the data.

The *Paper-Citation graph* is a directed graph since an edge connecting two nodes (a reference that links two papers) is always a directed edge from a source paper (S) to a target paper (T). In the rest of this paper, such an edge will be referred to as "S references T" or "T is cited by S" and will be denoted by $S \rightarrow T$.

One might make the assumption that the *Paper-Citation graph* is acyclic and therefore paths of the form $S \rightarrow \cdots \rightarrow S$ will not be present in the graph. This assumption could be valid since a paper (usually) only references already published /

older papers and can receive citations from (usually) younger papers, but this is not always the case for scientific publications. There exist cases where a paper is cited before its official publication (for example, an online-first edition of a paper or a pre-published version of a paper appearing in a personal web page or repository may receive citations). Therefore, even though we would expect that no cycles are present in the *Paper-Citation graph,* in practice, we can detect them and in many cases cycles that include more than two papers may be present in a particular graph.

Some of the basic concepts discussed so far are demonstrated in Figure 1, where five papers, labeled P_1, P_2, P_3, P_4, P_5, are included. By considering paper P_3 as the paper under scrutiny, we can identify the two types of relationships "References" and "Is cited by", and we can also identify the "Backward" and "Forward" citations and their affinity with time.

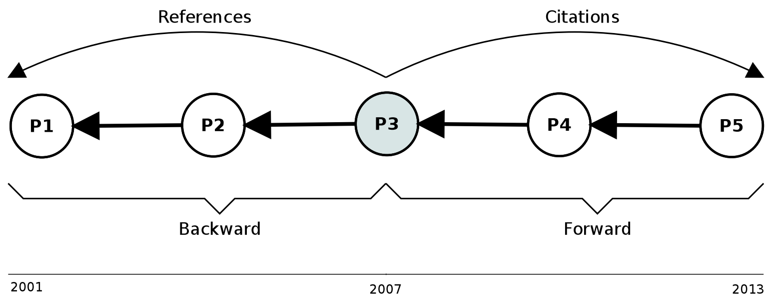


Fig. 1 A paper references other (usually older) papers via its list of references, while it receives citations from (usually younger) papers. Backward citations are defined as the papers referenced by a paper and Forward citations are defined as the citations received by a paper.

A more detailed description of the *Paper-Citation graph* can be provided with the use of a number of mathematical notations. The notations used are described in the following list:

- $\mathbf{P} = \{\mathbf{P_1}, \mathbf{P_2}, ..., \mathbf{P_{NP}}\}$ denotes the closed set of papers participating in a *Paper-Citation graph* and **NP** is the total number of papers included in the collection.
- $\mathbf{A} = \{\mathbf{A_1}, \mathbf{A_2}, ..., \mathbf{A_{NA}}\}$ denotes the set of authors that have participated in any of the papers included in the *Paper-Citation graph*. **NA** denotes the total number of authors participating in the *Paper-Citation graph*.
- $\mathbf{P(A_k)} = \{\mathbf{P_i} | \mathbf{P_i} \in \mathbf{P} \bigwedge \mathbf{A_k}\, \mathbf{has authored}\, \mathbf{P_i}\}$ denotes the set of papers authored by authorA_k.
- $\mathbf{J} = \{\mathbf{J_1}, \mathbf{J_2}, ..., \mathbf{J_{NJ}}\}$ denotes the set of journals in which the papers of the *Paper-Citation graph* where published. **NJ** denotes the total number of journals participating in the *Paper-Citation graph*.
- $\mathbf{C} = \{\mathbf{C_{P_iP_j}} | \mathbf{P_i}, \mathbf{P_j} \in \mathbf{P}\}$ denotes the set of citations between the papers included in the *Paper-Citation graph*. $C_{P_iP_j}$ denotes that paper P_j is cited by paper P_i and **NC** denotes the total number of citations (edges) present in the *Paper-Citation graph*.

- $\mathbf{a(P_i)}$ denotes the total number of authors that have co-authored paper P_i.
- $\mathbf{c(P_i)}$ denotes the total number of citations received by paper P_i.
- $\mathbf{r(P_i)}$ denotes the total number of papers referenced by paper P_i.
- $\mathbf{w(C_{P_iP_j})}$ denotes the weight of citation $C_{P_iP_j}$.

An example of a *Paper-Citation graph* is shown in Figure 2. The total number of papers (NP) is 5, and the papers included in the collection is the set $P = \{P_1, P_2, P_3, P_4, P_5\}$. Meta-data information for each paper is presented in the form of boxes around each paper.

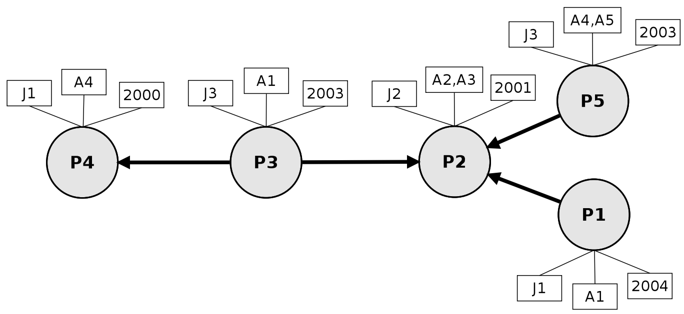


Fig. 2 The Paper-Citation graph

The meta-data information included for each paper are the author list, the year of publication and the journal of publication. The set of all co-authors of the papers in the graph is $A = \{A_1, A_2, A_3, A_4, A_5\}$, and the set of journals in which the papers were published is $J = \{J_1, J_2, J_3\}$. The relevant information is also presented in Table 1.

Paper	Publication Year	Journal	Co-Authors	References
P_1	2004	J_1	A_1	P_2
P_2	2001	J_2	A_2, A_3	-
P_3	2003	J_3	A_1	P_2, P_4
P_4	2000	J_1	A_4	-
P_5	2003	J_3	A_4, A_5	P_2

Table 1 The Paper-Citation graph information

If we apply the notations defined earlier to better describe paper P_2, it holds that:

- $a(P_2) = 2$, since P_2 was co-authored by authors A_2, A_3.
- $P(A_1) = \{P_1, P_3\}$, since author A_1 has co-authored papers P_1 and P_3.

- $c(P_2) = 3$, since P_2 is cited by papers P_1, P_3, P_5 (citations C_{12}, C_{32}, C_{52}).
- $r(P_2) = 0$, since P_2 does not reference any of the papers included in the graph.

In the *Paper-Citation graph* of Figure 2, papers P_2 and P_4 do not reference any other paper in the graph, papers P_1 and P_5 reference just one other paper, whereas paper P_3 references two papers. One might also notice that all citations appear to have the same weight, namely 1, but this is not necessarily true when examining a *Paper-Citation graph.* Papers very rarely do not provide any citation. Instead, they usually cite multiple other papers, in which case citations may not always be fully counted for. Fractional counting might be applied, thus the *Paper-Citation graph* becomes weighted, with each edge originating from paper P_i receiving a weight equal to $\frac{1}{r(P_i)}$, where $r(P_i)$ is the total number of papers referenced by paper P_i. For example, for P_2 discussed earlier, if all citations have weight 1, then the weighted citations are $\sum_{P_i \in P} \mathrm{w}(\mathrm{C}_{P_i P_2}) = c(P_2) = 3$, since P_2 is referenced by three papers, thus receiving the value of 1 from each citation. But, if fractional counting were applied to the graph, the weighted citations received by P_2 would be 2.5. The weight of citation $C_{P_3 P_2}$ would have changed from 1 to $\frac{1}{2}$, because P_3 references two papers, thus the weight of each citation would be divided by a factor of 2.

3.1.1 Derived graphs

The *Paper-Citation graph* has been used as the basic source of information for several bibliometric indicators. A number of different graphs can be constructed by extracting and transforming the information present in a *Paper-Citation graph.* We refer to these graphs as *Derived,* and, later on in this paper, we examine some of their uses in defining other types of indicators.

More specifically, the *Paper-Citation graph* also includes information about the authors of a paper and the journal in which each paper was published. We can extract these information in order to construct the *Author-Citation graph* (for author based indicators) and the *Journal-Citation graph* (for journal based indicators). The construction of both types of graphs follows the same principles and the differences lie in the distribution of the (usually) unit weight value of the original citations and the normalization method used (if any).

In all cases, the outcome is a directed, weighted graph. The resulting graphs are directed since the citations they are constructed from are directed and weighted since authors and journals are not unique in a *Paper-Citation graph,* whereas papers are. This means that even though only a single edge may exist between two papers in the *Paper-Citation graph* (a paper can only reference another paper once), this is not true for authors and journals. An author (or journal) may cite many times another author (or journal) in the paper collection, hence the unique edges between papers are replaced by multiple ones between authors (or journals). It is therefore common practice to combine the multiple edges between pairs of authors (or journals) to a single edge with an appropriate weight.

We should also mention that *Journal-Citation graphs* are structurally similar to *Author-Citation graphs* and the two types of derived graphs share a number of common attributes. Therefore, we will use the *Author-Citation graph* to demonstrate the construction of derived graphs and we will point out any differences that arise between the two types of derived graphs.

Let's consider again the *Paper-Citation graph* of Figure 2. For the construction of the *Author-Citation graph* the required information includes papers, citations and the authors of each individual paper. Therefore, we can omit the publication year and the journal of publication since they do not carry any relevant information. For the construction of the *Journal-Citation graph* on the other hand, the required information includes papers, citations and the journals in which each paper was published, thus, we can omit the publication year and the authors of the individual papers. The resulting simplified *Paper-Citation graphs* are shown in Figures 3(a) and 4(a) respectively.

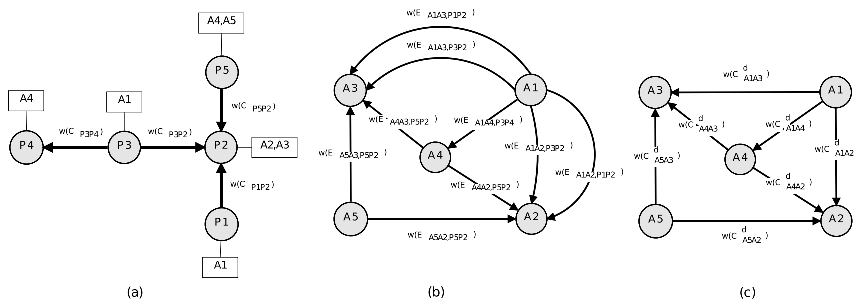


Fig. 3 (a) The simplified *Paper-Citation graph,* (b) the intermediate *Author-Citation graph* shown with all the edges between the author pairs, and (c) the *Author-Citation graph* after the replacement of the multiple edges between author pairs with unique edges (or author citations).

The nodes included in the resulting *Author-Citation graph* will be the set of authors present in the original *Paper-Citation graph.* A directed edge between authors A_k and A_l, which will be referred to as "A_k references A_l" or "A_l is cited by A_k", exists in the *Author-Citation graph* if at least one citation (edge) exists in the original *Paper-Citation graph* from a paper co-authored by A_k to a paper co-authored by A_l. Multiple such edges may exist between two authors, and the weight of each individual edge depends on the normalization method used during the construction of the *Author-Citation graph*. Figure 3(b) shows the intermediate *Author-Citation graph* where multiple edges may exist between author pairs . These edges are in turn combined to a single directed edge with a suitable weight (Figure 3(c)).

Let us define some new notations to better explain *Author-Citation graphs*. Exactly the same notations can be used for *Journal-Citation graphs* by replacing authors that author papers with journals that contain papers.

- $\mathbf{E} = \{\mathbf{E_{A_k A_l, P_i P_j}} | \exists \mathbf{C_{P_i P_j}} \in \mathbf{C} \bigwedge \mathbf{P_i} \in \mathbf{P(A_k)} \bigwedge \mathbf{P_j} \in \mathbf{P(A_l)}\}$ denotes the set of edges between authors in the intermediate *Author-Citation graph*. $E_{A_k A_l, P_i P_j}$ denotes an edge from author A_k to author A_l that exists because there exists a citation from paper P_i co-authored by A_k to paper P_j co-authored by A_l.
- $\mathbf{e(A_k)}$ denotes the total number of outgoing edges, originating from author A_k, towards the rest of the authors included in the intermediate *Author-Citation graph*.

- $\mathbf{w(E_{A_kA_l,P_iP_j})}$denotes the weight of edge $E_{A_kA_l,P_iP_j}$.
- $\mathbf{C^d} = \{\mathbf{C^d_{A_kA_l}} | \exists \mathbf{E_{A_kA_l,P_iP_j}} \in \mathbf{E}\}$denotes the set of edges between authors, or derived author citations, in the final *Author-Citation graph*.
- $\mathbf{r(A_k)}$ denotes the total number of authors referenced by author A_k in the final *Author-Citation graph*.
- $\mathbf{w(C^d_{A_kA_l})}$denotes the weight of author citation $C^d_{A_kA_l}$.

Using Figure 3(b) we can compute the $e(A_k)$ values for the authors included in this *Author-Citation graph*. For example $e(A_1) = 5$, since $A1$ has 2 edges to author $A2$, 2 to author $A3$, and 1 to author $A4$. Using Figure 3(c) we can also compute the $r(A_k)$ values of the authors. For example, $r(A_1) = 3$, since author $A1$ cites the three authors mentioned earlier, namely authors $A2$, $A3$ and $A4$.

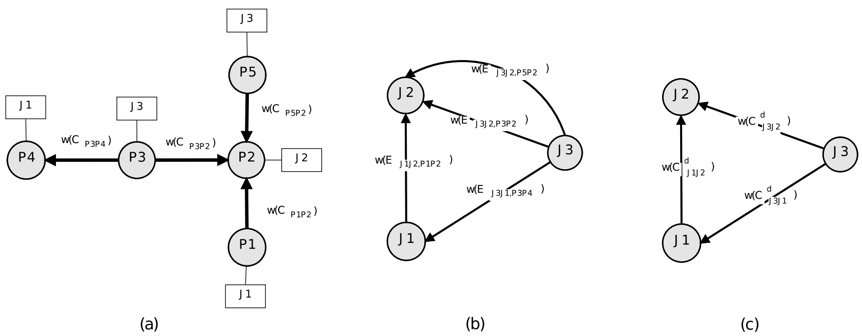


Fig. 4 (a) The simplified *Paper-Citation graph*, (b) the intermediate *Journal-Citation graph* shown with all the edges between the journal pairs, and (c) the *Journal-Citation graph* after the replacement of the multiple edges between journal-pairs with unique edges (or journal citations).

Following the same principles, the nodes included in the resulting *Journal-Citation graph* will be the set of journals present in the original *Paper-Citation graph* and the directed edges between the journals will signify the presence of citations in the original *Paper-Citation graph* between papers published in the corresponding journals. Figure 4(b) shows the intermediate *Journal-Citation graph* with the multiple edges between journals and Figure 4(c) shows the resulting *Journal-Citation graph*, after the replacement of the multiple edges with a journal citation with an appropriate weight.

To summarize, in order to construct an *Author-Citation graph* (or *Journal-Citation graph)* the weight of each paper citation in the original *Paper-Citation graph* needs to be defined. Then, the edges between the distinct author (or journal) pairs need to be specified and the weight of each edge should be calculated. In the case of *Author-Citation graphs*, the calculation depends on whether one wishes to account for the fact that more than two authors may participate in a paper citation, which results into a paper citation being translated into more than one edges in the intermediate *Author-Citation graph*. While constructing a *Journal-Citation graph* this step is redundant since a paper citation in the *Paper-Citation graph* is always translated into a single edge in the intermediate *Journal-Citation graph*. Finally, the multiple edges between the author (or journal) pairs should be replaced with a unique author (or journal) citation with an appropriate weight.

The first step defines whether Full or Fractional counting is applied to the original *Paper-Citation graph.* During the second step each paper citation is translated into edges that connect the co-authors of the papers. Since a paper can have multiple co-authors a single paper citation can correspond to multiple edges in the Author-Citation (or Journal-Citation graph). Each citation is assigned an appropriate weight. In the intermediate *Author-Citation graph,* the weights of the edges can be normalized and the actual values depend on the normalization method selected. We identify two types of normalizations that one can apply to each edge between authors A_k, A_l, namely, *No Normalization* and *Normalize per citation*.

No normalization: In this approach the weight of each edge is equal to the weight of the paper citation from which the edge is derived from.

Normalize per citation: In this approach the weight on an edge is normalized based on the number of authors participating in the formation of each paper citation present in the original *Paper-Citation graph.* So, a citation $C_{P_iP_j}$ between papers P_i with $a(P_i)$ co-authors and P_j with $a(P_j)$ co-authors, defines a total of $a(P_i) * a(P_j)$ edges between the authors of the two papers. We normalize the weight of each edge to be equal to $w(C_{P_iP_j}) \cdot \frac{1}{a(P_i)*a(P_j)}$.

As already mentioned, edge weights in the intermediate *Journal-Citation graph* do not require any normalization since a paper appears in a single journal. Finally, in the third step, the weight of each author (or journal) citation in the *Author-Citation (*or *Journal-Citation) graph* is calculated based on the sum of the weights of all the edges between the particular pair of authors (A_k, A_l) (or journals) and its final value depends on whether Full or Fractional counting is applied to the resulting graph. During this step, we distinguish three types of Fractional counting. One can divide the sum of the weights of the edges from author A_k to author A_l by either (a) the total number of authors A_k references, or (b) the total number of edges originating from A_k, or (c) the sum of the weights of the edges originating from A_k.

Summarizing the process described above and expressing each step with the mathematical notations introduced earlier, we conclude with the following equations (for clarity we present only the case of the *Author-Citation graph*; it is straightforward to replace authors with journals to get the corresponding equations for the case of the *Journal-Citation graph*; also, notice that *Normalize per citation* is meaningless in *Journal-Citation graphs*): The weight of each paper citation $C_{P_iP_j}$ between papers P_i, P_j of the original *Paper-Citation graph* is

$$w(C_{P_iP_j}) = \begin{cases} 0 & , citation\, does\, not\, exist \\ \frac{1}{r(P_i)} & , fractional\, counting \\ 1 & , full\, counting \end{cases} \tag{1}$$

The weight of each individual edge $E_{A_kA_l,P_iP_j}$ between authors A_k, A_l of papers P_i, P_j respectively of the intermediate *Author-Citation graph* is

$$w(E_{A_kA_l,P_iP_j}) = \begin{cases} 0 & , edge\,does\,not\,exist \\ w(C_{P_iP_j}) & , No\,normalization \\ w(C_{P_iP_j}) \cdot \frac{1}{a(P_i)*a(P_j)} & , Normalize\,per\,citation \end{cases} \quad (2)$$

The weight of each derived author citation $C^d_{A_kA_l}$ between authors A_k, A_l in the resulting *Author-Citation graph* is

$$w(C^d_{A_kA_l}) = \begin{cases} 0 & , citation\,does\,not\,exist \\ \sum\limits_{P_i,P_j \in P} w(E_{A_kA_l,P_iP_j}) & , full\,counting \\ \frac{1}{r(A_k)} \cdot \left(\sum\limits_{P_i,P_j \in P} w(E_{A_kA_l,P_iP_j}) \right) & , fractional\,citation\,counting \\ \frac{1}{e(A_k)} \cdot \left(\sum\limits_{P_i,P_j \in P} w(E_{A_kA_l,P_iP_j}) \right) & , fractional\,edge\,counting \\ \frac{1}{\sum\limits_{P_i,P_j \in P \wedge A_M \in A} w(E_{A_kA_M,P_iP_j})} \cdot \left(\sum\limits_{P_i,P_j \in P} w(E_{A_kA_l,P_iP_j}) \right) & , fractional\,weight\,counting \end{cases} \quad (3)$$

Based on the steps provided, let us calculate the weights of each author citation included in the derived *Author-Citation graph* of the P*aper-Citation graph* presented in Figure 2. Table 2 shows the weights of the paper citations in the original *Paper-Citation graph* depending on whether Full or Fractional counting has been applied. In addition, for each paper citation the list of all edges of the intermediate *Author-Citation graph* are listed along with the weight based on the two types of normalizations. We observe that if Full counting and *No normalization* are applied, all edges have a weight equal to 1. On the other hand, if Full counting is applied, but we choose to use the *Normalize per citation* method the results are quite different since now the unit weight of the paper citations is split equally among the receiving authors. In the case of Fractional counting, the differences lie in edges derived from paper citations of papers that cite more than one other paper, as in the case of paper P_3. Now, the weight of each citation provided by P_3 is half of what it was in the previous scenario because P_3 cites two papers in the graph.

The last step is to identify the derived author citations that should be included in the resulting *Author-Citation graph* and calculate the appropriate weights. In order to do so, we first need to locate the author pairs for which at least one edge exists, and then, replace these edges with a directed edge whose weight will originally be the sum of the weights of the individual edges. Finally, these weights should be adjusted based on whether Full counting (FUC), Fractional citation counting (FRCC), Fractional edge counting (FREC) or Fractional weight counting (FRWC) is chosen to be applied to the resulting *Author-Citation graph*. This process is presented in Tables 3 and 4.

Paper-Citation graph	Citation		Author		Edge weight	
	Notation	Weight	From	To	No normalization $\mathbf{w(C_{P_iP_j})}$	Normalize per citation $\mathbf{w(C_{P_iP_j}) \cdot \frac{1}{a(P_i)*a(P_j)}}$
Full counting	C_{P1P2}	1	A_1	A_2	1	$1 \cdot \frac{1}{1*2} = \frac{1}{2}$
	C_{P1P2}		A_1	A_3	1	$1 \cdot \frac{1}{1*2} = \frac{1}{2}$
	C_{P5P2}	1	A_4	A_2	1	$1 \cdot \frac{1}{2*2} = \frac{1}{4}$
	C_{P5P2}		A_4	A_3	1	$1 \cdot \frac{1}{2*2} = \frac{1}{4}$
	C_{P5P2}		A_5	A_2	1	$1 \cdot \frac{1}{2*2} = \frac{1}{4}$
	C_{P5P2}		A_5	A_3	1	$1 \cdot \frac{1}{2*2} = \frac{1}{4}$
	C_{P3P2}	1	A_1	A_2	1	$1 \cdot \frac{1}{1*2} = \frac{1}{2}$
	C_{P3P2}		A_1	A_3	1	$1 \cdot \frac{1}{1*2} = \frac{1}{2}$
	C_{P3P4}	1	A_1	A_4	1	$1 \cdot \frac{1}{1*1} = 1$
Fractional counting	C_{P1P2}	1	A_1	A_2	1	$1 \cdot \frac{1}{1*2} = \frac{1}{2}$
	C_{P1P2}		A_1	A_3	1	$1 \cdot \frac{1}{1*2} = \frac{1}{2}$
	C_{P5P2}	1	A_4	A_2	1	$1 \cdot \frac{1}{2*2} = \frac{1}{4}$
	C_{P5P2}		A_4	A_3	1	$1 \cdot \frac{1}{2*2} = \frac{1}{4}$
	C_{P5P2}		A_5	A_2	1	$1 \cdot \frac{1}{2*2} = \frac{1}{4}$
	C_{P5P2}		A_5	A_3	1	$1 \cdot \frac{1}{2*2} = \frac{1}{4}$
	C_{P3P2}	$\frac{1}{2}$	A_1	A_2	$\frac{1}{2}$	$\frac{1}{2} \cdot \frac{1}{1*2} = \frac{1}{4}$
	C_{P3P2}		A_1	A_3	$\frac{1}{2}$	$\frac{1}{2} \cdot \frac{1}{1*2} = \frac{1}{4}$
	C_{P3P4}	$\frac{1}{2}$	A_1	A_4	$\frac{1}{2}$	$\frac{1}{2} \cdot \frac{1}{1*1} = \frac{1}{2}$

Table 2 Edge weights derived from the *Paper-Citation graph* presented in Figure 2. The table presents the weights based on whether Full or Fractional counting has been applied to the original *Paper-Citation graph* and the two types of normalizations presented in Section 3.1.1.

In Tables 3 and 4, there are 16 different approaches in the creation of the derived *Author-Citation graph.* In each case the nodes as well as the directed edges included in the resulting *Author-Citation graph* are the same, since the only aspect of the graph that varies depending on the chosen approach is the weighting of the directed edges. Tables 5 and 6 present the 16 distinct *Author-Citation graphs* produced. Each graph is named after the choices made at each step during its creation. For example, a FUC-NN-FUC graph is constructed when Full counting is applied to the original *Paper-Citation graph*, No normalization is used for calculating the weights of the edges of the intermediate *Author-Citation graph* and, finally, Full counting is applied to the resulting *Author-Citation graph.*

Similarly, in the case of *Journal-Citation graphs* where *Normalize per citation* does not apply in the second step, there are 8 distinct types of graphs.

Author Edge			No normalization			
Author-Citation			FUC	FRCC	FREC	FRWC
C^d_{A1A2}	A_1	A_2	$1+1=2$	$\frac{1}{3} \cdot 2 = \frac{2}{3}$	$\frac{1}{5} \cdot 2 = \frac{2}{5}$	$\frac{1}{5} \cdot 2 = \frac{2}{5}$
C^d_{A1A3}	A_1	A_3	$1+1=2$	$\frac{1}{3} \cdot 2 = \frac{2}{3}$	$\frac{1}{5} \cdot 2 = \frac{2}{5}$	$\frac{1}{5} \cdot 2 = \frac{2}{5}$
C^d_{A1A4}	A_1	A_4	1	$\frac{1}{3}$	$\frac{1}{5}$	$\frac{1}{5}$
C^d_{A4A2}	A_4	A_2	1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
C^d_{A4A3}	A_4	A_3	1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
C^d_{A5A2}	A_5	A_2	1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
C^d_{A5A3}	A_5	A_3	1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$

Author Edge			Normalize per citation			
Author-Citation			FUC	FRCC	FREC	FRWC
C^d_{A1A2}	A_1	A_2	$\frac{1}{2}+\frac{1}{2}=1$	$\frac{1}{3} \cdot 1 = \frac{1}{3}$	$\frac{1}{5} \cdot 1 = \frac{1}{5}$	$\frac{1}{3} \cdot 1 = \frac{1}{3}$
C^d_{A1A3}	A_1	A_3	$\frac{1}{2}+\frac{1}{2}=1$	$\frac{1}{3} \cdot 1 = \frac{1}{3}$	$\frac{1}{5} \cdot 1 = \frac{1}{5}$	$\frac{1}{3} \cdot 1 = \frac{1}{3}$
C^d_{A1A4}	A_1	A_4	1	$\frac{1}{3}$	$\frac{1}{5}$	$\frac{1}{3}$
C^d_{A4A2}	A_4	A_2	$\frac{1}{4}$	$\frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}$	$\frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}$	$\frac{1}{\frac{1}{2}} \cdot \frac{1}{4} = \frac{1}{2}$
C^d_{A4A3}	A_4	A_3	$\frac{1}{4}$	$\frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}$	$\frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}$	$\frac{1}{\frac{1}{2}} \cdot \frac{1}{4} = \frac{1}{2}$
C^d_{A5A2}	A_5	A_2	$\frac{1}{4}$	$\frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}$	$\frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}$	$\frac{1}{\frac{1}{2}} \cdot \frac{1}{4} = \frac{1}{2}$
C^d_{A5A3}	A_5	A_3	$\frac{1}{4}$	$\frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}$	$\frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}$	$\frac{1}{\frac{1}{2}} \cdot \frac{1}{4} = \frac{1}{2}$

Table 3 Author citation weights for the derived *Author-Citation graph* of the *Paper-Citation graph* of Figure 2, when using Full counting in the original *Paper-Citation graph*.

3.2 Author-citation graph examples

A FUC-NC-FUC derived graph can be found in (Radicchi, Fortunato, Markines, and Vespignani 2009; Radicchi, Fortunato, and Vespignani 2012), where the *Author-Citation graph* is constructed from the *Paper-Citation graph* and is called the *Weighted Author Citation Network (WACN)*. The authors use a *Paper-Citation graph* with Full citation counting and normalize the weights of the individual edges between authors using the Normalize per citation method. Full citation counting is also applied for the author citations to the resulting *Author-Citation graph*.

A FRC-NC-FUC derived graph can be found in (West et al 2013), where the authors apply Fractional citation counting to the original *Paper-Citation graph* and

Author Edge			No normalization			
Author-Citation			**FUC**	**FRCC**	**FREC**	**FRWC**
C^d_{A1A2}	A_1	A_2	$1+\frac{1}{2}=\frac{3}{2}$	$\frac{1}{3}\cdot\frac{3}{2}=\frac{1}{2}$	$\frac{1}{5}\cdot\frac{3}{2}=\frac{3}{10}$	$\frac{1}{\frac{7}{2}}\cdot\frac{3}{2}=\frac{3}{7}$
C^d_{A1A3}	A_1	A_3	$1+\frac{1}{2}=\frac{3}{2}$	$\frac{1}{3}\cdot\frac{3}{2}=\frac{1}{2}$	$\frac{1}{5}\cdot\frac{3}{2}=\frac{3}{10}$	$\frac{1}{\frac{7}{2}}\cdot\frac{3}{2}=\frac{3}{7}$
C^d_{A1A4}	A_1	A_4	$\frac{1}{2}$	$\frac{1}{3}\cdot\frac{1}{2}=\frac{1}{6}$	$\frac{1}{5}\cdot\frac{1}{2}=\frac{1}{10}$	$\frac{1}{\frac{7}{2}}\cdot\frac{1}{2}=\frac{1}{7}$
C^d_{A4A2}	A_4	A_2	1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
C^d_{A4A3}	A_4	A_3	1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
C^d_{A5A2}	A_5	A_2	1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
C^d_{A5A3}	A_5	A_3	1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$

Author Edge			Normalize per citation			
Author-Citation			**FUC**	**FRCC**	**FREC**	**FRWC**
C^d_{A1A2}	A_1	A_2	$\frac{1}{2}+\frac{1}{4}=\frac{3}{4}$	$\frac{1}{3}\cdot\frac{3}{4}=\frac{1}{4}$	$\frac{1}{5}\cdot\frac{3}{4}=\frac{3}{20}$	$\frac{1}{2}\cdot\frac{3}{4}=\frac{3}{8}$
C^d_{A1A3}	A_1	A_3	$\frac{1}{2}+\frac{1}{4}=\frac{3}{4}$	$\frac{1}{3}\cdot\frac{3}{4}=\frac{1}{4}$	$\frac{1}{5}\cdot\frac{3}{4}=\frac{3}{20}$	$\frac{1}{2}\cdot\frac{3}{4}=\frac{3}{8}$
C^d_{A1A4}	A_1	A_4	$\frac{1}{2}$	$\frac{1}{3}\cdot\frac{1}{2}=\frac{1}{6}$	$\frac{1}{5}\cdot\frac{1}{2}=\frac{1}{10}$	$\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{4}$
C^d_{A4A2}	A_4	A_2	$\frac{1}{4}$	$\frac{1}{2}\cdot\frac{1}{4}=\frac{1}{8}$	$\frac{1}{2}\cdot\frac{1}{4}=\frac{1}{8}$	$\frac{1}{\frac{1}{2}}\cdot\frac{1}{4}=\frac{1}{2}$
C^d_{A4A3}	A_4	A_3	$\frac{1}{4}$	$\frac{1}{2}\cdot\frac{1}{4}=\frac{1}{8}$	$\frac{1}{2}\cdot\frac{1}{4}=\frac{1}{8}$	$\frac{1}{\frac{1}{2}}\cdot\frac{1}{4}=\frac{1}{2}$
C^d_{A5A2}	A_5	A_2	$\frac{1}{4}$	$\frac{1}{2}\cdot\frac{1}{4}=\frac{1}{8}$	$\frac{1}{2}\cdot\frac{1}{4}=\frac{1}{8}$	$\frac{1}{\frac{1}{2}}\cdot\frac{1}{4}=\frac{1}{2}$
C^d_{A5A3}	A_5	A_3	$\frac{1}{4}$	$\frac{1}{2}\cdot\frac{1}{4}=\frac{1}{8}$	$\frac{1}{2}\cdot\frac{1}{4}=\frac{1}{8}$	$\frac{1}{\frac{1}{2}}\cdot\frac{1}{4}=\frac{1}{2}$

Table 4 Author citation weights for the derived *Author-Citation graph* of the *Paper-Citation graph* of Figure 2, when using Fractional counting in the original *Paper-Citation graph*.

normalize the weights of the edges between authors using the Normalize per citation method. Full citation counting is applied for the author citations to the resulting *Author-Citation graph.* Self citations are completely removed from the resulting *Author-Citation graph*.

3.3 Journal-citation graph examples

A FUC-NN-FUC derived graph can be found in Bollen, Rodriguez, and Van de Sompel (2006), while a FUC-NN-FREC derived graph is found in Bergstrom (2007);

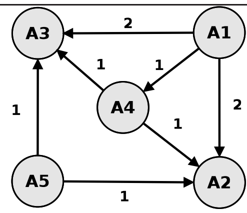
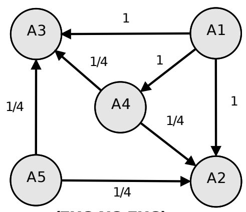
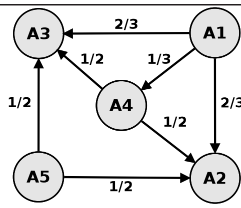
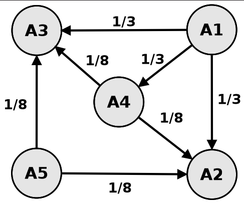
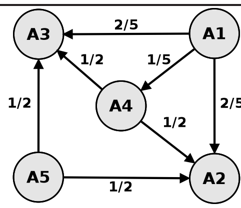
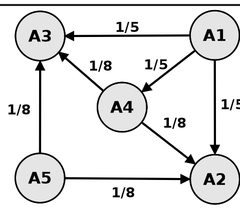
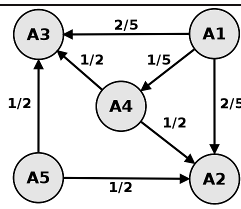
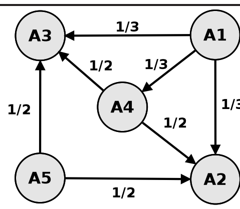
Paper-Citation	Full counting (FUC)	
Author edge	No normalization (NN)	Normalize per citation (NC)
Author-Citation		
FUC	 (FUC-NN-FUC)	 (FUC-NC-FUC)
FRCC	 (FUC-NN-FRCC)	 (FUC-NC-FRCC)
FREC	 (FUC-NN-FREC)	 (FUC-NC-FREC)
FRWC	 (FUC-NN-FRWC)	 (FUC-NC-FRWC)

Table 5 The eight approaches for the construction of the *Author-Citation graph,* when using Full counting in the original *Paper-Citation graph.*

González-Pereira, Guerrero-Bote, and Moya-Anegón (2010); Guerrero-Bote and Moya-Anegón (2012) with slight differences. The differences lie in the way journal self-citations are treated and in the time-constraints imposed during the construction of the graph.

More specifically, Bollen, Rodriguez, and Van de Sompel (2006) do not consider journal self-citations while constructing the graph. On the other hand, in the methods used to construct the *Journal-citation graph* for the computation of the Eigen-Factor scores (eigenFACTOR.org 2008) journal self-citations are completely removed, whereas González-Pereira, Guerrero-Bote, and Moya-Anegón (2010) restrict the number of journal self-citations to 33% of the journal's overall citation count. The

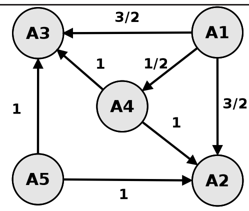
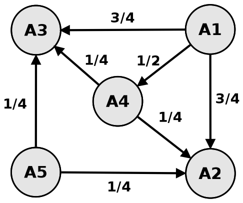
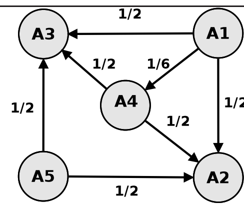
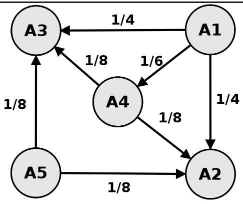
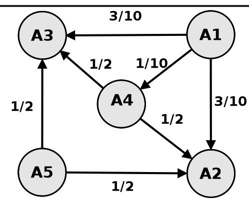
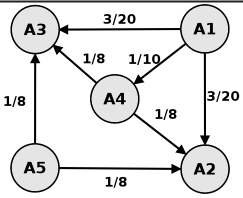
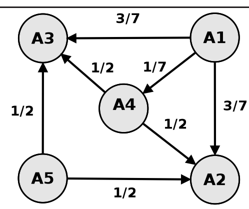
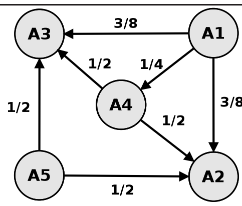
Paper-Citation	Fractional counting (FRC)	
Author edge	No normalization (NN)	Normalize per citation (NC)
Author-Citation		
FUC	 (FRC-NN-FUC)	 (FRC-NC-FUC)
FRCC	 (FRC-NN-FRCC)	 (FRC-NC-FRCC)
FREC	 (FRC-NN-FREC)	 (FRC-NC-FREC)
FRWC	 (FRC-NN-FRWC)	 (FRC-NC-FRWC)

Table 6 The eight approaches for the construction of the *Author-Citation graph,* when using Fractional counting in the original *Paper-Citation graph.*

same 33% limit can be implied for the self-citations included in the *Journal-citation graph* used in Guerrero-Bote and Moya-Anegón (2012) based on the fact that the authors follow a similar procedure with the one presented in González-Pereira et al (2010) in order to propose two indicators that extend the ones presented in the later paper.

Timewise, Bollen, Rodriguez, and Van de Sompel (2006) use the generated graph as is, whereas for the calculations of the EigenFactor scores (eigenFACTOR.org 2008) only the subset of citations falling into a specific five-year window are included in the graph. By imposing this limitation, the produced graph contains a subset of the available information local to a specific time-period. We refer to this property of the particular graphs as *time-awareness.* A similar time constraint is also imposed by González-Pereira et al (2010) and Guerrero-Bote and Moya-Anegón (2012)with the time-window set to three years instead of five.

4 Citation generations

Although the ideas presented in this section can be applied to any citation graph, we focus our discussion on Paper-citation graphs. A generation of citations is defined as the collection of papers that cite a target paper either directly (first generation) or indirectly (via a path in the citation graph originating from a source paper and ending to the target paper). Thus, taking into account the entire citation graph, we define backward generations as the collection of papers referenced directly or indirectly from the current paper and forward generations as the collection of papers directly or indirectly citing the current paper.

Backward and forward generations were originally defined by Rousseau (1987). They have also been studied as such by Dervos and Kalkanis (2005) and expanded to also include self-citations and chords, which we analyze later in this section, in Dervos, Samaras, Evangelidis, and Folias (2006). In addition, second generation citations were discussed by Kosmulski (2010), and finally, a generalization of the forward and backward citation generations was presented by Hu, Rousseau, and Chen (2011).

Backward and forward generations share exactly the same properties and are defined in exactly the same manner. For the rest of the paper we only consider forward generations, but everything mentioned applies to backward and forward generations alike. There are four different definitions for generations depending on whether a paper may appear more than once in a generation and whether a paper that has already participated in a generation can be used in another (later) generation (Hu, Rousseau, and Chen 2011). Based on the notations used in (Hu, Rousseau, and Chen 2011), we present the eight different types of generations in Table 7:

- Forward and Backward generations are denoted with a subscript n, with n being either a positive natural number (Forward generations) or a negative whole number (Backward generations).
- H denotes that the generations are to be defined independently and G denotes that the generations can only include papers not already included in a previous generation.
- The superscript s denotes that a paper can only be included once in a particular generation (the generation is a set) and m denotes that a paper can be included more than once in a generation (the generation is a multiset).

Now, let us consider the *Paper-Citation graph* of Figure 5 in which we have only included the papers and their citations. Table 8 shows the different sets of forward citation generations according to the definitions listed above for paper P_1, which we consider to be the only paper included in generation 0.

Type	Relation	Papers	Notation
Forward	Independent	Unique papers per generation	H^s_n
		Non-unique papers per generation	H^m_n
	Restricted	Unique papers per generation	G^s_n
		Non-unique papers per generation	G^m_n
Backward	Independent	Unique papers per generation	H^s_{-n}
		Non-unique papers per generation	H^m_{-n}
	Restricted	Unique papers per generation	G^s_{-n}
		Non-unique papers per generation	G^m_{-n}

Table 7 Generations definitions

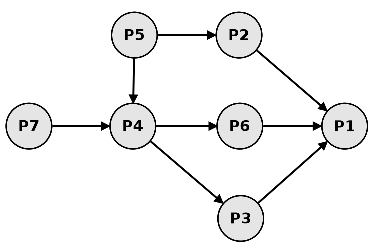


Fig. 5 Example *Paper-Citation graph*.

	Non-unique	Unique
Independent	$H^m_0 = \{P_1\}$	$H^s_0 = \{P_1\}$
	$H^m_1 = \{P_2, P_3, P_6\}$	$H^s_1 = \{P_2, P_3, P_6\}$
	$H^m_2 = \{P_4, P_4, P_5\}$	$H^s_2 = \{P_4, P_5\}$
	$H^m_3 = \{P_5, P_5, P_7\}$	$H^s_3 = \{P_5, P_7\}$
Restricted	$G^m_0 = \{P_1\}$	$G^s_0 = \{P_1\}$
	$G^m_1 = \{P_2, P_3, P_6\}$	$G^s_1 = \{P_2, P_3, P_6\}$
	$G^m_2 = \{P_4, P_4, P_5\}$	$G^s_2 = \{P_4, P_5\}$
	$G^m_3 = \{P_7\}$	$G^s_3 = \{P_7\}$

Table 8 Different types of forward citation generations for paper P_1.

Some citations that are worth noting are the ones originating from paper P_5. P_5 provides P_1 with a second generation citation via P_2 (path $P_5 \to P_2 \to P_1$) and with two third generation citations via paths $P_5 \to P_4 \to P_6 \to P_1$ and $P_5 \to P_4 \to P_3 \to P_1$. This means that P_5 will be included in the second generations for all four definitions, but for the third generations the following are true: (a) P_5 is included in H^s_3 once since it provides at least one third generation citation to P_1, (b) P_5 is included in H^m_3 twice and this is the exact number of third generation citations provided to P_1, and (c) P_5 is not included at all in G^s_3 and G^m_3, since it has already been included in the corresponding G^s_2 and G^m_2 due to the second generation citation it provides to P_1.

The generations defined by Rousseau 1987 correspond to the G^m_{-n} generations defined here. In that paper, backward citation generations are used to determine the influence that references have on the paper under scrutiny.

In Dervos and Kalkanis 2005; Dervos, Samaras, Evangelidis, and Folias 2006, the authors define the *Cascading Citations Indexing Framework (cc-IF)* in which citation generations are defined as H^m_n generations. In these papers, self-citation generations are also defined at the (paper, author) level. The definition follows the same pattern but instead of considering the citations at the paper level, the author is examined as well. If a citation originates from a paper that includes the current author in its author list, then it is considered to be an n-generation self-citation for the current author. Another aspect of the citations examined in these papers is the existence of chords. For a target paper B, a chord is defined to be a first generation citation $A \to B$

that co-exists with a citation $A \to \cdots \to B$ of higher rank (second, third, fourth, etc. generation citation).

Finally, Kosmulski 2010 studies forward generations and particularly second generation citations of the G_n^s definition.

5 Indirect bibliometric indicators

The indicators presented in this section have been used either to measure the importance of a particular research unit or as a general methodology that applies to all types of research units as long as the appropriate unit-citation graph is used. Some of the indicators utilize the notion of citation generations as such, while others do so indirectly by utilizing the information present in the entire citation graph. For example, the *Gozinto theorem,* proposed by Rousseau (1987), specifically determines the citation generations, while the popular *PageRank* algorithm, proposed by Page, Brin, Motwani, and Winograd (1999) and applied to bibliometrics by a number of researchers, is based on the information present in the whole citation graph without specifically naming citation generations as such.

PageRank, originally inspired by citation analysis and used for ranking pages on the web, has again found its way back to bibliometrics with many researchers attempting to explore the interlinking of the research units via their citation patterns. *PageRank* is defined recursively by equally dividing the influence value of a web page to its connected pages via the outbound links found on the page. The model imitates a "random surfer" who chooses to blindly follow one of the outbound links of a page, and thus, navigates through the web in a number of random hops. The surfer from time to time chooses to end his current path and start a new one from a completely different point in the web, and does so by a probability defined by a preselected damping factor. The damping factor chosen in the original implementation of the algorithm was 0.85 (Page, Brin, Motwani, and Winograd 1999). The *PageRank* scores of the pages are calculated by the following formula:

$$PR_a = (1-d) + d \cdot \sum_i \frac{PR_i}{N_i} \tag{4}$$

where, PR_a is the score of the current page (page a), d is the damping factor, PR_i is the score of the individual pages directly citing paper a and N_i are the total pages cited by each page i.

5.1 Paper assessment

The Gozinto theorem (Rousseau 1987)**:** In this paper, Rousseau determines the papers that had the greatest influence in the creation of the paper under scrutiny. Papers included in the reference list of the current paper (first generation) had direct influence whereas papers included in the reference lists of those papers (second generation) are considered to have an indirect influence.

The direct influence of a paper can be given a weight or considered to have a weight equal to 1. The weights of all direct influences that papers had among each other forms an nxn matrix A, where n is the total number of papers. The author states that there are many ways in which the weights can be assigned and in an example included in the paper two different methods are presented. The first method assigns an integer value in the references mentioned in the different sections of the paper under scrutiny. The direct influence is then calculated as the sum of all distinct values of all occurences of the particular reference within the paper. A derived method uses this weight to calculate a weight that transforms all weights to numbers between 0 and 1.

Rousseau utilizes matrix A in his calculations of the total influence along with the Gozinto theorem. Based on the theorem, the total influence of paper A_i on A_j is the sum examined over all papers (z) of the direct influence of A_i on A_k (a_{ik}) times the total influence of A_k on A_j (c_{kj}) and is given by:

$$c_{ij} = \sum_{k=1}^{z} a_{ik} \cdot c_{kj} + \delta_{ij} \tag{5}$$

Where δ_{ij} denotes the Kronecker delta and is defined as

$$\delta_{ij} = \begin{cases} 1 & , i = j \\ 0 & , i \neq j \end{cases} \tag{6}$$

For more details on the calculations of the total influence we refer the reader to the original paper (Rousseau 1987).

SCEAS Rank (Sidiropoulos and Manolopoulos 2005)**:** A recursive scoring algorithm that wishes to minimize some of the side effects of the original PageRank algorithm. According to the authors the proposed score meets two conditions that are not present in the original PageRank algorithm. More precisely, the following are true: (a) the factor that should have the greatest influence over the score of a particular paper should be the number of direct citations and, (b) the addition of new citations in the *Paper-Citation graph* should have a greater effect in the scores of nearby rather than distant papers. In that respect, they proposed the following scoring formula:

$$S_a = \sum_i \frac{S_i + b}{N_i} a^{-1} \; (a \geq 1, b > 0) \tag{7}$$

where, S_a is the score of the current paper (paper a), S_i is the score of the individual papers directly citing paper a, N_i is the total number of papers cited by each paper i, b denotes the *direct citation enforcement factor* (which controls the effect that direct citations have to the calculated score) and a denotes the speed with which an indirect citation enforcement converges to zero.

The authors also propose a generalization of the above formula and the original PageRank algorithm that introduces a damping factor (d) in the SCEAS rank:

$$S_a = (1 - d) + d \cdot \sum_i \frac{S_i + b}{N_i} a^{-1} \; (a \geq 1) \tag{8}$$

PageRank (Ma, Guan, and Zhao 2008)**:** The original PageRank algorithm applied to the *Paper-Citation graph* by changing the damping factor from its original value to 0.5. The selection of this value was based on an earlier study that, according to the authors of the paper, indicates that in the random surfer model for scientific papers the path followed is much shorter (in particular two papers).

Cumulative patent citations (Atallah and Rodríguez 2006)**:** The proposed indicator was used as a means to measure the importance and quality of patents. It uses the *Patent-Citation graph* that is identical to the *Paper-Citation graph* with the only difference being that the nodes of the graph are patents instead of papers. Thus, this indicator can be used in the context of paper assessment and, therefore, is included in this study.
The *Cumulative patent citations* measure represents the sum of all direct and indirect citations received by a given patent. So, for a *Patent-Citation graph* with N patents the score of generation j of citations received by patent x is given in (9), where, $a_i(x) = 1$ if a path exists between patents i and x and $a_i(x) = 0$ otherwise. In more detail, the generational score of a patent is the sum of all direct citations received by the patents included in the previous generation.

$$S_j(x) = \sum_{i=1}^{N} a_i(x) \cdot S_{j-1}(i) \tag{9}$$

The score of patent x is then calculated by adding the individual generation scores from 0 to M, where, M is the maximum generation of citations for the patent x, i.e., the length of the longest path present in the *Patent-Citation graph* that terminates at x, and is given in (10).

$$S_T(x) = \sum_{i=0}^{M} S_i(x) \tag{10}$$

Weighted cumulative patent citations (Atallah and Rodríguez 2006)**:** This is a weighted version of the *Cumulative patent citations* indicator and its purpose was to account for the closeness of citations to the cited patent. The weight of a generation i is calculated based on its distance from the patent, which means that generations closer to the patent have a greater influence to the score of the patent in question. The calculation of the *Weighted cumulative patent citations* is given in (11).

$$S_w(x) = \sum_{i=0}^{M} \left(1 - \frac{i}{M+1}\right) \cdot S_i(x) \tag{11}$$

CiteRank (Walker, Xie, Yan, and Maslov 2007; Maslov and Redner 2008)**:** This is an adaptation of the original PageRank algorithm that takes into consideration the fact that researchers usually traverse papers starting from a relatively new paper and following its references. So, apart from including the damping factor as the probability that the researcher will drop their current search path and start a new one, they also include a decay time (T_{dir}) that controls the probability that a paper will be selected as the start of a new research path. This probability is defined in (12)

$$p_i = e^{-age_i/T_{dir}} \tag{12}$$

where, age_i denotes the age of the paper. Therefore, more recent papers have a higher probability to be selected as the starting point of a random walk.

f-value (Fragkiadaki, Evangelidis, Samaras, and Dervos 2011)**:** An indicator based on recursive calculations across the entire citation graph that takes into account the impact of the citing articles in order to identify the impact of the article in question. A reducing factor, whose value could account for the different scientific fields, is used to mitigate the impact transferred from the different generations of citations. For paper P_i with $c(P_i)$ direct citations, the *f-value* is defined as

$$f(P_i) = 1 + \frac{1}{2.2} \cdot \sum_{j=1}^{c(P_i)} f(P_j) \tag{13}$$

The fraction $\frac{1}{2.2}$ is the reducing factor calculated based on the data set used and $f(P_j)$ represents the f-value of paper P_j directly citing paper P_i. For more details on the calculations of the reducing factor, we refer the reader to the original paper (Fragkiadaki et al 2011).

P-Rank (Yan, Ding, and Sugimoto 2011)**:** This is a proposition to evaluate articles taking into account the heterogeneity of the citation networks. It uses a citation network that includes papers, citations, authors and journals and the final value is calculated as a combination of the importance of all three factors. For the papers aspect, the PageRank algorithm is used, whereas for the author and journal aspects the adjacency matrices are utilized.

PrestigeRank (Su, Pan, Zhen, Ma, Yuan, Guo, Yu, Ma, and Wu 2011)**:** The authors discuss the problem of the incompleteness of bibliometric databases, in the sense that usually not all citations to (or from) papers included in the database are always present in the *Paper-Citation graph* used to calculate the paper scores based on PageRank. This poses a problem in the computations, since a paper providing one internal citation (to a paper included in the database) and five external citations (to papers not included in the current system), in the standard PageRank algorithm will transfer all its influence to the single paper included in the system instead of diving its score to all the referenced papers.

In order to solve the problem, they introduce a “virtual node” that accumulates all citations originating from papers within the database to external papers and that is also responsible for providing all citations from external papers. The “virtual node” instantly solves the problem of diving the influence of a paper to the referenced papers, but in order to also account for all external citations some more computations are needed. Based on the assumption that the more internal citations a paper has received the more external citations it should have received they divide the score of the "virtual node" to the papers in the database.

5.2 Author assessment

SARA (Radicchi, Fortunato, Markines, and Vespignani 2009)**:** SARA (Science Author Rank Algorithm) utilizes a FUC-NC-FUC *Author-Citation graph* in order to calculate author scores based on a PageRank approach. The citations used to construct the *Author-Citation graph* are defined dynamically using a set of M_I overlapping homogeneous intervals of the full date ordered list of citations. The intervals are defined as homogeneous because they contain the exact same number of citations (M_R), and overlapping because each qth interval shares the first $M_R/2$ citations with the $(q-1)$th interval and its last $M_R/2$ citations with the $(q+1)$th interval. The algorithm differs from the original PageRank since part of the credit attributed to each author includes a contribution from the whole network. The authors define the "scientific dept" of a scientist as the knowledge gained by the whole field and the value attributed to each author is proportional to his productivity and the number of co-authors of each one of his papers. The SARA score of author i is given by

$$P_i = \left((1-q) \cdot \sum_j \frac{P_j}{s_j^{out}} \cdot w_{ji}\right) + \left(q \cdot z_i + (1-q) \cdot z_i \sum_j P_j \cdot \delta(s_j^{out})\right) \tag{14}$$

The first term is the distribution of the value of all authors citing i, which is weighted based on the total number of authors cited by each author, while the second term is the sum of the scientific dept received by the current author from all other authors in the community whether they do cite other authors or not (dangling nodes). In the equation, q represents the damping factor, P_i is the score of node i, w_{ji} is the weight of the directed connection from j to i, s_j^{out} is the sum of the weights of all links outgoing from the jth vertex, and if $x = 0$ then $\delta(x) = 1$ otherwise $\delta(x) = 0$. z_i is a factor that considers the normalized scientific credit given to the author i based on his productivity. For a more detailed description of equation (14) and of the variables used, we refer the reader to Radicchi, Fortunato, Markines, and Vespignani (2009). Finally, the authors propose two rankings for the authors, an absolute one and a relative one (the relative one being used to account for different historical periods).

hfg-index (Kosmulski 2010)**:** The hfg-index is a successive Hirsch-type index that is based on first calculating the h-index values for the papers a scientist has co-authored (Schubert 2009). The h-index of a paper is defined as the number h of papers citing the current paper that have received at least h citations each. Having calculated the h-index values of the papers co-authored by the scientist, his hfg-index is defined in the same way as the h-index as the largest hfg of papers that each have an h-index of at least hfg, whereas the remaining papers of the scientist do not have an h-index greater than hfg. This indicator may not be recursive in its calculation but it does utilize more information from the citation graph than the direct citations (one extra generation of citations is used).

Indirect H-index (Egghe 2011a,b)**:** This is an indicator very similar to the hfg-index (Kosmulski 2010), as it is based on the same concept of creating a listing of the h-indices of the papers of an author, and then using these values instead of the number of citations to calculate the indirect H-index of the author. The main difference is that this indicator was proposed as a complementary indicator and it is calculated for the papers participating in the Hirsch-core of the author rather than on his whole set of papers. Thus, the papers included in the hfg-core may very well be completely different from the ones included in this indirect H-index core, since papers with very few first generation citations will probably not be included in the indirect H-index core, whereas if they have received few citations from high-impact papers they may be included in the hfg-core.

Generational indices (Hu, Rousseau, and Chen 2011)**:** A general term used to describe indicators that are calculated for a particular generation of citations. An example given for a Generational index is the generational h-index. The authors having defined each generation in accordance with one of the four ways described in Section 4, they define the h-index of generation $n, n \geq 0$ as the number h of papers included in generation n each of which has received at least h citations.

Cross-generational indices (Hu, Rousseau, and Chen 2011)**:** A general term used to describe indicators that are calculated based on values calculated per generation of citations. The authors define the general generational indicator C_n for a sequence of $A = (a_k)_{k=0,..,n-1}$ of forward generational indicators as $max(a_0, a_1, .., a_{n-1})$. For example, if the generational indicator used is the h-index then the C_n cross-generational indicator is equal to $max(h_0, h_1, .., h_{n-1})$. The definition also applies when backward generational indicators are considered.

The authors also describe a case where the cross-generational index considers the generations in a cumulative manner instead of using them as standalone values. An example provided is the *Total influence indicator,* defined as the sum of the generational indices divided by the generation factorial. Again, the generational h-index can be used as a generational indicator. In both cases the selection of the number of generations considered is selected arbitrarily.

Eigenfactor score for authors (West et al 2013)**:** The EigenFactor score for authors is an adaptation of the original Eigenfactor score algorithm (Bergstrom 2007; Bergstrom et al 2008; eigenFACTOR.org 2008) used for Journal assessment. It uses a FRC-NC-FUC *Author-Citation graph* from which all self-citations are excluded. As discussed in Section 5.3, EigenFactor imitates the original PageRank algorithm.

Influence weight of the journal (Pinski and Narin 1976)**:** This indicator was described by means of journals but, as the authors of the paper state, the same concept can be applied to any unit participating in a citation network as long as the corresponding derived graph is used. The Influence weight is meant to be a size-independent measure of the number of citations a journal receives from all other journals participating in the citation network, normalized by the number of references the journal gives to other journals. The Influence weight is a measure of the influence that each reference provided by this journal has.

Influence per publication for the journal (Pinski and Narin 1976)**:** To calculate the Influence per publication for a journal, one needs to know the Influence weight of the journal, the number of references that the journal made and the papers published within a year. Since the Influence weight is a measure of influence per reference, by multiplying by the number of references and dividing by the number of publications, we get the Influence per publication. This indicator is included in this category as a derived indicator based on the calculations for the Influence weight of a journal.

Total influence of the journal (Pinski and Narin 1976)**:** The Total influence of a journal is its Influence per publication multiplied by the number of publications within a year. This indicator is included in this category as a derived indicator based on the calculations for the Influence weight of a journal.

Weighted PageRank (Bollen, Rodriguez, and Van de Sompel 2006)**:** A modification of the original PageRank algorithm that instead of equally dividing the credit of a journal to the references it gives to other journals, it uses a weighting that is proportional to the amount of references given per journal. So, if a journal cites a particular journal more often, that should be done via links that weigh more. The formula used to calculate the Weighted PageRank score for journal a is

$$PR_a^w = \frac{(1-d)}{N} + d \cdot \sum_i PR_i^w \cdot w(i,a) \tag{15}$$

where, $w(i,a)$ is the weighting function, which according to the authors is given by

$$w(v_j, v_i) = \frac{W(v_i, v_j)}{\sum_k W(v_j, v_k)} \tag{16}$$

where, $W(v_i, v_j)$ maps each edge between the journal v_i and v_j to a positive, citation frequency. This means that the *Journal-Citation graph* used in this version of PageRank is constructed based on the method that does not normalize the weights of the edges, which are later normalized via the weighting function (16).

Y-factor (Bollen, Rodriguez, and Van de Sompel 2006)**:** The Y-factor is calculated as the product of the Weighted PageRank value of a journal and its Impact factor (Garfield 1999, 2005). It is listed in this category because it is a derived indicator based on the Weighted PageRank. Equation (17) shows the formula used to calculate the Y-factor

$$Y_a = ISIIF_a \cdot PR_a^w \tag{17}$$

where $ISIIF_a$ denotes the Impact Factor value for journal a, and PR_a^w denotes the Weighted PageRank value for the same journal.

Eigenfactor™ score (Bergstrom 2007; Bergstrom, West, and Wiseman 2008)**,** (eigenFACTOR.org 2008)**:** The EigenFactor score is an indicator of the total influence of a journal. It uses a FUC-NN-FREC *Journal-Citation graph*, with a five-year citation window and with all journal self-citations excluded. EigenFactor imitates the original PageRank algorithm by calculating the journal influence vector, which in turn is used to calculate the EigenFactor score as the percentage of citations received by the journal in question from all other journals included in the graph. The authors mention that the EigenFactor metrics can be used at the article and author levels as well.

Article influence™ score (Bergstrom 2007; Bergstrom, West, and Wiseman 2008)**:** The Article influence score is a derived indicator, whose calculation is based on the EigenFactor Score of a journal divided by the number of papers published in the journal for the five year period normalized by the total number of papers published in the same period. This yields the per-article influence of the journal which can be compared to the Impact Factor.

PSJR - Prestige SJR (González-Pereira, Guerrero-Bote, and Moya-Anegón 2010)**:** PSJR is a size-dependent metric used to calculate the overall journal prestige and influence. Its calculation is based on a FUC-NN-FREC *Journal-Citation graph*, with a three year citation window and the number of self-citations per journal restricted to 33% of its overall citations.
It is recursively calculated for each journal and the final value depends upon three terms. The first term is constant and represents a minimum value assigned to each journal in the citation graph. The second term, also constant, represents the *prestige of the papers* that is the number of the papers included in the journal normalized by the total number of papers published by all journals included in the graph. Finally, the third term represents the *prestige of the citations* and is given by the weighted PSJR values of the citing journals and a constant value that represents the portion of the PSJR value of the dangling nodes of the graph (journals that do not cite any other journal) assigned to the current paper. The overall outcome of the calculations can be tuned by two constants (namely, e and d) that control the effect of the prestige of the papers and citations respectively. The PSJR is given by

$$PSJR_i = \left(\frac{1-d-e}{N}\right) + \left(e \cdot \frac{Art_i}{\sum_{j=1}^{N} Art_j}\right) + \left(d \cdot \sum_{j=1}^{N} C_{ji} \cdot \frac{PSJR_j}{C_j} \cdot CF + \frac{Art_i}{\sum_{j=1}^{N} Art_j} \cdot \sum_{k \epsilon DN} PSJR_k\right) \tag{18}$$

In the equation, Art_i is the number of primary items of journal i, C_{ji} represents the references from journal j to journal i, C_j is the number of references of journal j, CF is a correction factor used to spread the undistributed prestige and DN represents the dangling nodes. For a more detailed description of each element we refer the reader to González-Pereira, Guerrero-Bote, and Moya-Anegón (2010).

SJR indicator (González-Pereira, Guerrero-Bote, and Moya-Anegón 2010)**:** The SJR indicator is a size-independent metric that calculates the average prestige per paper published in a specific journal and, as such, it can be used to compare journals that publish different number of items. It is calculated by dividing the PSJR value of a journal by the number of papers published and multiplying the result by a constant value c that makes the outcome more easily readable. The SJR indicator for journal i is given by

$$SJR_i = c \cdot \frac{PSJR_i}{Art_i} \tag{19}$$

PR_{sum} **- Total authority** (Su, Pan, Zhen, Ma, Yuan, Guo, Yu, Ma, and Wu 2011)**:** This is a derived indicator based on the scores of the papers included in the *Paper-Citation graph* and calculated with the PrestigeRank indicator. It is defined as the sum of all PrestigeRank values of all papers published in a journal and for journal i is given by

$$PR_{sum} = \sum_{i=1}^{N} PR_i \tag{20}$$

The authors consider this indicator to be the equivalent of the citation counts.

PR_{ave} **- Authority factor** (Su, Pan, Zhen, Ma, Yuan, Guo, Yu, Ma, and Wu 2011)**:** It is a derived indicator based on the scores of the papers included in the *Paper-Citation graph* and calculated with the PrestigeRank indicator. It is defined as the average PrestigeRank values for the papers published in a journal and for journal i is given by

$$PR_{ave} = \frac{1}{N}\sum_{i=1}^{N} PR_i \tag{21}$$

The authors consider this indicator to be the equivalent to the Impact Factor for journals.

Recursive Mean Normalized Citation Score (Recursive MNCS) (Waltman et al 2011b): The Recursive MNCS is based on the non-recursive MNCS indicator originally defined to account for differences among scientific fields (Waltman et al 2011a). The MNCS indicator is defined over a set of papers and is equal to the average *Normalized citation score* of the papers in the set. The *Normalized citation score* for each paper is defined as the Total number of citations received, divided by the *Expected number of citations* for papers published in the field, which is equal to the average number of citations received per paper in the field. In Waltman et al (2011b) the authors define the described MNCS indicator as the first-order MNCS indicator used to recursively calculate higher-order MNCS indicators for journals and institutions. For the calculations of these higher order indicators the authors assign varying weights to each citation based on the previous order MNCS value of the citing journal. In an empirical study presented in the same paper, the authors, apart from the scientific field, also consider the publication year of the papers included in the calculations. It is should be mentioned though that the authors conclude that the combination of normalized citation counts, to account for differences among scientific fields, along with recursive citation weighing does not produce satisfactory results. For a detailed presentation of the empirical results and conclusions we refer the reader to the original paper Waltman et al (2011b).

PSJR2 - Prestige SJR2 (Guerrero-Bote and Moya-Anegón 2012): PSJR2 was proposed as an improvement of the PSJR indicator proposed by González-Pereira et al (2010) and is a size-dependent metric. The indicator considers both the prestige and closeness of the citing journal. PSJR2 is also recursively calculated and its value depends again upon three terms. The first two terms are the same as in the original PSJR indicator. The difference between the two indicators lies in the calculations of the third term which represents the citation prestige. This term is dependent upon a set of coefficients, named $Coef_{ji}$ and a factor, named PSJR2D, in the calculations of which the cosine of the co-citation profiles of the journals are used. The co-citations received by two journals are used as a measure of their closeness and the introduction of the cosine of the cocitation vectors is used as a measure of the thematic relationship between journals. The PSJR2 indicator is given by

$$PSJR2_i = \left(\frac{1-d-e}{N}\right) + \left(e \cdot \frac{Art_i}{\sum_{j=1}^{N} Art_j}\right) + \left(\frac{d}{PSJR2D} \cdot \left[\sum_{j=1}^{N} Coef_{ji} \cdot PSJR2_j\right]\right) \tag{22}$$

For a detailed description of each element and for the equations used to calculate the PSJR2D factor and the $Coef_{ji}$ coefficients we refer the reader to Guerrero-Bote and Moya-Anegón (2012).

SJR2 indicator (Guerrero-Bote and Moya-Anegón 2012): The SJR2 indicator is a size-independent metric whose calculations are based on the PSJR2 indicator. It is calculated by dividing the PSJR2 value of a journal with the ratio of citable documents that each journal has relative to the total. The SJR2 indicator is given by

$$SJR2_i = \frac{PSJR2_i}{\left(Art_i / \sum_{j=1}^{N} Art_j\right)} = \frac{PSJR2_i}{Art_i} \cdot \sum_{j=1}^{N} Art_j \tag{23}$$

6 Classification of the indirect indicators

In this section, we attempt a classification of the indicators included in each category. The classifications derive from the definitions of the indicators and the factors considered in their calculations.

6.1 Paper-based indicators

All the discussed paper-based indicators consider the information present in the entire *Paper-Citation graph* and are either *independently defined* or defined as *modifications* *of the PageRank algorithm*. One may also categorize some of them based on whether they consider: (a) the *distance of citations from the current paper*, (b) the *age of the current paper*, (c) the *scientific field* and, (d) the *incompleteness of the Paper-Citation graph*. Table 9 presents the classification scheme described and lists the indirect indicators that belong to each class.

Relation to PageRank	Additional factors considered	Indicator
Independently defined	-	Gozinto theorem
	Distance of citations from current paper	SCEAS Rank
	-	Cumulative patent citations
	Distance of citations from current paper	Weighted cumulative patent citations
	Scientific field	f-value
Modifications of PageRank	Age of current paper	CiteRank
	-	PageRank
	-	P-Rank
	Incompleteness of the Paper-Citation graph	Prestige-Rank

Table 9 Classification of the paper-based indirect indicators

6.2 Author-based indicators

The indirect author-based indicators either use the information present in the *Author-Citation graph* (SARA and EigenFactor score for authors) or are based on information present in the *Paper-Citation graph* (Cross generational indices, hfg-index and the Indirect H-index). From the indicators examined, the EigenFactor score for authors uses the entire graph, whereas SARA makes use of the entire graph within the specified citations interval. The cross-generational indices may (or may not) utilize the entire graph depending on the number of generations examined, and the hfg- and indirect H-indices only examine two generations of citations. Table 10 presents the classification scheme described and lists the indirect indicators that belong to each class.

Type of graph	Use of graph	Indicator
Author-Citation	Complete within interval	SARA
	Complete	EigenFactor score for authors
Paper-Citation	Complete or Partial	cross generational indices
	Two generations	hfg-index
		Indirect H-index

Table 10 Classification of the author-based indirect indicators

6.3 Journal-based indicators

One can assign the indirect journal-based indicators based on their relation (if any) to PageRank in the following categories: (a) the ones that are *adaptations of PageRank*, (b) the ones that are *derived from indicators that use or adapt PageRank* and, finally, (c) the ones that are *independently defined*. Another possible categorization can be based on whether the indicators are dependent on the size of the journal in question or not. Table 11 presents the classification scheme described and lists the indirect indicators that belong to each class.

Relation to PageRank	Journal-size dependent	Indicator
PageRank adaptations	X	Weighted PageRank
	X	EigenFactor score
	X	Prestige SJR
	X	Prestige SJR2
Derived from indicators that use or adapt PageRank	X	Y-factor
	-	Article influence score
	-	SJR indicator
	-	SJR2 indicator
	X	PR_{sum}-Total authority
	-	PR_{ave}-Authority factor
Independently defined	-	Influence weight of the journal
	-	Influence per publication of the journal
	X	Total influence of the journal
	-	Recursive Mean Normalized Citation Score

Table 11 Classification of the journal-based indirect indicators

7 Discussion

In this paper an overview of the concepts of indirect impact and generations of citations was given. The *Paper-Citation graph* was defined along with the derived *Author-Citation* and *Journal-Citation graphs.* The procedure used to construct the derived graphs was presented in the form of three distinct steps: (a) Define the type of citation counting to be applied to the original *Paper-Citation graph* (Full citation counting or Fractional citation counting), (b) Define the type of normalization to be used in the calculations of edge weights in the intermediate *Author-Citation* (or *Journal-Citation*) graphs, and, (c) Define the type of citation counting to be applied to the resulting *Author-Citation* (or *Journal-Citation)* graph. The aforementioned procedure produces sixteen (or eight in the case of *Journal-Citation graphs*) distinct types of derived graphs that all share the same structure (nodes and edges) and differ only in the citation weights.

The concept of generations of citations was also discussed with the eight different types of generations of citations found in the literature described in detail. More specifically, the concepts discussed included (a) backward and forward generations for references and citations, respectively, (b) independent and restricted generations, where papers are included in the current generation of citations only if they have not been included in any prior generation or they are always included, and (c) unique and non-unique papers per generation, where a paper is included only once per generation or as many times as the paths originating from this paper towards the paper under scrutiny.

A number of indirect indicators for papers, authors and journals were listed and briefly explained. Finally, classification schemes for all categories of indirect indicators were proposed.

Acknowledgements We would like to thank the reviewers for their helpful and constructive comments.

References

Anderson TR, Hankin RKS, Killworth PD (2008) Beyond the Durfee square: enhancing the h-index to score total publication output. Scientometrics 76(3):577–588

Antonakis J, Lalive R (2008) Quantifying Scholarly Impact: IQp Versus the Hirsch h. Journal of the American Society for Information Science and Technology 59(6):956–969

Atallah G, Rodríguez G (2006) Indirect patent citations. Scientometrics 67:437–465

Batista PD, Campiteli MG, Kinouchi O (2006) Is it possible to compare researchers with different scientific interests? Scientometrics 68(1):179–189

Bergstrom CT (2007) Eigenfactor: Measuring the value and prestige of scholarly journals. C&RL News 68(5)

Bergstrom CT, West JD, Wiseman MA (2008) The Eigenfactor ^{TM}Metrics. The Journal of Neuroscience 28(45):11,433–11,434

Bollen J, Rodriguez MA, Van de Sompel H (2006) Journal status. Scientometrics 69(3):669–687

Cabrerizo F, Alonso S, Herrera-Viedma E, Herrera F (2010) q2-Index: Quantitative and qualitative evaluation based on the number and impact of papers in the Hirsch core. Journal of Informetrics 4(1):23–28

Costas R, Bordons M (2008) Is g-index better than h-index? An exploratory study at the individual level. Scientometrics 77(2):267–288

Dervos DA, Kalkanis T (2005) cc-IFF: A Cascading Citations Impact Factor Framework for the Automatic Ranking of Research Publications. In: Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, 2005. IDAACS 2005. IEEE, pp 668–673

Dervos DA, Samaras N, Evangelidis G, Folias T (2006) A New Framework for the Citation Indexing Paradigm. Proceedings of the American Society for Information Science and Technology 43(1):1–16

van Eck NJ, Waltman L (2008) Generalizing the h- and g-indices. Journal of Informetrics 2(4):263–271

Egghe L (2006) Theory and practise of the g-index. Scientometrics 69(1):131–152

Egghe L (2011a) The single publication H-index and the indirect H-index of a researcher. Scientometrics 88:1003–1004

Egghe L (2011b) The single publication H-index of papers in the Hirsch-core of a researcher and the indirect H-index. Scientometrics 89:727–739

eigenFACTORorg (2008) Eigenfactor ^{TM}Score and Article Influence ^{TM}Score: Detailed methods [version 2.01]. (Accessed via wwweigenfactororg/methodspdf)

Fragkiadaki E, Evangelidis G, Samaras N, Dervos DA (2011) f-Value: measuring an article's scientific impact. Scientometrics 86(3):671–686

Garfield E (1999) Journal impact factor: a brief review. Canadian Medical Association Journal 161(8):979–980

Garfield E (2005) The Agony and the Ecstasy— The History and Meaning of the Journal Impact Factor

González-Pereira B, Guerrero-Bote VP, Moya-Anegón F (2010) A new approach to the metric of journals' scientific prestige: The SJR indicator. Journal of Informetrics 4(3):379–391

Guerrero-Bote VP, Moya-Anegón F (2012) A further step forward in measuring journals' scientific prestige: The SJR2 indicator. Journal of Informetrics 6(4):674–688

Hirsch JE (2005) An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences of the United States of America 102(46):16,569–16,572

Hirsch JE (2007) Does the h-index have predictive power? Proceedings of the National Academy of Sciences of the United States of America 104(49):19,193–19,198

Hu X, Rousseau R, Chen J (2011) On the definition of forward and backward citation generations. Journal of Informetrics 5(1):27–36

Jin B, Liang L, Rousseau R, Egghe L (2007) The R- and AR-indices: complementing the h-index. Chinese Science Bulletin (52(6)):855–863

Kosmulski M (2006) A new Hirsch-type index saves time and works equally well as the original h-index. ISSI Newsletter pp 4–6

Kosmulski M (2010) Hirsch-type approach to the 2nd generation citations. Journal of Informetrics 4(3):257–264

Ma N, Guan J, Zhao Y (2008) Bringing PageRank to the citation analysis. Information Processing & Management 44(2):800–810

Maslov S, Redner S (2008) Promise and Pitfalls of Extending Google's PageRank Algorithm to Citation Networks. The Journal of Neuroscience 28(44):11,103–11,105

Page L, Brin S, Motwani R, Winograd T (1999) The PageRank Citation Ranking: Bringing Order to the Web. Technical Report 1999-66, Stanford InfoLab

Pinski G, Narin F (1976) Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics. Information Processing & Management 12(5):297–312

Radicchi F, Fortunato S, Markines B, Vespignani A (2009) Diffusion of scientific credits and the ranking of scientists. Phys Rev E 80:056,103

Radicchi F, Fortunato S, Vespignani A (2012) Citation Networks, Springer Berlin Heidelberg, pp 233–257. Understanding Complex Systems

Rousseau R (1987) The Gozinto theorem: Using citations to determine influences on a scientific publication. Scientometrics 11:217–229

Schubert A (2009) Using the h-index for assessing single publications. Scientometrics 78:559–565

Sidiropoulos A, Manolopoulos Y (2005) A citation-based system to assist prize awarding. SIGMOD Rec 34(4):54–60

Sidiropoulos A, Katsaros D, Manolopoulos Y (2007) Generalized Hirsch h-index for disclosing latent facts in citation networks. Scientometrics 72:253–280

Su C, Pan Y, Zhen Y, Ma Z, Yuan J, Guo H, Yu Z, Ma C, Wu Y (2011) PrestigeRank: A new evaluation method for papers and journals. Journal of Informetrics 5(1):1–13

Tol RS (2008) A rational, successive g-index applied to economics departments in Ireland. Journal of Informetrics 2(2):149–155

Walker D, Xie H, Yan KK, Maslov S (2007) Ranking scientific publications using a model of network traffic. Journal of Statistical Mechanics: Theory and Experiment 2007(06):P06,010

Waltman L, van Eck NJ (2012) The inconsistency of the h-index. Journal of the American Society for Information Science and Technology 63(2):406–415

Waltman L, van Eck NJ, van Leeuwen TN, Visser MS, van Raan AF (2011a) Towards a new crown indicator: Some theoretical considerations. Journal of Informetrics 5(1):37–47

Waltman L, Yan E, van Eck N (2011b) A recursive field-normalized bibliometric performance indicator: an application to the field of library and information science. Scientometrics 89:301–314

West JD, Jensen MC, Dandrea RJ, Gordon GJ, Bergstrom CT (2013) Author-level Eigenfactor metrics: Evaluating the influence of authors, institutions, and countries within the social science research network community. Journal of the American Society for Information Science and Technology 64(4):787–801

Wu Q (2010) The w-index: A measure to assess scientific impact by focusing on widely cited papers. Journal of the American Society for Information Science and Technology 61(3):609–614

Yan E, Ding Y, Sugimoto CR (2011) P-Rank: An indicator measuring prestige in heterogeneous scholarly networks. Journal of the American Society for Information Science and Technology 62(3):467–477