BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES, Vol. 70(4), 2022, Article number: e141988
DOI: 10.24425/bpasts.2022.141988

Fuzzy logic in knowledge dissemination due to citation trees. Contribution to discipline vector

Aleksandra MRELA[1]*, **Veslava OSIŃSKA**[2], and **Oleksandr SOKOLOV**[3]

[1] Institute of Informatics, Kazimierz Wielki University in Bydgoszcz, Kopernika 1, 85-074 Bydgoszcz, Poland
[2] Institute of Information and Communication Research, Nicolaus Copernicus University in Torun, W. Bojarskiego 1, 87-100 Torun, Poland
[3] Department of Informatics, Faculty of Physics, Astronomy and Informatics, Nicolaus Copernicus University in Toruń, Grudziądzka 5, 87-100 Toruń, Poland

Abstract. Large sets of articles are evaluated by predefined measures such as the article numbers and h-indexes. All of these indicators are scalars and refer rather to one discipline or the comprehensive science. Thus, according to disciplinary categories in scientific databases, the distribution has become too rigid for current science needs, dynamically growing towards inter- and trans-disciplinarity. We propose a new method of calculating the impact on knowledge of articles and their citations, creating citation networks, and using one of the optimistic fuzzy aggregation norms to estimate the contribution to the knowledge considering the citation inheritance of citing papers to cited papers (paper-children to the paper-parents). Due to this method, we produced the contribution vectors for various disciplines/subdisciplines based on articles and their citations of publications belonging to the considered disciplines. We can prepare the scientific profiles of papers and disciplines based on the contribution vectors. Moreover, we can evaluate how much citations matter in the development of science. Applying this method, we can estimate the contribution to the considered research field caused by papers and their citations from different areas of science. The proposed method might be applicable in the assessment of developing concepts.

Key words: knowledge dissemination; citation analysis; fuzzy aggregation norm; contribution to discipline.

1. INTRODUCTION

The dissemination scope of scientific ideas is assumed to be measured by citations of the original publication – novel conceptions sources and carriers. Besides citation numbers, another measure such as the h-index has become an essential parameter for measuring scholarly impact since the beginning of the XXI century. The h-index can be applied to evaluate the productivity and citation impact of a group of scientists employed at one department, university, or even country. Many datasets classify articles to scientific disciplines, and a few bibliometric and scientometric measures have been developed based on the categorical membership of scientific articles.

Scientometricians try to estimate the scientists' achievement by applying selected measures, while others study their usefulness to describe the development of disciplines or research areas. Since the tremendous scientific accomplishments appeared suddenly and revolutionary [1], the proposed measures could notice them only after some time by analyzing their citations and the numbers of people following their methods. Moreover, when authorities of research institutes choose measures to evaluate the attainments, this causes the researchers to try to reach high levels of these measures and sometimes may lead some scientists to unethical behavior [2].

In the current paper, we consider the evolution of a scientific field displayed by the publications appearing in the area and endowed by the corresponding references with focus on understanding the overall knowledge dissemination process. As is well-known, the distribution of publications according to disciplinary categories in appropriate databases poorly reflects current scholars' needs, which are dynamically growing towards inter-, trans- and multidisciplinarity [3, 4]. The two global bibliographic databases, Web of Science (WoS) and Scopus use a general classification system of the journals indexed. WoS categories reveal just the higher-level systematic domains, but the Scopus classification exposes the research areas and research disciplines. Similarly, the national and local databases function by grouping the journals into one or two levels of hierarchy. However, these distributions are not sufficiently informative to understand the knowledge being disseminated and the prediction of its development.

This paper proposes a new method for constructing and evaluating citation trees. The approach is based on a novel fuzzy logic attitude constructing a diagram for each given paper with follow-up encapsulation into a discipline general distribution. In recursive tree building, fuzzy aggregation norms are applied and confronted with the other currently used bibliometric measures. A comparative study defines the essential differences in the progress of the scientific field and might therefore be used to predict potential changes and tendencies in the development of science. The hierarchical citation inheritance from article-parent to articles-children is taken into account employing a

*e-mail: a.mrela@ukw.edu.pl

Manuscript submitted 2021-11-14, revised 2022-06-03, initially accepted for publication 2022-06-28, published in August 2022.

model using citation trees. Assuming that both the publication of the article and its citations add some small unit of knowledge, we propose to aggregate all of these units of knowledge with one of the fuzzy aggregation norms along with the citation trees concerning the research fields receiving the vectors of the contribution of this publication to the considered research fields. Based on these vectors, using the fuzzy aggregation norm, we estimated the contribution-to-knowledge vectors, the components of which are contributions to the considered research fields. Through the application of these vectors, we can also include the contribution to knowledge of the citations of publications not indexed to the chosen research field.

2. STATE OF THE ART

The majority of scholars' evaluation metrics are based on an analysis of the citations received by scientific publications. Bibliographic databases, including reference data and primarily global indexes, such as WoS, Scopus, or Google Scholar, are the essential source for this study. Citation-based indicators (for example, journal impact factor or h-index) may provide information about the impact of an individual publication and publications corpora.

One of the critical subjects in scientific knowledge studies consists of evaluating a possible influence of the information derived from the bibliography or research characteristics on citation "output" [5]. In this connection, co-authors' attitudes can normalize citation occurrences to obtain a general view on scholarly output like conditions, collaborations, or funding. Citations can also serve as a multi-perspective vision on the knowledge transfer in science and technology. Additionally, due to significant differences in the citation count in various disciplines, a problem of "a limited understanding of how disciplinary knowledge is used and diffused" is still perceptible (see also [6]).

The mentioned paper [5] splits the fundamental citation indicators into size-dependent and size-independent. The first group includes three matters: the total number of citations, the highly cited publications number, and h-index of a scientist or scholar's publication. Obviously, these measures do not decrease with the addition of the following publication. The second group highly depends on dataset modification because it consists of factors like the average citations per publication and the proportion of highly cited publications. These factors facilitate comparative analysis between a small and an extensive research group or a small and prominent institution after an appropriate normalisation. Bibliometricians use the first group to estimate the scientific achievements of researchers. However, the h-index may be modified, e.g., the departmental h-index, to evaluate groups of researchers' activities [7]. Lazaridis [8] uses the averages of the academics' h-indexes to express university department ranks.

Lipitakis *et al.* [9] suggest citation rates (CR) of research areas (yearly averages of citations per paper) during 2008-2014, concluding that in 2012, CR of Molecular Biology and Genetics articles possess the highest (24.29) value. In comparison, the CR of Math publications is significantly smaller (4.16). Aiming to make it possible to compare different disciplines, the WoS database developer proposed in 2018 Category Expected Citations (CEC) as the "average cites to items of the same document type (article), year, and category" [10]. This factor used for extensive and longitudinal bibliographic data analysis is involved in micro and macro studies based on WoS data (Expected citation rates, half-life, and impact ratio, 1994).

Citations characteristics of a particular discipline or research field are served by InCites Benchmarking and Analytics [11], that is an analytics module demonstrating a significant discrepancy in normalized citation values. A few publications are highly cited, but most of them (poorly cited) compose a long tail in the citation distribution. Researchers also report the lack of stability of normalized citation indicators on aggregation level of classification, such as, for example, well-known WoS categories [5, 12, 13].

A new evaluation concerning mainly paper reassignments according to the multidisciplinary thematic is given in [3, 12]. Milojević [4] proposes reclassifying WoS articles based on the references and adapting them towards citations. Another direction uses matching citation normalisation metrics to classification. The idea was to examine whether some metrics (average, median) are stabilized during the articles' diminishing clusters. Kostoff and Martinez [14] showed that "the citation characteristics became increasingly stratified," and normalisation studies start losing their credibility.

Thus, the connection between citation information and publication classification was not used sufficiently during the science macrostructure study, particularly in multi- and interdisciplinary research. The reason may lie in divergent levels of generated information: more top or more down of documents organization-level making these data hardly compatible. Unlike thematic study or paper classification, most citation models rely more on data about future publications than other scientometric research methods. As was mentioned earlier in this paper, a fuzzy logic-based approach is applied. A WoS query ("scientometric" or "bibliometric") AND "fuzzy" returns 123 records. Simultaneously, the "Result Analysis" module of WoS demonstrates a continuous growth in the number of publications using fuzzy methods [15], from 3 in 2011 to 38 in 2020. Most of these papers relate to bibliometric studies in the fuzzy sets algorithms (see e.g., "Bibliometric Analysis of Fuzzy Logic Research" ([16]). The fuzzy methods serve to analyze citation networks presenting the relations between information sciences and fuzzy systems [17, 18].

3. MATERIALS AND METHODS

3.1. Optimistic fuzzy aggregation norms

Fuzzy logic is a kind of the general many-valued logic allowing the truth value to take values in interval $[0,1]$ in contrast to classical Boolean logic where this value can be just 0 or 1. Lotfi Zadeh [19] introduced this term in 1965 within the proposed fuzzy set theory. Let X be the universal suite (a set under consideration). A fuzzy set A given on X is conveyed as ordered pairs $(x, \mu_A(x))$, where $\mu_A(x)\colon X \to [0,1]$ is a membership function. The value $\mu_A(x)$ embodies the degree to which x belongs to A. Fuzzy sets and relations (fuzzy sets in a Cartesian product of spaces) are applied in many scientific disciplines and practical applications. For example, this idea is used to search for tools for assessing some quantities levels in social sciences. For instance, we can use it to analyze levels of learning outcomes [20] or design tools for the assessment of gait problems [21].

When a set consists of several values, one of the aggregation functions might be used to calculate one summary value. However, when the impact of a chosen article is estimated, the more significant the number of papers citing the chosen one, the higher the impact on the knowledge or discipline. Let $\mathcal{A}$ be a set of papers. For each paper A from $\mathcal{A}$, we assign the membership function indicating the level of contribution of A to the chosen discipline. Thus, we have looked for an aggregation function with a unique property such that the result is more significant than the arguments indicating the level of contribution of the set of papers to the discipline. We have decided to use optimistic fuzzy aggregation norms [22] since we assume that each new paper and each citation add value to scientific fields.

Let us introduce optimistic fuzzy aggregation norms. The function $S : [0, 1] \times [0, 1] \to [0, 1]$ is called an optimistic fuzzy aggregation norm if, for each $x, y \in [0, 1]$, it fulfills the following conditions:

- **(S1)** $S(0, 0) = 0$ (border condition),
- **(S2)** $S(x, y) = S(y, x)$ (commutativity),
- **(S3)** $S(x, y) > \max\{x, y\}$ if $x, y \notin \{0, 1\}$ and $S(x, y) \geq \max\{x, y\}$ (optimism).

We will use such functions to study the knowledge dispersion between scientific fields, considering the publications and their citations. From this standpoint, the variables x, y are indeed values of membership functions representing the development of the scientific field or the estimation of the importance of the publications and their citations to this scientific field's growth. According to our optimistic suggestion, a function intended for aggregation must return a value more significant than both x and y, presuming that more citations of an article in any field increase the contribution to the applicable discipline. We will apply these functions to aggregate values of a membership function to calculate the impact of articles on the scientific field and determine the impact of one paper considering its citations. There are many functions fulfilling conditions **(S1)–S(3)**, for example, $S(0,0) = 0$ and $S(x,y) = \ln\left((e-1)\cdot\min\left\{\sqrt{x^2+y^2}, \frac{\min\{x,y\}}{\max\{x,y\}}\right\}+1\right)$ for $0 < x \leq 1$ and $0 < y \leq 1$. However, because of simplicity, we prefer to deal with the function

$$S(x,y) = x + y - xy \tag{1}$$

for $x, y \in [0, 1]$ to aggregate values. This function admits a natural interpretation connected to its probabilistic meaning as the "probability" of unity of two independent events. On the other hand, it is simple and can be efficiently implemented to create fuzzy citation trees.

3.2. Impact on the scientific field caused by affiliation values

Let $A \in \mathcal{A}$, where $\mathcal{A}$ denotes a space of papers from a chosen database, and A is a paper. Let F represent the scientific field. Article A can add importance to F in two ways - by its publication and citations - so let the membership function indicating the impact of a paper assigned to F at the time of its publication be called the affiliation function. Its values are denoted by $Af(A,F)$.

Let $a \in (0, 1)$ be a small value representing the impact of one published article. Moreover, we assume that at the time of publication, the membership value of each article is the same, and is equal to a. Hence, we can apply the optimistic fuzzy aggregation norm to calculate the impact of two or more articles to F. Let $Af(F)$ denote the value of the membership function representing the impact on scientific field F caused by the publication of some articles assigned to F.

Consider the example, let $a = 0.01$ and assume that in one year, three articles A_1, A_2, and A_3, assigned to scientific field F are published, so $Af(A_i,F) = a$ for $i = 1,2,3$. Let $Af(F) = 0$. Thus, using (1), we can calculate the impact of articles A_1, A_2, and A_3 on F. Hence, after the publication of A_1, we have $Af(F) = S(Af(F), Af(A_1,F)) = S(0,a) = 0.01$. Next, after publishing paper A_2, $Af(F) = S(Af(F), Af(A_2,F)) = 0.01999$. Similarly, we can show that the impact on F calculated after the publication of the considered three papers equals $Af(F) = S(Af(F), Af(A_3,F)) = 0.02997$. Thus, assuming affiliation values for the set of articles assigned to considered scientific field F, we can estimate their impact on F. Because of the properties of function S, $Af(F) \in [0,1]$. Moreover, when the following paper is published, $Af(F)$ will be more significant.

3.3. Contribution to scientific field caused by papers and their citations

Next, we can apply function S to estimate the impact to F caused by publication paper A and its citations.

Let $A, A_1, \ldots, A_n \in \mathcal{A}$ be papers assigned to F such that $A_1, \ldots, A_n$ cite A. The membership function indicating the impact of an article and its citations to F be called the contribution function. Its values are denoted by $C(A,F)$. At the time of publication, the following relationship is fulfilled – $C(A,F) = Af(A,F)$. Let the depth d_i $(i = 1,2,\ldots,n)$ be defined as follows $d_i = 1$ if A_i is published in the same period as A and $d_i = k+1$ if A_i is published k periods after A. Let us assume that each citation adds to F the value a; however, the more years lapse between papers A and A_i being published, the less significant the contribution to F of A caused by the citation of A_i. Thus, we can propose the following formula

$$C(A,F) = S\left(C(A,F), \frac{a}{d_i}\right) \quad \text{for } i = 1,2,\ldots,n. \tag{2}$$

Consider an example: assume that three papers cite A, the first one in the following year, the second one – after three years, and the last one – after five years. In the year of publication, the contribution of A to F equals $C(A,F) = Af(A,F) =$

0.01. After one year, paper A_1 cites A, so $d_1 = 2$ and, by (2), we have $C(A,F) = S\left(C(A,F), \frac{a}{d_1}\right) = 0.014995$. After three years, article A_2 is published, so $d_2 = 4$ and $C(A,F) = S\left(C(A,F), \frac{a}{d_2}\right) = 0.02498$. After five years, article A_3 is published, so $d_3 = 6$, and $C(A,F) = S\left(C(A,F), \frac{a}{d_3}\right) = 0.0266425$. Thus, we can apply the optimistic fuzzy aggregation norm to find the contribution to F caused by a paper and its citations considering the number of periods between the publication and its citations.

3.4. The citation tree

We have considered two simple situations where we consider only the affiliation values of articles or one article with several citations. However, the citation tree, including a paper and its citations, may be much more complicated.

Let us recall some definitions connected with trees, which, in the computer sciences, are an important way of storing and presenting hierarchically ordered non-linear data (comp. Fig. 1 and Fig. 2). Tree components are named after natural tree elements: roots, branches, and leaves; however, the root is placed at the top of the tree. A node stores a component of data, and it consists of two parts: one is the name of the stored data and the other consists of links (edges) to other nodes. The node where the tree starts is called the root. The nodes placed below others in the tree are children; the ones that are above the others are called parents. The nodes that have no children are leaves. You can walk the tree only from parents to children; these ways are called paths. The number of nodes (excluding the considered node) on the path is called the depth.

Let us choose one paper as the root of a tree as it has been cited by a few articles at different depths. The set of second-level publications cites this first-level article. The longer the period we choose, the more complicated the tree becomes. Moreover, papers from different scientific fields can cite articles from our tree, so citations facilitate the data aggregation in the context of knowledge growth. Fuzzy logic methods enable and simplify calculations of the contribution to disciplines. Thanks to the normalisation of the fuzzy aggregation norms, the results belong to interval $[0, 1]$.

Now let $\mathcal{A}$ be a space of articles that are assigned to one of the predefined n scientific fields F_i $(i = 1,2,\dots,n)$. Hence, for each article $A \in \mathcal{A}$, we assign a vector of length n such that its components are contribution values to scientific fields F_i. Assume that $F = \{F_i, i = 1,2,\dots,n\}$ is a discipline. Then $C(A,F) = [C(A,F_1), C(A,F_2),\dots,C(A,F_n)]$ is called a contribution vector of article A to discipline F. Moreover, let $C(F)$ denote the contribution vector of discipline F to the knowledge. Thus, if root $A \in \mathcal{A}$ is a cited article, we can build the citation tree. For each child and finally the root, we calculate the contribution vectors as follows: For each node (and the root as well), we assign the affiliation vector such that $[Af(A,F_1), Af(A,F_2),\dots,Af(A,F_n)]$, where $Af(A,F_i) = a$ if an article is set to scientific field F_i and 0 otherwise $(i = 1,2,\dots,n)$. For all leaves, the contribution vectors and affiliation vectors are equal. For all parents, we use the optimistic fuzzy aggregation norm to increase the contribution to all of the scientific fields of the parent. Let us consider two cases:

- the child is a leaf on the depth d, then

$$C(parent, R_i) = \begin{cases} S(a, C(child, F_i)/d) & \text{if parent is assigned to } F_i, \\ S(0, C(child, F_i)/d) & \text{otherwise} \end{cases}$$

  hence, $C(parent, F_i) = S(Af(parent, F_i), C(child, F_i)/d)$,
- the parent (P) has got K children $\{ch_1, ch_2, \dots, ch_K\}$ located on the depths $d_1, d_2, \dots, d_K$, respectively, then

$$C(P,F_i) = S\left(\dots S\left(S\left(Af(P,F_i), \frac{C(ch_1,F_i)}{d_1}\right), \frac{C(ch_2,F_i)}{d_2}\right), \dots, \frac{C(ch_K,F_i)}{d_K}\right).$$

Let us consider the following example. The database assigns each article to one of the two scientific fields F_1 and F_2 (represented in Fig. 1 as grey and white colours). Paper A_1 (the root of the tree) was published in year X and was classified as F_1. Assume that in the next year $(X+1)$, two papers citing A_1 were published: A_2 assigned to F_1 and A_3 classified to F_2. In the following year $(X+2)$, three articles were published: A_4 citing A_2 assigned to F_1; A_5 citing A_3 classified to F_2 and A_6 citing A assigned to F_2.

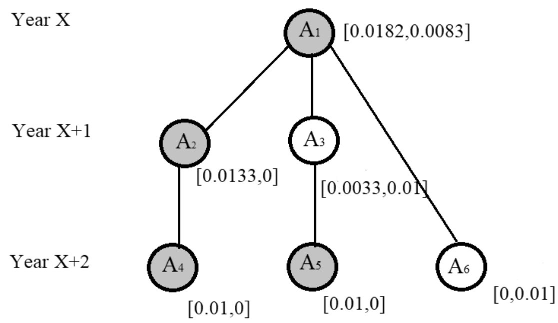

Fig. 1. The citation tree of article A_1 with contribution vectors of articles

At the beginning, for each node, we assign the affiliation vectors. Leaves represent articles A_4, A_5 and A_6. Because they are not cited, their contribution vectors are equal to the affiliation vectors. Thus, $C(A_4,F) = [C(A_4,F_1), C(A_4,F_2)] = [Af(A_4,F_1), Af(A_4,F_2)] = [0.01,0], C(A_5,F) = [0.01,0]$ and $C(A_6,F) = [0,0.01]$. Next, for article A_2, we can notice that this article together with A_4 are assigned to the same scientific field F_1, so $C(A_2,F_1) = S(0.01, Af(A_4,F_1)/3) = S(0.01,0.00333) = 0.0133$ and $C(A_2,F_2) = S(0, C(A_4,F_2)/3) = S(0,0) = 0$. Hence, $C(A_2,F) = [0.0133,0]$. For article A_3, we observe that this article together with A_5 are assigned to different research areas. Thus, so $C(A_3,F_1) = 0.0033$ and $C(A_3,F_2) = 0.01$. Hence, $C(A_3,F) = [0.0033,0.01]$. Finally, we calculate the contribution of the root. The contribution is increased due to the inclusion of article A_2 and its citations, we have $C(A_1,F_1) = S(0.01, C(A_2,F_1)/2) = S(0.01, 0.0133/2) = 0.01658$ and $C(A_1,F_2) = S(0, C(A_2,F_2)/2) = S(0,0) = 0$. Next, we include the citation subtree of A_3, hence $C(A_1,F_1) = S(0.01658, C(A_3,F_1)/2) = S(0.01658, 0.0033/2) = 0.0182$ and $C(A_1,F_2) = S(0, C(A_3,F_2)/2) = S(0, 0.01/2) = 0.005$. Finally, we include article A_6. Thus, $C(A_1,F_1) = 0.0182$ and $C(A_1,F_2) = 0.0083$. Hence, $C(A_1,F) = [0.0182, 0.0083]$. Let us consider the next example (Fig. 2). The structure of this citation tree is different than the previous one. In this case all articles (apart from the root A_1) cite directly A_1 and they are all leaves. Thus, $C(A_2,F) = C(A_4,F) = C(A_5,F) = [0.01,0]$ and $C(A_3,F) = C(A_6,F) = [0,0.01]$. After the similar calculation, we get $C(A_1,F) = [0.02477, 0.00998]$.

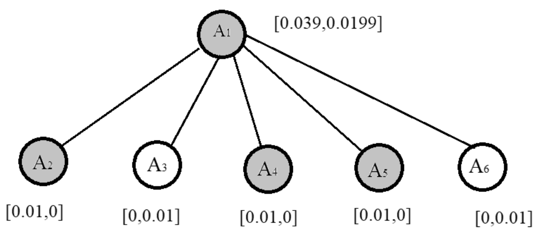

Fig. 2. The citation tree of article A_1 with contribution vectors of articles

Let us consider the next example (Fig. 2). The structure of this citation tree is different than the previous one. In this case all articles (apart from the root A_1) cite directly A_1 and they are all leaves. Thus, $C(A_2,F) = C(A_4,F) = C(A_5,F) = [0.01,0]$ and $C(A_3,F) = C(A_6,F) = [0,0.01]$. After the similar calculation, we get $C(A_1,F) = [0.02477, 0.00998]$.

Summing up, we can see that different structures of the citation trees cause distinct contribution vectors. Using the optimistic fuzzy aggregation norms, we can calculate the vectors of the article contribution to all considered scientific fields (the root of a tree and all items in the tree). Because of the properties of the S operation, the contribution values belong to interval $[0,1]$, and we can observe the increasing contribution to scientific fields as the citation tree is growing. Moreover, we can compare the results for different articles (and their citation trees).

3.5. Description of data

The presented method of evaluating researchers' scientific achievements is based on the data retrieved from the CORA dataset [23]. The relational dataset consists of 3 tables: cites, papers and content. The tables include fields describing the main bibliographic entities and such as:

- cites: cited_paper_id, citing_paper_id,
- paper: paper_id, class_label,
- content: paper_id, word_cited_id.

As we can see, the $paper_id$ field serves as both the primary and secondary keys linking the tables. The CORA consists of 1446 scientific publications published in 1987–1997 and 4062 direct citations. All of the articles are classified into one of seven scientific fields of computer science:

1. Case-Based (CB),
2. Genetic Algorithms (AL),
3. Neural Networks (NN),
4. Probabilistic Methods (PM),
5. Reinforcement Learning (RFL),
6. Rule Learning (RL),
7. Theory (Th).

A simple database structure is convenient to present the citation network of 1446 nodes and 4062 edges. Sufficiently large numbers of nodes and edges allow complex patterns of citation links inside and outside of each category to be explored. The citation relationship of the CORA dataset is shown in Fig. 3 (left).

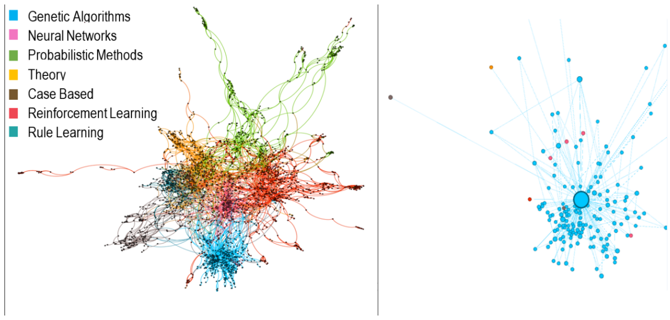

Fig. 3. Citation network of the CORA articles (left). The citation network of article ID = 35 as a root (right)

An article is represented by a node, which is scaled by the primary measure used in network analysis (NA), such as eigenvector centrality. The colour relates to the appropriate scientific field. Regarding colour overloading on the spectacular graph using the spring layout in Fig. 3 (right), we can observe that the citations between different disciplines occur in the central network area. The distribution of degrees is characterised by a long tail that follows a power law. Therefore we can conclude the CORA belongs to a free-scale network.

Figure 4 shows how the distribution of articles published within the considered scientific areas changes over time. As we can see, most papers are assigned to Neural Networks (total 413) during the whole considered period. The number of articles assigned to other computer science subfields was changing;

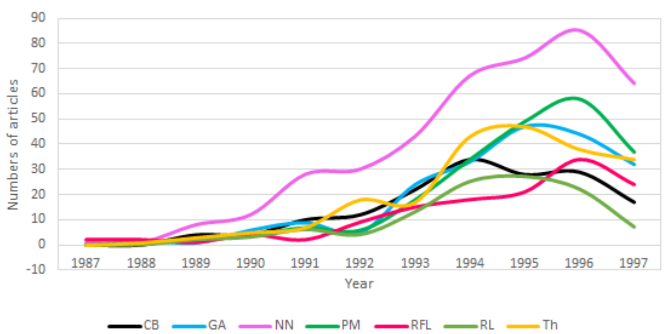

Fig. 4. The growth of numbers of articles categorized into seven subfields during the years 1987–1997

different years show varying numbers of publications. Interestingly, the maximum quantity of articles occurs around 1995–1996 in all discussed scientific fields. The year 1997 is the last year of our research period, and usually the finishing year data is underrepresented, and we can observe this decline in the number of published articles.

However, in our study, we separately analysed the citation trees of each article instead of the whole network. We specified several papers and their connected nodes for convenient analysis, but the tree of only one paper is presented in this article due to space limitations.

An NA application such as Gephi will run the Ego network filter. Figure 3 (right) illustrates the selected node Ego network (ID = 35), which was considered a root as it was cited by papers at all depths. The item and its children are assigned mainly to Genetic Algorithms (Fig. 3 (right), blue circles). However, some citing articles are allocated to Reinforcement Learning, Theory, and Case-based subfields, and they appear on the deepest levels in the citation network (Fig. 3 (right), different colours).

For each article A, we assign the vector $V(A,D) = [CB,GA,NN,PM,RFL,RL,Th]$ such that its components indicate the class affiliation at the time of publication and D denotes computer science. For example, if article A is assigned to the Genetic Algorithms class, $V(A,D) = [0,1,0,0,0,0,0]$. In the case of the CORA dataset, we can aggregate, for example, the number of articles belonging to classes or the number of citations of these articles.

The paper aimed to compare the initial distribution of the articles from the CORA dataset with the distribution based on citation information. For this purpose, we built trees of citations for each article. The trees of articles were built as follows. The tree root is the current article, the nodes of the first level of the tree are articles citing the present article, the second level nodes are built from the papers citing the papers from the first level so on, as described above. The leaves of the tree are the articles that are not cited anywhere. Let us consider one of the trees (Fig. 5). The root of this tree is the following paper: T.R. Martinez, "Models of Parallel Adaptive Logic," [24].

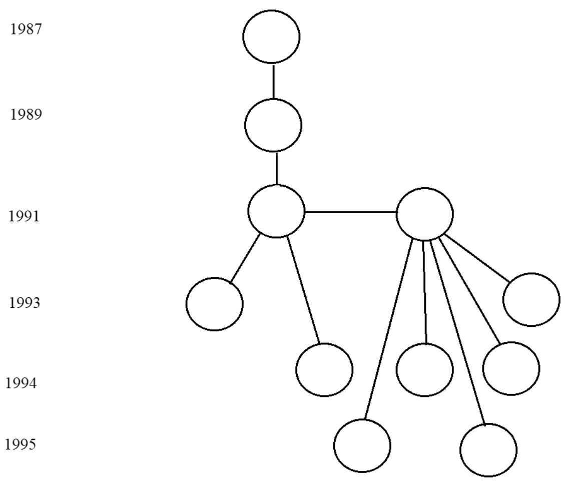

Fig. 5. Example of a tree of citations of T.R. Martinez, "Models of Parallel Adaptive Logic" [24]

This tree presents the hierarchy of papers that have cited the root and then the nodes indexed by the CORA database during 1987–1995. The tree shows that taking into account only the first level of citations when calculating the contribution to computer science subfields diminishes the true scientific impact of the root article.

4. RESULTS

4.1. The dynamic scientific profile of a paper

Let us apply the concept of contributions to discipline/research areas to prepare the dynamic scientific profiles of a chosen paper or a research field. For each paper from the considered set of articles, we can prepare its dynamic scientific portrait. For each year, starting from the year of publication, we can describe the vector of the contributions to the development of subfields of computer science. Thus let $C(article,D) = [C(CB),C(GA),C(NN),C(PM),C(RL),C(RL),C(Th)]$ where D denotes computer science, $C(article,D)$ – the vector of contributions to computer science of the article, $C(scientificfield)$ – the contribution to the considered field of computer science.

We can prepare the vector of the article contribution to the considered scientific fields for the given period. We can also calculate the set of these vectors for each year from the considered period. In the second case, we call this set of contributions a dynamic scientific profile of an article. Suppose paper A is assigned to scientific field P at the time of publication. In that case, the vital thing that we can see while observing the dynamic scientific profile of this paper is the possibility that it can also contribute to subdisciplines other than P.

As an example, we prepared the dynamic scientific portrait of the following publication during 1989–1997: Goldberg D.E., Genetic algorithms in search, optimization, and machine learning, [25]. In the CORA dataset, this book was assigned to Genetic Algorithms. According to our calculations, this book contribution has increased in Genetic Algorithms (leading scientific field), Reinforcement Learning, and later in Neural Networks and Case-Based (comp. Table 1).

Table 1
The values of the contribution of D.E. Goldberg's book to the subfields of computer science

Class	CB	GA	NN	PM	RLF	RL	Th
1989	0	0.269	0	0	0.066	0	0
1990	0	0.681	0	0	0.339	0	0
1991	0	0.893	0	0	0.423	0	0
1992	0	0.93	0	0	0.533	0	0
1993	0	0.993	0	0	0.822	0	0
1994	0.149	1	0.105	0	0.917	0	0
1995	0.276	1	0.17	0	0.965	0	0
1996	0.312	1	0.251	0	0.988	0	0
1997	0.351	1	0.251	0	0.994	0	0

As can be observed (Table 1), we can monitor its contribution to various scientific fields over time based on the paper citations. Hence, we can see that this book has influenced the development of Genetic Algorithms; however, it has had an impact on the other subdisciplines as well. The dynamic portrait gives us a better image of the paper's influence during the several years on its contribution to computer science.

The vectors of the article's contribution to science help us observe its impact on different scientific fields over time. Considering only the scientific field to which the paper is assigned during publication gives us an inadequate, non-dynamic, less compound picture of this publication's impact on the development of the discipline.

4.2. Dynamics of changes in scientific achievements over time

The vector $C(D) = [C(CB), C(GA), C(NN), \ldots, C(RL), C(Th)]$ representing the contribution to computer science can be calculated with the application of aggregation norm S based on the papers' citations in the following way. After preparing the citation tree with the contribution vectors of each article (nodes of the tree), we aggregate the contributions of all articles assigned to the computer science scientific fields with formula (2). As in the case of a paper, we can prepare a scientific field's dynamic profile (Table 2).

Table 2
The contribution to computer science its scientific fields based on paper citations

Class	CB	GA	NN	PM	RLF	RL	Th
1987	0	0	0.1	0	0.1	0	0
1988	0	0.19	0.1	0	0.466	0	0
1989	0.231	0.408	0.546	0	0.501	0	0.1
1990	0.511	0.981	0.838	0.1	0.973	0.307	0.342
1991	0.843	0.9998	0.9996	0.19	0.995	0.307	0.678
1992	0.897	1	1	0.467	1	0.377	0.995
1993	0.998	1	1	0.913	1	0.887	1
1994	1	1	1	1	1	1	1
1995	1	1	1	1	1	1	1
1996	1	1	1	1	1	1	1
1997	1	1	1	1	1	1	1

All dynamic profiles of computer science subfields show increasing contributions to science; however, with different ratios. Neural Networks impact on science is the most rapid during this period, but its domination is not significant enough compared to the growth in the number of articles.

5. DISCUSSION

5.1. Normalisation

In bibliometric research, normalising most of the measures related to cumulative knowledge output is assumed. Thus, to evaluate the scholarly impact, they use the annual number of publications to calculate the growth rate, the yearly number of citations to get the citation impact, and the average number of documents per journal to estimate its influence. Normalisation at the current stage and time, i.e., division by the maximum value of the measured variables, in most cases operates retrospectively because we do not know future numbers of articles. Therefore, normalisation used at any point in time evokes a redistribution of the ratio between the components and changes the disciplinary relations at the aggregation level. Table 3 presents the example (sample, hypothetical) numbers of articles for which normalised and affiliation values were calculated based on the number of articles. In this period, the total number of published articles was 61, so we normalised all papers with this number. The affiliation vectors were calculated using (1).

Table 3
Normalized and affiliation values for sample number of articles

Number of articles	0	1	2	3	4
Normalized values	0	0.016	0.033	0.049	0.066
Affiliation values	0	0.01	0.02	0.03	0.039
Number of articles	5	6	7	8	9
Normalized values	0.082	0.098	0.115	0.131	0.148
Affiliation values	0.049	0.059	0.068	0.077	0.086

The following graph (Fig. 6) presents the normalised numbers of articles and affiliation values of the discipline. As we can observe, the impact on the knowledge is comparable; however, we have not added the impact caused by citations.

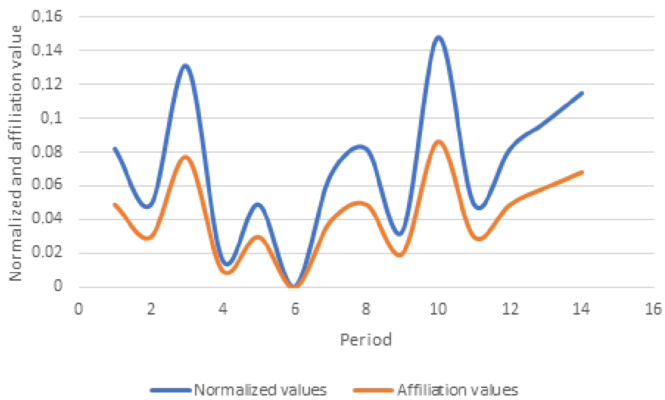

Fig. 6. Normalized values of numbers of articles and affiliation values of the impact on the discipline

In the case of normalisation, after choosing the set of citing articles of one paper, we can calculate the citation count CT(d) [26]; the primary measure of scientific impact. However, it is hard to use it for comparison studies because they are unpredictable, jumping, and irregular [5]. When we calculate the citation count, the results are the same if we consider two trees with the same number of papers but with different structures. Moreover, comparing two trees needs normalisation, which is not accessible if we are dealing with dynamic situations and following the new citation papers.

Therefore, instead of citation count, we propose building citation trees of articles to calculate their contribution to scientific fields producing contribution vectors. Each component represents the contribution of a root of the citation tree to the chosen scientific fields. This method also meets the requirement that the contribution of this article to the knowledge increases as the number of articles citing it grows, which is the specificity of scientific achievements. Furthermore, the impact of a paper citing the root depends on the number of years between the root and the paper-child being published. The more years that have passed between the publications of both articles; the less significant the child's contribution to the science of the root.

5.2. Category expected citation

To better understand the CORA citation data and their distribution by scientific fields, we refer to the scientometrically recognised indicators mentioned in Section 2. One of them is the Category Expected Citations (CEC), the number of citations of articles published in the same subject categories and the same year. According to Clarivate [27], the CEC measure was designed to compare articles' citation data with the average citation for similar publications in the same research field/journal/database in the same year. Let us notice that the CEC is a scalar whereas we received vectors, the components of which estimate the contribution of the considered paper to the discussed scientific fields. Table 4 shows the CECs for computer science scientific subfields in the years 1987–1997. This parameter indicates the following shortcomings. When there are more articles without citation in the considered period, the Category Expected Citations diminishes and might suggest that the contribution to these research fields knowledge also decreases. Moreover, when an article with many citations is published, it might stand out and distort the picture (in 1989 in GA, one publication was published with 165 citations during this period).

We can analyse the dynamics of development of scientific fields based on the CEC coefficient. In this case, when a new idea arises, this is reflected by the high growth of articles describing it and the rapid growth of citations of these articles. The citations do not keep up with the number of papers. But only a few articles will receive the greatest number of citations related to a new idea; the most widely known pioneer works directly reporting the discovery [28]. However, the vast majority of papers are not cited, which is commonly explained by the Pareto distribution (80% with no citations).

Nonetheless, it can be expected that each publication adds some value to the development of science. So the use of the cumulation method for each article and its citations is justified. Of course, some papers are cited after a few years, and we can observe some shorter or longer delays. However, generally citations "favour" the revelatory or fashionable articles (Matthew effect). We can see similar situation on the CEC plot (Fig. 7). Only one paper receives a vast number of citations (165), requiring the log scale to be used. According to subfields, comparing the dynamics of articles contributions (Table 2) and Category Expected Citations (Table 4), we can observe that CEC gradually decreases whereas the contributions to science are not decreasing values. Hence, the values of the contributions to science over ten years might reach some point of balance.

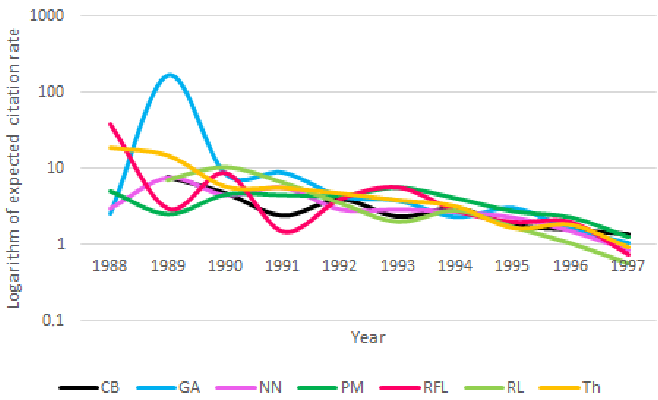

Fig. 7. Category expected citations of computer science subfields in logscale during 1988–1997

Table 4
Category Expected Citation of computer science articles from the CORA during 1988–1997

Class	CB	GA	NN	PM	RLF	RL	Th
1988	0	2.5	3	5	38	0	19
1989	7.75	165	7.625	2.5	3	7	15
1990	4.75	8.333	4.5	4.5	8.75	10.333	5.8
1991	2.4	8.667	5.643	4.429	1.5	6.5	5.571
1992	3.917	4.2	2.9	4.333	4	3.5	4.722
1993	2.318	3.708	2.907	5.556	5.667	2	3.824
1994	2.971	2.273	2.701	4.029	2.833	2.8	3.186
1995	1.786	2.957	2.284	2.714	2	1.704	1.638
1996	1.552	1.568	1.506	2.259	2	1.045	1.816
1997	1.353	1.031	0.844	1.243	0.75	0.571	0.912
Total	2.444	3.567	2.345	2.713	2.546	2.45	2.329

Nowadays, the analysis of citations and co-citations to reveal the relative importance or impact of an author or publication is the primary method in science studies. But there are a series of difficulties concerning citations. Usually, counting the citations of publications does not consider the different research fields the documents are ascribed to. The citation density and citation-based indicators in the natural sciences are also incomparably higher than in the humanities. The method of collaboration in the natural sciences, reflected in multi-specialist and multi-task teams (for instance, consisting of theorists, experimenters, analysts, and others), generates more extensive responses from researchers. As was mentioned above, to minimise these differences among research fields, the Category Expected Citations index was introduced and adopted by ClarivateAnalytics [29].

Comparing the results achieved from two approaches (Fig. 8): normalisation of the number of articles according to computer science subfields and contributions of computer science subfields, we can observe that these two factors are not equal in the overall distribution of computer science publications. For some research areas, the number of articles is higher (for example, NN); for others, the contribution of individual subfields is higher (for example, RL), and for one research area (Th), they are almost equal. Suppose we strive to present the growth of the scientific fields. In that case, we should consider that the use of the number of publications does not reproduce the proper development of knowledge which is mainly based on scholar communication in the form of direct citations and complex and interdisciplinary trees of citations.

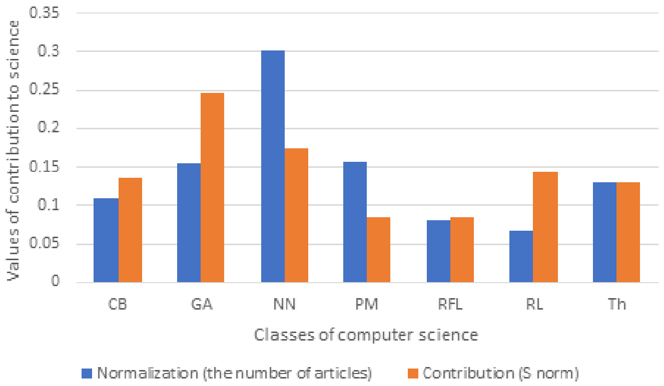

Fig. 8. Values of normalisation and contribution to scientific fields of computer science in the years 1987–1997

To calculate the normalisation values, we determine the proportional ratios of the articles of the computer science classes (blue bars in Fig. 8). Thus, we can observe that the NN papers had a higher impact on computer science during these years. Next, we calculate the proportional ratios of the contributions to computer science in all of its subfields (orange in Fig. 8). Here, a higher impact is observed for total GA contributions. Hence, comparing these two results, we can see that considering the normalisation, the picture of the actual contribution to science is blurred and incomplete. It does not consider the information connected with the considered subfield citations and the citations of articles assigned to other computer science subfields. For example, the NN normalisation ratio is the highest. However, its contribution ratio is much smaller than GA because of the larger number of more minor citations. Summing up, due to the cases of overlays of subfields in the citation trees, these normalisation ratios are not-complete. Hence, we used the non-linear calculation of the contribution ratio, and the larger orange bars mean GA and RL were very intensively cited, comparably to all of the other subfields of computer science.

The authors' approach tries to integrate information about citations and subdisciplines. By considering each citation tree separately, we can track the backpropagation of specific knowledge from the leaves (cited articles) to the root, i.e., down-up. The articles-leaves give only (for some time) their affiliation values to the development of knowledge, but they increase the contribution to the science articles when they are cited. At each successive step, we observe the cumulation of the knowledge of the considered disciplines. Thus the root article will contain the cumulative information about all of the scientific fields of the cited children. It is worth noting that aggregation of the data was performed for each year in which citations were found. So, for aggregating knowledge, we should investigate the citations and, therefore, the disciplinary contribution recursively. Using optimistic fuzzy aggregation norms, we can calculate science aggregation every year and find the root article multidisciplinary profile. Such a profile will reflect the actual distribution of knowledge. If the publication-root is ascribed to one discipline, the calculation will show the future propagation of the authors' ideas among many research fields and their non-linear contribution.

As Waltman [5] stressed, additional trouble with citations is comparing publications from different years because of the possible large variance and their undetermined values. To overcome these drawbacks, normalisation and averaging of measures have been developed to make comparisons regarding selected variables. Normalisation, however, has the following weakness. It needs to be updated with every new considered period (year) in a dataset, and any comparison leads to permanent rescaling. In the case of citations, a significant variance is usually observed, and linear normalisation by a maximum value generates a considerable divergence of the unit contribution. Instead of averaging the quantities of citations and publications, the authors propose to cumulate them from one year to another using the contribution concept. This strategy results from the necessity to construct citation trees while maintaining the information about citation inheritance. A similar approach to evaluating the disciplinary contribution to the whole field in the case of journals set by citation trees in "descending order" was applied for calculations using a maximum spanning tree [30]. In the current article, the calculations were performed by optimistic fuzzy aggregation norms, the advantage of that being their natural built-in normalisation (the range is interval $[0,1]$). Moreover, the "amount of knowledge" is continuously increasing, adding new publications and citations of articles by the property of optimism. However, in normalisation, article citation impact might be decreasing over several years when there are more unquoted articles.

We can analyse the dynamics of cumulative citation knowledge based on the proposed calculation, considering the ratios of particular disciplinary areas. As Table 2 presents, the contribution of all examined subfields of computer science is continuously increasing from 1987 to 1997. From 1995, the saturation of this indicator occurs. We can make inferences only within a given period because of the limited citation years of the CORA dataset. Every citation and publication, contributes to specified knowledge, defined by subfields. This way, despite the unknown future contribution (the number of forthcoming publications and in particular citations), we will not go beyond interval $[0,1]$. Thus, while estimating the contribution to the research areas by "adding" the small value a to the knowledge assigned to these research areas because of fuzzy-optimistic-norm properties, we can reach saturation.

5.3. Discrete derivatives

It is also interesting to see the dynamics of the discrete derivatives of the contribution to computer science in its scientific fields (Table 2). By differentiating the results according by year for contributions to the science of the subfields of computer science, we achieve the values if the discrete derivatives values presented in Fig. 9. They can be considered as the annual impacts obtained from the cumulative knowledge.

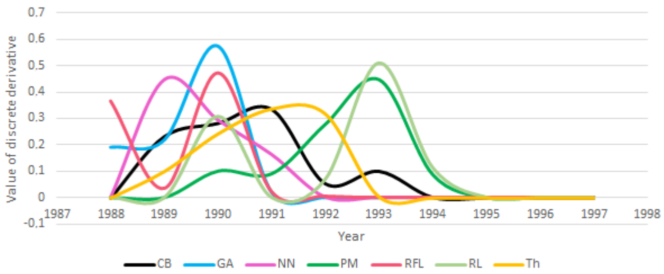

Fig. 9. Discrete derivatives of contribution to the science of computer science subfields during 1998–1997

As we can observe, the discrete derivatives behave very irregularly: some subfields increase to some maximum values (different in each case; however, they happen during 1994 – 1997), and some slowly decrease. But all have cycles of changes that are reflected by two or more hills. Furthermore, almost all disciplines present up and down cycles – we can evaluate them as wave-like fluctuations of varying intensities. These observations with comparison to the CEC distribution lead to the following conclusions during the selected period. While the citation frequencies fade away after ten years in all subfields, the nonlinear annual contributions reveal express growth in two cases, PM and RL.

As was mentioned above, the CEC is an indicator that can create comprehensive insight into the development of research fields using citation measures. The chart (Fig. 7) indicates the decreasing trend for all computer science subfields. It can be explained if we consider that articles containing new, valuable ideas sometimes occur. These articles get the most citations in the applicable field, but they constitute a small ratio of all publications (<20%). Although the number of citations exceeds the number of publications, their growth does not match the development of the number of publications.

It is challenging to assess the scientific fields because each field develops using individual paths (theoretically, they do not cross in most databases) and the publications servicing this area. Moreover, only their citations are included in the estimation of the development of this scientific field. However, even when the article is assigned to one research area, the authors refer to papers from different research areas or subdisciplines. Hence, the knowledge of the cited research areas (including articles assigned to other scientific fields or disciplines) should also increase, such as in Fig. 9. When we observe the citation trees of "scientific fields" of computer science, the paths are crossing. For example, some Neural Network articles cite Genetic Algorithm papers and vice versa. Hence, we propose to prepare the publication profile, including the contribution to the primary (assigned at the time of publication) scientific field and the contributions to research areas caused by its citations and further citations. Our method of preparing the dynamic profiles of articles shows their contribution to several (more or less general) scientific fields based on the contribution vectors during some period. These profiles can indicate papers that significantly impact research fields different from the initial ones. This is vital for some scientists whose achievements may be underestimated because their works are primarily applied in various areas, i.e., classified as multidisciplinary.

For many centuries, scientists have wanted to compare others' scientific achievements and locate themselves on the ranking lists. They are supported by economists, institutions funding science, business people, and awarding institutions. As Stallings noticed when an article has a few co-authors, it is difficult to estimate the real contribution to this paper of each author [31].

6. CONCLUSIONS

The authors of the article introduce a new method of measuring the contribution to knowledge, which analyses knowledge cumulation dynamics based on the concept of a contribution-to-scientific field. We use double cumulation for the scientific fields and years. We can calculate the contribution to the given research area caused by papers and their citations from different fields of science for each year. The proposed method is worth being considered in terms of developing concepts so that we will be able to predict the development of research areas.

In the current approach, the use of affiliation and, above all, contribution vectors are proposed instead of scalar measures concerning articles and scholars' disciplinary profiles. Therefore, a contribution to a discipline can be measured by a vector of contributions to subdisciplines or subfields. If papers are not cited, their contribution vectors are equal to the affiliation vectors. To create citation networks, we have considered each paper as a potential root, and this way, using optimistic fuzzy aggregation norms, we can evaluate the cumulative knowledge disseminated across direct and indirect citations instead of simple summarisation.

We integrate the data about the citations and affiliation of articles concerning disciplines/subdisciplines. Information about the disciplinary distribution can be helpful to present the development of scientific fields. Estimating the level of development of a scientific field only on the quantities of publications and direct citations among the discipline is insufficient. The research area is also developed by the indirect citations of papers assigned to other scientific areas or fields. Hence, it is essential to create complex and interdisciplinary citation trees and estimate the development of all scientific fields. Because of the properties of fuzzy aggregation norms, we do not need to normalise results, and the components of the contribution vectors always exist, so we can constantly evaluate scholar production. Furthermore, by comparing the affiliation and contribution vectors, we can state whether the citations of the articles matter while considering the contribution to research fields.

In comparison with linear counting, we proposed a method to flatten the dispersion of extremal values (Fig. 4, GA and NA). This way, the distribution based on traditional scientometric measures such as the h-index or g-index or originated from them will look different than if we use a proposed method of evaluation of the contribution to various scientific fields. Indeed, the CORA dataset is not robust enough to draw essential conclusions about the behaviour of disciplines in the retrospective view. In the future, we plan to extend the method using large-scale, accurate bibliographic data. It is essential to extend the citation tree to all research areas. Then, the contribution vectors provide the whole image of the given article contribution to science by the authors' research and the impact produced by its citations. One problem is knowledge saturation when the given article or scientific field contribution to science reaches 1. We can quickly improve the situation by lessening the value a and recalculating the contribution vectors.

In the future, we plan to prepare the citation tree of the articles assigned to a whole discipline, for example, computer science. Because of the application of the optimistic fuzzy aggregation norm, all achieved values belong to interval [0, 1], so they do not need to be normalised and provide more than just an image of the dynamics of the growth of knowledge in computer science. We hope that it will let us predict the development of the knowledge caused by computer science articles and their citations. Of course, there may and probably will be publications with so many citations as to disturb the study of continuous development of science. These publications provide impulses and directions for the development of science. After some time, like many other ideas, their problems will be solved, almost forgotten, or become a seed for new concepts.

REFERENCES

[1] T.S. Kuhn, *The structure of scientific revolutions*. Chicago University of Chicago Press, 1970, vol. 111.

[2] V. Barbour, "Perverse incentives and perverse publishing practices," *Sci. Bull.*, vol. 60, pp. 1225–1226, 2015, doi: 10.1007/s11434-015-0846-4.

[3] W. Glänzel, A. Schubert, and H. Czerwon, "An item-by-item subject classification of papers published in multidisciplinary and general journals using reference analysis," *Scientometrics*, vol. 44, no. 3, pp. 427–439, 1999, doi: 10.1007/BF02458488.

[4] S. Milojević, "Practical method to reclassify web of science articles into unique subject categories and broad disciplines," *Quant. Sci. Stud.*, vol. 1, no. 1, pp. 183–206, 2020, doi: 10.1162/qss_a_00014.

[5] L. Waltman, "A review of the literature on citation impact indicators," *J. Informetr.*, vol. 10, no. 2, pp. 365–391, 2016, doi: 10.1016/j.joi.2016.02.007.

[6] E. Yan, "Disciplinary knowledge production and diffusion in science," *J. Assoc. Inf. Sci. Technol.*, vol. 67, no. 9, pp. 2223–2245, 2016, doi: 10.1002/asi.23541.

[7] P.N. Tyrrell, A.R. Moody, J.O.C. Moody, and N. Ghiam, "Departmental h-index: evidence for publishing less?" *Can. Assoc. Radiol. J.*, vol. 68, no. 1, pp. 10–15, 2017, doi: 10.1016/j.carj.2016.05.005.

[8] T. Lazaridis, "Ranking university departments using the mean h-index," *Scientometrics*, vol. 82, no. 2, pp. 211–216, 2010, doi: 10.1007/s11192-009-0048-4.

[9] E. Lipitakis and A. Filip, "Bibliometrics in a nutshell," 2017.

[10] "Incites benchmarking and analytics: Understanding the metrics. types of classes and metrics: Normalized." [Online]. Available: https://clarivate.libguides.com/incites_ba/understanding-indicators (Accessed 2021-04-20).

[11] B. Clark, "Incites benchmarking and analytics, research impact reporting demands the consistent, quality data and proven impact indicators delivered by incites benchmarking and analytics." [Online]. Available: https://federation.edu.au/__data/assets/pdf_file/0007/466315/Research_Impact_Reporting_ In Cites_ Web-of-Science.pdf (Accessed 2021-04-20).

[12] N. Rons, "Partition-based field normalization: An approach to highly specialized publication records," *J. Informetr.*, vol. 6, no. 1, pp. 1–10, 2012, doi: 10.1016/j.joi.2011.09.008, 10.1016/j.joi.2012.09.001.

[13] H.F. Moed, F. de Moya-Anegon, V. Guerrero-Bote, and C. Lopez-Illescas, "Are nationally oriented journals indexed in scopus becoming more international? the effect of publication language and access modality," *J. Informetr.*, vol. 14, no. 2, p. 101011, 2020.

[14] R.N. Kostoff and W.L. Martinez, "Is citation normalization realistic?" *J. Inf. Sci.*, vol. 31, no. 1, pp. 57–61, 2005, doi: 10.1177/0165551505049260.

[15] "Web of science core collection: Data mining using analyze results," 2021. [Online]. Available: https://clarivate.libguides.com/woscc/analyze (Accessed 2021-04-21).

[16] V. Radu, F. Radu, A.I. Tabirca, S.I. Saplacan, and R. Lile, "Bibliometric analysis of fuzzy logic research in international scientific databases." *Int. J. Comput. Commun. Control*, vol. 16, no. 1, 2021, doi: 10.15837/ijccc.2021.1.4120.

[17] J.M. Merigó, W. Pedrycz, R. Weber, and C. de la Sotta, "Fifty years of information sciences: A bibliometric overview," *Inf. Sci.*, vol. 432, pp. 245–268, 2018.

[18] D. Yu and S. Shi, "Researching the development of atanassov intuitionistic fuzzy set: Using a citation network analysis," *Appl. Soft Comput.*, vol. 32, pp. 189–198, 2015, doi: 10.1016/j.asoc.2015.03.027.

[19] L. Zadeh, "Fuzzy collection," *Inf. Control*, vol. 8, pp. 338–356, 1965, doi: 10.1016/S0019-9958(65)90241-X.

[20] A. Mreła, O. Sokolov, and W. Urbaniak, "The method of learning outcomes assessment based on fuzzy relations," *Bull. Pol. Acad. Sci. Tech. Sci.*, vol. 67, no. 3, 2019, doi: 10.24425/bpasts.2019.129651.

[21] P. Prokopowicz, D. Mikołajewski, K. Tyburek, and E. Mikołajewska, "Computational gait analysis for post-stroke rehabilitation purposes using fuzzy numbers, fractal dimension and neural networks," *Bull. Pol. Acad. Sci. Tech. Sci.*, vol. 68, pp. 191–198, 2020. doi: 10.24425/bpasts.2020.131843.

[22] O. Sokolov, W. Osińska, A. Mreła, and W. Duch, "Modeling of scientific publications disciplinary collocation based on optimistic fuzzy aggregation norms," in *International Conference on Information Systems Architecture and Technology*. Springer, 2018, pp. 145–153.

[23] A. Barragao, "Cora, relational dataset repository." [Online]. Available: https://relational.fit.cvut.cz/dataset/CORA (Accessed 2020-10-20).

[24] T. Martinez, "Models of parallel adaptive logic," in *Proceedings of the 1987 IEEE Systems Man and Cybernetics Conference*. Citeseer, 1987, pp. 290–296.

[25] D.E. Goldberg and J.H. Holland, "Genetic algorithms and machine learning," *Mach. Learn.*, vol. 3, p. 95–99, 1988, doi: 10.1023/A:1022602019183.

[26] R. Yan, J. Tang, X. Liu, D. Shan, and X. Li, "Citation count prediction: learning to estimate future citations for literature," in *Proceedings of the 20th ACM international conference on Information and knowledge management*, 2011, pp. 1247–1252. doi: 10.1145/2063576.2063757.

[27] "Normalized indicators," 2020. [Online]. Available: https://incites.help.clarivate.com/Content/Indicators-Handbook/ih-normalized-indicators.htm (Accessed 2021-04-21).

[28] S. Fortunato *et al.*, "Science of science," *Science*, vol. 359, no. 6379, 2018, doi: 10.1126/science.aao0185.

[29] "Expected citation rates, half-life, and impact ratio," 2021. [Online]. Available: https://clarivate.com/webofsciencegroup/essays/expected-citation-rates-half-life-and-impact-ratio (Accessed 2021-02-20).

[30] E. Yan and Q. Yu, "Using path-based approaches to examine the dynamic structure of discipline-level citation networks: 1997–2011," *J. Assoc. Inf. Sci. Technol.*, vol. 67, no. 8, pp. 1943–1955, 2016, doi: 10.1002/asi.23516.

[31] J. Stallings *et al.*, "Scientific impact using a collaboration index," in *Proceedings of the National Academy of Sciences Jun. 2013*, 2013, pp. 9680–9685, doi: 10.1073/pnas.1220184110.