Journal of Informetrics 15 (2021) 101201

Contents lists available at ScienceDirect

Journal of Informetrics

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

A network embedding-based scholar assessment indicator considering four facets: Research topic, author credit allocation, field-normalized journal impact, and published time

Qing Xie[a,b], Xinyuan Zhang[c], Min Song[b,*]

[a] Department of knowledge resource management, School of public administration, Xiangtan University, Xiangtan, Hunan 411105, China
[b] Department of Library and Information Science, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul, 03722, Republic of Korea
[c] Zhengzhou Key Laboratory of Big Data Analysis and Application, Henan Academy of Big Data, Zhengzhou University, No. 75, Daxuebei-ro, Zhengzhou, Henan, 450000, China

ARTICLE INFO

Keywords:
Scholar performance assessment
Topic-dependent indicator
Authorship credit allocation, Field-normalized journal impact
Derek de Solla price memorial medal
The association for information science and technology's awards

ABSTRACT

Scholar performance assessment plays an important role in reward evaluation, funding allocation, and promotion and recruitment decisions. However, raw publication counts and raw citation count-based scholar performance assessment indicators, such as H-index or author citations, have shortcomings; for example, they ignore the impact of different citation patterns under different research topics, leading to authorship credit inflation due to full citation allocation to each author in multi-author publications. This study proposes a new scholar performance assessment indicator called the normalized scholar academic productivity (NSAP) indicator, which overcomes the issues posed by raw citation counts and publication counts-related scholar performance indicators by considering diverse aspects of scholar research achievements. The NSAP indicator considers the research topic, author sequence and author role in the author list, field-normalized journal impact when allocating citation counts to scholars, and published time. The research topic is generated by the co-keyword embedding and semantic relatedness of each keyword in order to make NSAP topic-dependent; the author sequence and role affect authorship credit allocation strategy; and field-normalized journal impact was used to assign different weights on raw publication counts and citation counts. Finally, awardees of the Derek de Solla Price Memorial Medal and the Association for Information Science and Technology's awards were used to evaluate the validity of NSAP for calculating scholar performance assessment. Results reveal outstanding topic-related scholar performance assessment properties compared to raw citation count indicators, such as H-index, author citations, and cited-by counts (i.e., total number of citing authors).

1. Introduction

Scholar performance assessment plays an important role in reward evaluation, funding allocation, and promotion and recruitment decisions (Bai et al., 2017). The number of publications and citations are the most commonly used measure of a scholar's scientific achievements. However, raw publication counts and citation counts-related scholar performance indicators cannot consider variation between topics, cannot evaluate scholars of varied academic ages, and have difficulty discriminating the author contribution of multi-author publications in practical evaluation applications. The reasons for these difficulties include that different research domains may have different citation patterns (Zitt et al., 2005), a growing number of multi-author articles across scientific fields make the

* Corresponding author.
E-mail address: min.song@yonsei.ac.kr (M. Song).

https://doi.org/10.1016/j.joi.2021.101201
Received 3 March 2021; Received in revised form 17 July 2021; Accepted 23 July 2021

co-authorship credit allocation more important for scholar impact assessment (Rahman et al., 2017; Wray, 2002), and raw citation counts lose comparability between cross-year article collections (Bornmann & Williams, 2013).

Consequently, bibliometrics scholars have tried to combine the research topic and publication date with the raw publication counts and raw citation counts. Some bibliometrics scholars have tried to assess author contributions using a reasonable authorship credit allocation model to determine multi-author performance assessment. There were some outstanding scholar indicators in past studies. PageRank-based methods like P-Rank ((Yan, Ding, & Sugimoto, 2011) Assimakis & Adam, 2010), SARA (Radicchi, Fortunato, Markines, & Vespignani, 2009), FutureRank (Sayyadi & Getoor, 2009), and BiRank (Jiang, Sun, & Zhuge, 2012) are good designed indicator to improve the raw citation counts. However, these graph-based methods are specifically designed for the ranking of authors in a citation network and/or co-authorship network, which need complicated preliminary calculation to ensure the feasibility for all field of science. Additionally, author-level publishing performance Index (PPI) considers citations, field of research, academic age, and number and sequence of authors as well. But the research filed of PPI was to limit the nominee within certain disciplines (Simoes & Crespo, 2020). Additionally, the author list does not follow any universally agreed-upon standard; it requires empirical testing to validate new scholar impact indicators. Even though the large publication databases, such as Web of Science (WoS), Elsevier Scopus, and arXiv.org, classify the indexed journals into comprehensive subject categories, a journal could belong to more than one category. In some fields, it may be especially difficult to find one complete paper classification scheme (Bornmann et al., 2013).

This work proposes a scholar performance assessment indicator, called the normalized scholar academic productivity (NSAP) indicator, that considers topic-dependence, author sequence and author role in the author list, and field-normalized journal impact in library and information science (LIS). In LIS disciplines, journal papers are important for scholar performance assessment. The literature types indexed in the WoS for the LIS category include: 69.50% journal papers, 5.40% book reviews, 4.13% proceedings papers, and some other types of literature. We focus on journal papers in this study. Many studies have focused on scholar impact evaluation to measure the scientific contribution of each scholar. The evaluation result could influence scientific resource distribution, such as academic awards and foundation applications (Bollen, Van de Sompel, Hagberg, & Chute, 2009); (Zhang, 2017). Traditional scholar impact evaluation methods fall into two categories: citation count-based methods (Egghe, 2006; Hirsch, 2005) and network-based methods (Amjad et al., 2016; Li et al., 2014). By contrast, performance assessment involves conveying a rational design and execution procedure with the tasks of designing, executing, interpreting, or evaluating a performance assessment for use by government, business, industry, the military, certification and licensing agencies, and educational institutions (Berk, 1986). In this context, scholar performance assessment may include not only citation counts but also their educational background, courses feedback, research topics, publications, and research funding experience. Scholar performance assessment is mainly applied between nation, university departments, or individuals (Moed et al., 1995). The NSAP aims to overcome the shortcomings of raw citation count-based indicators for scholar rank by combining with topic, author sequence, author role, and journal-level normalized indicators. Therefore, we defined NSAP as a scholar performance assessment indicator.

NSAP considers four factors because that is how it can overcome the raw citation problems. Features that compute for those factors are easy to collect or generate. The author impact can be compared on the basis of productivity, author impact, article impact, and article quality (Garfield & Welljams-Dorof, 1992). For multi-author articles, author impact was allocated in accordance with author sequence and author role in the multi-author list. Bibliometrics evaluates the impact and quality of papers. Impact is measured via citations collected by papers, while quality is determined by the impact factor. Experience has demonstrated that in each specialty the best journals are those in which it is most difficult to have an article accepted and these are the journals that have a high impact factor (Hoeffel, 1998). Furthermore, citations collected by articles increase over time, which means time-normalization of citations in necessary for scientific evaluation. Author impact assessment should consider the publication environment of specific research topics. With these objectives in mind, we designed the NSAP indicator. For research topic generation, we train keywords occurrence network embedding and word embedding to represent the structural and semantic features, while each keyword is represented using the concatenated vector of these two embeddings. K-means cluster algorithm is applied to cluster the keyword. For the calculation design of NSAP, the author sequence and role are identified to allocate authorship credit, following which the corresponding author and first author are identified using the WoS record. Then, the authorship credit is tagged with the paper topics. Later, the topic-level authorship credit is weighted by multiplying it with the field-normalized indicator of the corresponding journal. The journal indicator includes three field-normalized indicators proposed by the Centre for Science and Technology Studies (CWTS): Source Normalized Impact per Paper (SNIP), SCImago Journal Rankings (SJR), and CiteScore. Journal impact factor (JIF) proposed by Web of Science is also considered. NSAP is the sum of every topic-level authorship credit. Finally, the NSAP score is time-normalized using citations divided by published date per publication. This research's main contribution is the inclusion of research topic, author sequence and role in the author list, and field-normalized journal impact simultaneously for scholar performance assessment.

The rest of the paper is organized as follows. The second section briefly describes the main field-normalization journal impact indicators. We then review some popular authorship credit allocation approaches and their properties. Then, we introduce the state-of-the-art topic cluster approach. In the third section, we describe the calculation process of the NSAP indicator. Finally, an empirical test of the validity of the NSAP indicator was conducted upon the Price Medal awardees and the Association for Information Science and Technology's awards.

2. Related work

In this section, we present an overview of various approaches that have been developed to deal with the problem of journal-level field normalized indicators and the properties of these indicators and summarized in Table 1. Then, we discuss the advantages and

Table 1
Calculation process of common field-normalization indicators.

Indicator	Formulation	Description
JIF	JIF(k) = $\frac{C_{k-1(k)}+C_{k-2(k)}}{N_{k-1}+N_{k-2}}$	The JIF is all journal citations ($C_{k-1(k)} + C_{k-2(k)}$) in the current JCR year to items published in the previous two years, divided by the total number of scholarly items ($N_{k-1} + N_{k-2}$) published in the journal in the previous two years based on the Web of Science database.
CiteScore	CiteScore(k) = $\frac{C_{k-1(k)}+C_{k-2(k)}+C_{k-3(k)}}{N_{k-1}+N_{k-2}+N_{k-3}}$	The CiteScore is all journal citations ($C_{k-1(k)} + C_{k-2(k)} + C_{k-3(k)}$) in the current year to items published in the previous 3 years, divided by the total number of scholarly items ($N_{k-1} + N_{k-2} + N_{k-3}$) published in the journal in the previous 3 years based on the Scopus database.
SJR	$SJR_i = \frac{ave_i}{total_i}$	The variable ave_i is the average number of weighted citations received in a year of journal *i*; $total_i$ is the number of documents published in the previous 3 years of journal *i*.
SNIP	SNIP= $\frac{RIP}{RDCP}$ RIP: journal's raw impact per paper $RDCP = \frac{DCP}{median(DCP)}$ [1]	For a given year of analysis, the raw impact per paper value of a journal equals the average number of times the journal publications in the three preceding years were cited in the analysis year. The *DCP* denotes a journal database citation potential and *median* (*DCP*) denotes the median DCP value of all journals in the database.

Notes: JIF = journal impact factor of Journal Citation Report, SJR = SCImago Journal Rankings, SNIP = Source Normalized Impact per Paper.
[1] http://www.journalindicators.com

disadvantages of the popular author credit allocation approach. Finally, we review the recent co-words cluster method for topic identification using article keywords.

2.1. Journal-level field-normalized indicators

Traditionally, the "cited-side" normalization journal indicators would normalize based on the general number of publications in the field. These indicators compare citation rates per paper to the mean of such rates across a defined research field. The list of journals in the field could be taken from WoS categories or Scopus subject areas. Indicators such as the journal impact factor (JIF) of WoS and the impact per publication of Scopus provide a field comparison within the predefined subject categories.

More recently, citing-side normalization indicators were proposed using a predefined definition of the relevant field rather than the journal subject field. The SNIP is the representative citing-side normalization indicator (Moed, 2010). SJR is the PageRank-based indicators and consider the citing propensity of journals for a given cited journal.

We chose JIF, Citescore, SJR, and SNIP as journal impact indicators considering that Journal Citation Report and Scopus provides public and official indicator results for most peer-review journals with clear discipline categories. In this section, we present an overview of various approaches that have been developed to deal with the problem of journal-level field normalized indicators and the properties of these indicators and summarized in Table 1.

2.2. Authorship credit allocation model

The full counting authorship credit allocation approach has an undesirable inflationary effect; therefore, some form of fractional counting that could better represent the contribution of each author is a popular topic of research. No consensus exists about which is the most adequate authorship credit allocation approach. This section presents a review of the properties of the existing authorship credit allocation approach.

The most basic credit-assignment approach is fractional counting, allocating equal credit to each author (Oppenheim, 1998; Price, 1981). Fractional counting is easy to integrate into the variants of the h-index and raw citations.

Moreover, Geometric (Egghe et al., 2000), arithmetic (Trueba & Guerrero, 2004) and harmonic (Hagen, 2008, 2010), methods have been proposed to assign credit to avoid the ignorance of coauthor roles. Among these, harmonic counting is more accurate, fair, and flexible (Hagen, 2013). When applying harmonic counting, authors receive credit according to their byline rank. Osório (2018) proved that harmonic credit allocation could be considered a robust credit allocation scheme in scientific publications. (Hodge et al., 1981). There are also opinions against the author sequence-based authorship credit allocation model because of alphabetical author signature in some disciplines and the statement of author contribution. The use of alphabetical authorship is declining over time; it commonly exists in mathematics, economics (including finance), and high-energy physics. The consistently declining trend is proved via an empirical analysis of the use of alphabetical authorship in scientific publishing covering all fields of science (Waltman, 2012). In other fields with major papers, authors' names are ranked in descending order of the authors' contributions except for the corresponding authors (Trueba & Guerrero, 2004). Research publications in a number of journals, including some top-ranking journals (based on the journal's impact factor) such as Nature, Lancet, Proceedings of National Academy of Science USA, and the British Medical Journal, require a statement of each author's contribution. The listing of contributions gives a valuable text mining corpus to allocate authorship credit. Nonetheless, this growing trend of recognizing author's contribution is still currently non-quantitative (Rahman et al., 2017).

Shen and Barabási's work (2014)(Shen & Barabási, 2014) is discipline-independent and does not depend on the order of authors in the author list. Their result shows 86% accuracy on the citation data from several Nobel laureates and their award-winning papers. However, they claimed that further research must focus on learning to account for age- and time-dependent factors in credit allocation. Wu and his colleagues' work (2019) proved that citation count is not equal to scientific disruption or innovation. Bigger teams are

Table 2
Calculation of various authorship credit allocation models.

Authorship credit allocation model	r^{th} author's credit in an N-authored paper
Fractional	$\frac{1}{N}$
Harmonic	$\frac{\frac{1}{r}}{(1+\frac{1}{2}+\ldots+\frac{1}{N})}$

not always better in science and technology. Therefore, future work remains to be done to enrich the scientific impact indicator by considering not only citations but also creativity.

Table 2 presents the calculation of Fractional and Harmonic authorship credit allocation models. Here, N is the number of authors in a publication, r is the rank of author in the author list, and C is the credit allocation score.

2.3. Topic generation

For topic generation, Latent Dirichlet Allocation (LDA) (Blei et al., 2003) was widely used as the topic generation method to label scholar research, such as author–topic model, and author–conference–topic model. Each author is associated with a multinomial distribution over topics and each topic is associated with a multinomial distribution over words in the author–topic model, making it hard to allocate their research works in different fields (Rosen-Zvi, Griffiths, Steyvers, & Smyth, 2012). The traditional LDA model based on word co-occurrences cannot achieve a satisfactory semantic understanding of short texts because only very limited word co-occurrence information is available in short texts (Kastrati et al., 2019).

2.3.1. Network and word representation

Keywords can effectively represent topics and the main ideas of articles (Khasseh, Soheili, Moghaddam, & Chelak, 2017). In traditional bibliometric analysis, author keywords play a critical role in co-term analysis and capturing topic terms (Haunschild et al., 2019). Keywords that accompany an article are usually abstract definitions of the main research context on which the article focuses (Behrouzi et al., 2020). Recent co-term analysis used for topic identification is a content analysis method combining bibliometrics and deep learning technology to reveal the deep meaning of documents. Bibliometrics counts the frequency of word co-occurrence in a document and clusters these words based on co-occurrence to reveal the closeness between them. This analysis usually suggests that the higher the frequency of words' co-occurrence within a document, the closer they are (Milojević et al., 2011). Measuring the distance of keywords in a co-keyword network can indicate the research similarity of each keyword. This research similarity is slightly different from semantic relatedness as the context word is not a sentence but is another keyword studied in the same paper. Those research similarities are necessary for research topic identification. Thus, a co-keywords network was constructed. To capture more keyword relation characteristics, the network is represented as vector space using network-embedding algorithms (Grover & Leskovec, 2016). Furthermore, deep learning technology measured semantic distance between keywords based on the document as a corpus. Word embedding, as one such application of deep learning, maps words from a textual vocabulary to numeric vectors and, in doing so, creates a way to extract topics and discover the latent semantics in large-scale text (Mikolov et al., 2013). The key idea is that, words with similar meanings often appear in similar contexts. Word2Vec method is a well-recognized word embedding technique (Le & Mikolov, 2014). Furthermore, a keyword vector is represented by co-keywords embedding as well as Word2Vec semantic-relatedness embedding concatenated as a new vector. Research topic identification aims to sort keywords into groups or clusters so that the degree of correlation is strong between members of the same cluster and weak between members of different clusters (Braam et al., 1991). Those concatenated vectors are applied as input data for the k-means clustering model (Zhang et al., 2018).

2.3.2. Clustering algorithms

In previous work, some clustering algorithms have been widely used. The basic idea of Density-Based Spatial Clustering Algorithm with Noise (DBSCAN) is that for each point of a cluster, the neighborhood of a given radius (ε) has to contain at least a minimum number of points (MinPts) where ε and MinPts are input parameters (Duan et al., 2007). However, DBSCAN cannot efficiently detect clusters when there is significant spatial heterogeneity in the dataset, as is the case for Twitter data where the distribution of users as well as the intensity of publishing tweets varies over the study areas (Ghaemi & Farnaghi, 2019). The publication distribution of different research topics also presents spatial heterogeneity in that hotspot topics always have many more articles than less popular ones.

Affinity Propagation (AP) Clustering algorithm views each data point as a node in a network and recursively transmits real-valued messages along edges of the network until a good set of exemplars and corresponding clusters emerges (Frey & Dueck, 2007).

Agglomerative hierarchical clustering (AHC) is a bottom-up clustering algorithm that regards each data point as a single cluster and calculates the distance between all clusters to merge them until they are all merged. The choice of which clusters to merge or split is determined by a linkage criterion, which is a function of the pair-wise distances between observations (Sembiring et al., 2011). The cluster distances of AHC could be single-link, complete-link, UPGMA, WPGMA, Centroid, or Ward (Murtagh & Contreras, 2017). K-means is another commonly used document clustering method. There are studies indicating that agglomerative hierarchical clustering is better than K-means, although this result is just for a single dataset or non-document data (Douglass, 1992). In practice, K-means is used because of its efficiency and agglomerative hierarchical clustering is used because of its quality. There are also works showing

Table 3
Data collection.

Subject	Retrieval	Number of articles	Number of articles (Scopus)
LIS	(WC="Information Science & Library Science") AND LANGUAGE: (English) AND DOCUMENT TYPES: (Article) Indexes=SCI-EXPANDED, SSCI Timespan=1980-2020.9.12	103,305	67,440

that an efficient K-means variant would perform better than AHC (Steinbach et al., 2000) and K-Means is a good option for clustering large-scale data sets (Reddy et al., 2017).

The *K*-means algorithm (Likas, Vlassis, & Verbeek, 2003) uses the following steps. First, classes are selected, and the center points, which are the same length as each data point vector, are randomly initialized (although we must know the number of classes in advance). Second, the distance from each data point to the center point is computed, and the data point is classified into the class in which the center point is the closest. Third, the center point in each new class is calculated. Finally, Steps 1, 2, and 3 are run until each center is stable. The *K*-means algorithm is easy to calculate, but we must know how many classes exist in advance.

We chose the K-means method for clustering. In our case, the data do not have labels. In this situation, the silhouette coefficient (Rousseeuw, 1987) is always used to assess the number of classes, because it considers cohesiveness and separation simultaneously.

In addition, the author-defined keyword dataset may not be homogeneous. Thus, a standard and robust preprocessing method of keywords is a necessary natural language programming process. The Natural Language Toolkit (NLTK) is the leading platform for building Python programs to work with human language data.

To sum up, scholar performance assessment with a onefold bibliometric indicator has proven to be susceptible to rank bias. First, a journal impact indicator is designed to measure the average impact of papers included in journals. Using the journal impact indicator to evaluate authors directly is against the original intention of indicator design, and JIF itself has been misused in many situations (Garfield, 1999). Second, citation-based scholar indicators like H-index, which ignores citation time, are biased towards senior scholars and underestimates the scientific impact of junior scholars because a citation is a variable that keeps accumulating over time and articles with a longer citation window will tend to have higher citation counts than articles with shorter citation windows (Thelwall & Fairclough, 2015). Third, all-author citation analysis become increasingly important with the increase of multi-author scientific collaboration. Previous studies have revealed that all-author analysis was a more efficient way to identify authors' research fields compared with the first-author type (Eom, 2008; Persson, 2001). Finally, citation behavior across research areas, i.e., topic-level scientific evaluation, proved to be a helpful strategy for fine-grained academic evaluation. However, few contemporary scholar impact indicators try to address all of the above problems.

3. Method

Our proposed indicator to assess author impact is illustrated in Fig. 1. The keywords, titles, and abstracts of articles were collected from WOS; details are provided in Section 3.1. Regarding topic generation (Section 3.2), we created and trained a co-keyword network embedding. The keywords considered are the author-provided keywords; if a paper did not list keywords, we extracted the potential keywords from the paper title. In addition, all the abstracts were collected for training the word embedding to calculate semantic relatedness. The same keyword has a different representation in each embedding. Then, for each keyword, we concatenated the vector from those two embeddings as the new representation. In Section 3.2.2 we tested five different cluster algorithms, K-mean algorithm performed best. Thus, K-means algorithm was applied for topic clustering by using the new representation. In this manner, each topic was assigned a group keywords collection (e.g., Topic 1, Topic 2, Topic 3). For indicator calculation, we collected the paper citation count, published year, author role and sequence, and field-normalized journal impact of each publication from every researcher. The author sequence and author role were used for computed authorship credit score (Section 3.3). Furthermore, we accumulated the score of publication, if the keyword of that publication belongs to the same topic. Details are provided in Section 3.4.

We ranked the awardees of the Derek de Solla Price Memorial Medal (DSPM) and the Association for Information Science and Technology (ASIS&T) awards using the proposed indicator and other indicators and compared the results.

3.1. Data collection

In Table 3, we have summarized the formulated queries to retrieve the articles from WoS. The document object identifiers (DOIs) of all articles were then used to obtain the author_id and other author information from the Scopus database. Two databases were used because Scopus provides automatic author disambiguation with the Scopus author ID, while WoS provides the corresponding author identification. If the paper had no DOI, we used the journal name and title of the paper to find the EID field in Scopus to obtain the author_ID.

Considerable effort is required to deal with the author name disambiguation because of distorted author ranking in citation analysis as well as information retrieval (Strotmann & Zhao, 2012). Especially in co-authorship networks, disambiguated data would bring down the average shortest paths and clustering coefficient while inflating the network density and component size (Kim & Diesner, 2014). Author name disambiguation is rather complicated. A single author may publish under multiple names because of spelling errors or variants, and name changes may occur because of marriage or national conversion. Furthermore, many individuals

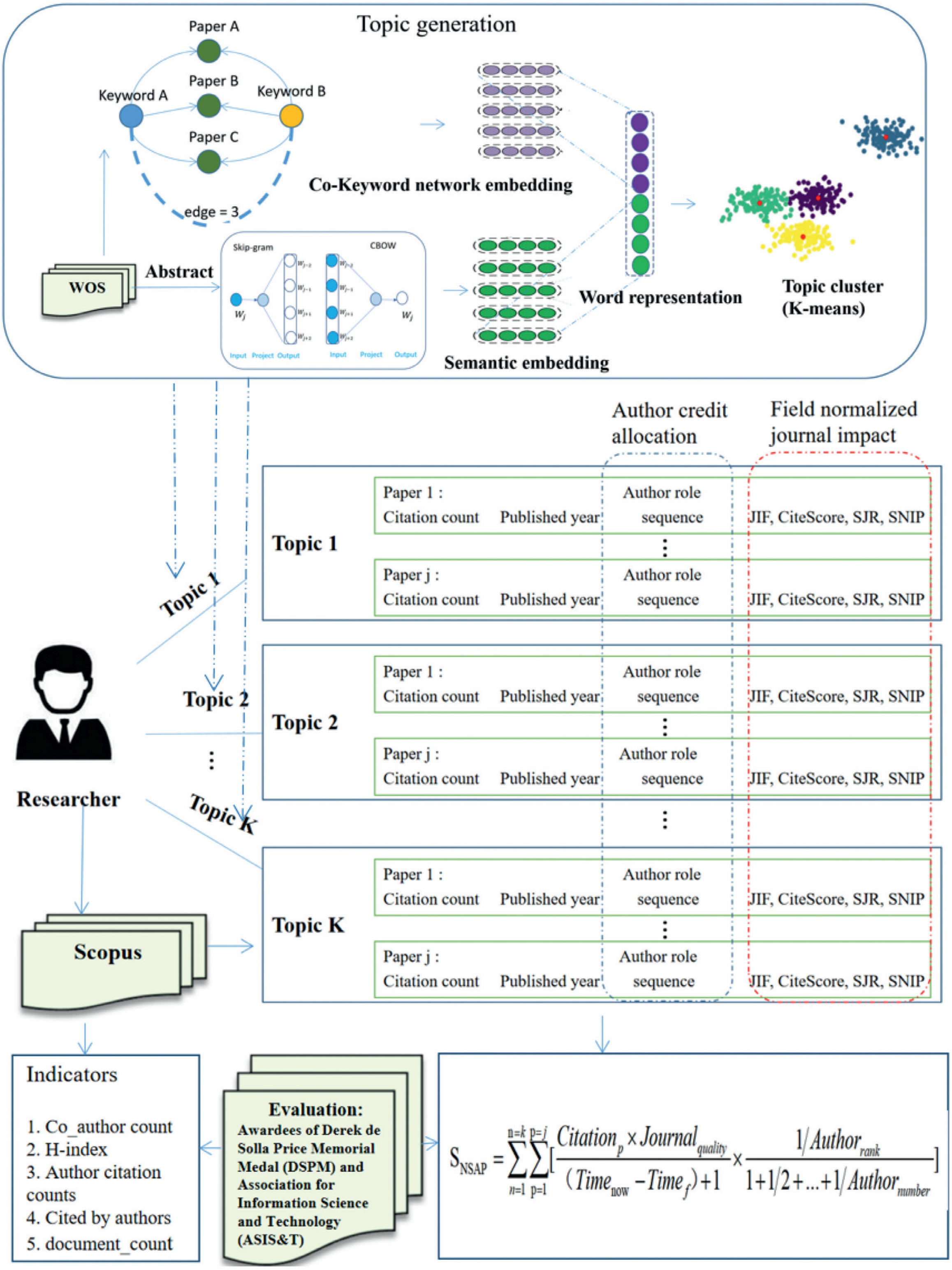

Fig. 1. Overview of the proposed method.

Table 4
Example of extract keyword from paper title.

comparative evaluation of three continuous speech recognition software packages in the generation of medical reports	
Trigrams term in the dictionary	speech recognition softwar
comparative evaluation of three continuous packages in the generation of medical reports	
Bigrams term in the dictionary	comparative evalu
of three continuous packages in the generation of medical reports	
unigram	continu
term	packag
in	gener
the	medic
dictionary	report

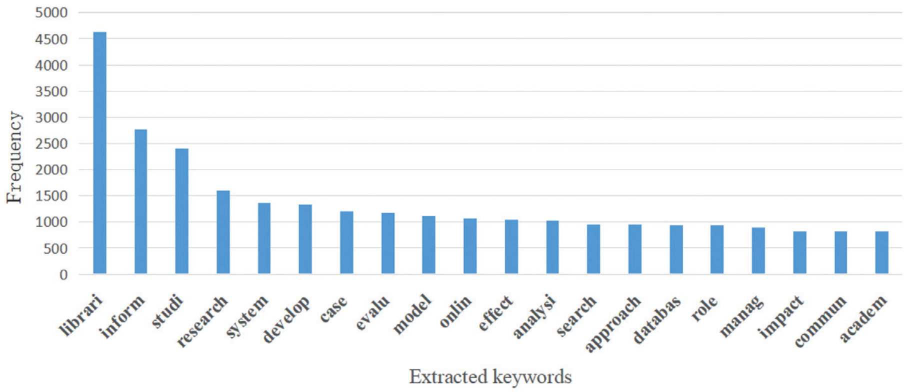

Fig. 2. Top 20 highest frequency extracted keywords.

share the same popular names. Bibliographic database metadata records including the author's first names, geographical locations, degrees, or positions are not absolutely complete. Scopus works well on researcher disambiguation with its Scopus author identifiers by assigning a unique ID for anyone who has created a scholarly work. It covers more than 17 million universal author profiles. Scopus also allows researchers to associate their Scopus Author Identifier to their ORCID and import papers from Scopus to his/her ORCID profile. Through the Scopus Author ID, the Scopus API allows easy analysis and tracking of an individual's citation history.

3.2. Topic generation

3.2.1. Keyword extraction

Keywords are regarded as the main ideas of the topic and the article content (Hu & Zhang, 2015) (Khasseh, Soheili, Moghaddam, & Chelak, 2017) Therefore, we used the keywords to generate the topics because some similar words exist in the keyword list. To this end, we used the NLTK package (https://www.nltk.org/) for stemming. Stemming is the process of reducing a word to its word stem. For example, using a stemming algorithm, words like "models" and "modeling" are stemmed to just "model." Those keywords were used to create a co-keyword network. If a paper does not provide keywords, we created a dictionary with all keywords from the collected dataset. The paper titles were used to extract the keywords. We extracted keywords by considering the trigrams, bigrams, and unigram of the paper title. Table 4 presents an example.

Fig. 2 presents the top 20 highest frequency extracted keywords.

Node2vec algorithms (Grover & Leskovec, 2016) trained the co-keyword network into vector space. In addition, Word2vec (Mikolov et al., 2013) was applied for training the word embedding for the semantic-relatedness calculation. Thus, the same keyword had different representations in each embedding. For each keyword, we learned the vector representation as 64 dimensions according to the empirical evaluation in each embedding. We concatenated the keyword vector from those two embeddings as a new representation; thus, every keyword was now represented by 128 dimensions. K-means algorithm was applied for clustering the topic based on those vectors of keywords. In this way, each topic is assigned with a group of keywords and without overlay.

3.2.2. Topic clustering algorithms comparison

In this section, K-means clustering, LDA topic model, DBSCAN, Affinity Propagation algorithm, hierarchical agglomerative clustering were applied for topic clustering algorithms comparison.

3.2.2.1. K-means cluster. K-means algorithm was applied for clustering; to do this, we need to identify the K value, i.e., the number of clusters in the dataset. The silhouette coefficient (Rousseeuw, 1987) was used to identify the number of clusters; the higher the silhouette coefficient value, the better the clustering effect. The vector of representation (co-keyword embedding and semantic

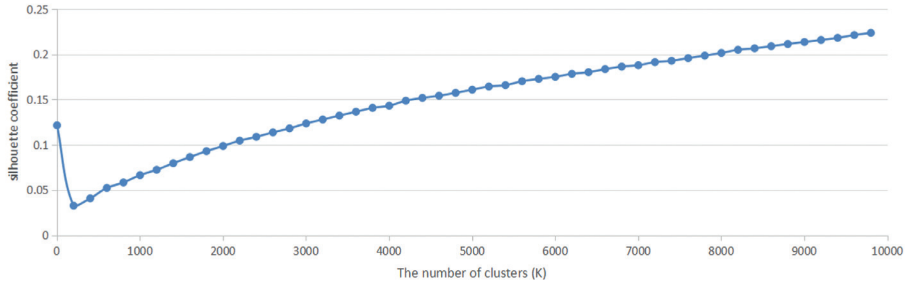

Fig. 3. Number of clusters (K) and its silhouette coefficient from 2 to 10000 at intervals of 200.

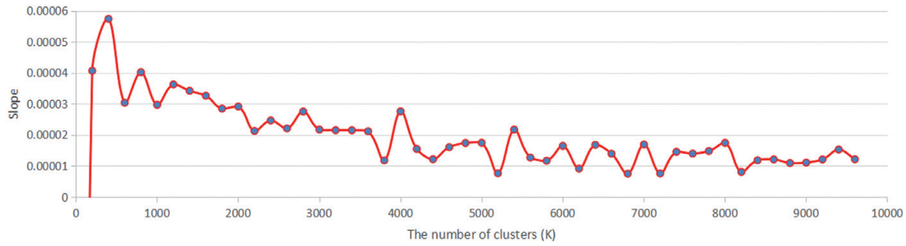

Fig. 4. Slope of silhouette coefficient from 2 to 10000 at intervals of 200.

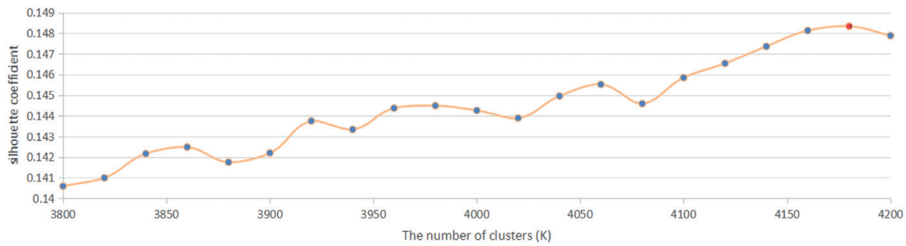

Fig. 5. Number of clusters (K) and its silhouette coefficient from 3800 to 4200 at intervals of 20.

embedding) is the input of the K-means model. The K-means model and silhouette score were computed using the scikit learn cluster module (https://scikit-learn.org/stable/modules/clustering.html) and scikit learn metric package. First, the number of clusters was tested from 2 to 10000 at intervals of 200 as shown in Fig. 3. We found the value increasing at this interval. However, setting k=10000, the average keyword in each topic is less than seven, and thus, the slope of each sampling is computed as shown in Fig. 4.

The higher the slope, the higher the change in unit time. A local maximum value was achieved at k=4000 and from 2 to 4000 the average slope is higher than the slope from 4000 to 10000. Thus, we identify the number of clusters to be around 4000.

Furthermore, we test the number of clusters from 3800 to 4200 at intervals of 20. Fig. 5 shows the relationship between the number of clusters and its silhouette coefficient. We identify the cluster with the highest silhouette coefficient as 4180.

Table 5 presents the topic cluster, which includes the top five highest frequency paper-provided keywords in our collected dataset.

As summarized in Table 5, “question ask”, “thesaurus construct”, and “text mining” appear under the information retrieval topic, which is reasonable.

3.2.2.2. LDA algorithm. Furthermore, we generated the topic using the LDA algorithm, with the number of topics set at 110 based on the empirical evaluation. Table 6 lists the result of the top five LDA topic clusters, with the number in the bracket being the significance of the word per topic. Although some related words are classified into the same topic, such as “education” and “students”, most words show less relatedness. In addition, some words appeared in several topics, such as “network” and “information”.

3.2.2.3. DBSCAN. We tested the combination of MinPts and ε parameter, MinPts from 2 to 10, ε from 0.1 to 0.9 at intervals of 0.1. Table 7 lists the result. The number of clusters is 1 when ε is 0.1, 0.2, 0.3. In this case, the silhouette coefficient may not be a good performance indicator, even MinPts as 10 and ε value as 0.8 or 0.9 achieved the highest silhouette coefficient, the number of clusters is 2, which is not meaningful. Thus, we set MinPts as 2 and ε value as 0.9 according to the empirical evaluation, however, the top 5 highest-frequency keywords in our collected dataset (e.g., Internet, Knowledge management, social media, Bibliometrics, Information retrieval) are identified as noise. this is because our data have different densities and the difference is very large, and DBSCAN difficult to cluster high-dimensional data, therefore, DBSCAN cannot provide a good clustering result in our case.

Table 5
Topic cluster of top 5 highest-frequency keywords (k-means).

Rank	Keyword (frequency)	Keyword (The top 10 degrees of keywords in the Co-keyword network)
1	Internet (1043)	Internet (2979), public domain (9), course evaluation (6), command language (5), online resource seek (5), text message interface (5), cell phone only household (5), general administration of the state (5), Spanish government (5), online social activities (5)
2	Knowledge management (976)	knowledge manage (2170), knowledge network (11), networked project (8), knowledge management model (8), governance reform (8), small and medium sized enterprises (8), defense sector (8), Indian work context (7), knowledge audit (7), adaptor-innovator style (7), non-government organization (7), manual work (7)
3	Social media (925)	social media (2680), gang (13), interpretive community (10). Infodemiology (8), efl (7), Strawson (7), scholarly read (7), heritage organ (7), local history (7), reactive attitude (7)
4	Bibliometrics (859)	Bibliometric (2388), eponym (15), trendmd (11), plate tectonic (10), derek jd price (9), history of scientometrics (9) jaeri (9), oldest old (9), etymology of scientometrics (9), jd bernal (9), publication trend (8)
5	Information retrieval (856)	information retrieve (1582), unconscious (17), question negotiation (15), google book (13), thesaurus construct (13), zprise (9), question ask (9), academic miss (8), automatic summarization system (7) text mining (7)

Table 6
LDA topic cluster.

Topic cluster	Words in topic cluster
1	education (0.340), students (0.204), standards (0.086), information (0.061), assessing(0.057), library(0.055), hierarchical (0.030), food (0.027), segmentation(0.024), terrain(0.011)
2	impact (0.373), publication (0.113), research (0.063), medication(0.048), accuracy (0.040), improving (0.038), credibility (0.032), city (0.028), environment (0.024),think (0.018)
3	network (0.299), community (0.295), scholar (0.040), crisis(0.036), reputation (0.030), rules (0.024), protocol(0.022), doctoral(0.019), bayesian (0.019), thesaurus (0.016)
4	user (0.289), satisfaction (0.085), needs (0.070), computing (0.066), information (0.060), comparative (0.047), perceptions (0.045), exploratory (0.040), space (0.039), end (0.029)
5	network (0.404), analysis (0.157), local (0.064), rights (0.046), flow (0.041), specific (0.029), rule (0.023), solution (0.019), repository (0.018), ownership (0.017)

Table 7
The number of clusters according to different parameters of DBSCAN clustering.

(MinPts, ε)	The number of clusters	silhouette coefficient	(MinPts, ε)	The number of clusters	silhouette coefficient	(MinPts, ε)	The number of clusters	silhouette coefficient
(2, 0.1)	1	-	(2, 0.4)	59	0.145102	(2, 0.7)	529	-0.07678
(3, 0.1)	1	-	(3, 0.4)	11	0.199796	(3, 0.7)	270	0.079171
(4, 0.1)	1	-	(4, 0.4)	2	0.269873	(4, 0.7)	123	0.134173
(5, 0.1)	1	-	(5, 0.4)	1	-	(5, 0.7)	40	0.159797
(6, 0.1)	1	-	(6, 0.4)	1	-	(6, 0.7)	16	0.183541
(7, 0.1)	1	-	(7, 0.4)	1	-	(7, 0.7)	5	0.215226
(8, 0.1)	1	-	(8, 0.4)	1	-	(8, 0.7)	3	0.239142
(9, 0.1)	1	-	(9, 0.4)	1	-	(9, 0.7)	3	0.239142
(10, 0.1)	1	-	(10, 0.4)	1	-	(10, 0.7)	2	0.304905
(2, 0.2)	1	-	(2, 0.5)	173	0.117825	(2, 0.8)	738	-0.24512
(3, 0.2)	1	-	(3, 0.5)	47	0.150841	(3, 0.8)	429	0.006262
(4, 0.2)	1	-	(4, 0.5)	14	0.195646	(4, 0.8)	230	0.087673
(5, 0.2)	1	-	(5, 0.5)	3	0.257217	(5, 0.8)	98	0.105981
(6, 0.2)	1	-	(6, 0.5)	1	-	(6, 0.8)	32	0.131941
(7, 0.2)	1	-	(7, 0.5)	1	-	(7, 0.8)	8	0.196737
(8, 0.2)	1	-	(8, 0.5)	1	-	(8, 0.8)	4	0.231902
(9, 0.2)	1	-	(9, 0.5)	1	-	(9, 0.8)	3	0.239266
(10, 0.2)	1	-	(10, 0.5)	1	-	(10, 0.8)	2	0.30518
(2, 0.3)	9	0.199053	(2, 0.6)	317	0.066648	(2, 0.9)	1106	-0.34883
(3, 0.3)	2	0.290353	(3, 0.6)	123	0.119131	(3, 0.9)	630	-0.12071
(4, 0.3)	1	-	(4, 0.6)	48	0.152376	(4, 0.9)	348	-0.00541
(5, 0.3)	1	-	(5, 0.6)	12	0.190168	(5, 0.9)	155	0.054301
(6, 0.3)	1	-	(6, 0.6)	4	0.218153	(6, 0.9)	55	0.100179
(7, 0.3)	1	-	(7, 0.6)	3	0.238569	(7, 0.9)	16	0.126603
(8, 0.3)	1	-	(8, 0.6)	1	-	(8, 0.9)	5	0.219463
(9, 0.3)	1	-	(9, 0.6)	1	-	(9, 0.9)	3	0.239266
(10, 0.3)	1	-	(10, 0.6)	1	-	(10, 0.9)	2	0.30518

Table 8
Silhouette coefficient according to damping parameter of affinity propagation clustering.

damping	The number of clusters	Silhouette coefficient
0.5	2886	0.197805211314118
0.6	2891	0.198086051182374
0.7	2888	0.197110478364329
0.8	2890	0.197215916216905
0.9	2892	0.197024578912345

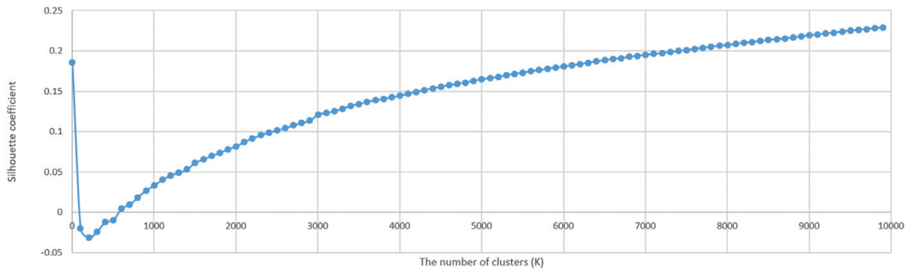

Fig. 6. Number of clusters (K) and its silhouette coefficient from 2 to 10000 at intervals of 100.

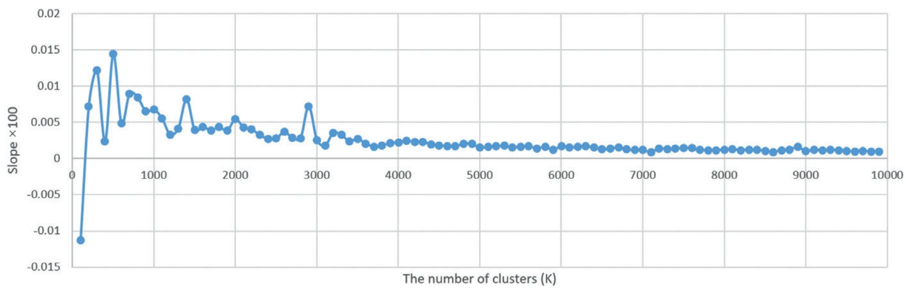

Fig. 7. Slope of silhouette coefficient from 2 to 10000 at intervals of 100.

3.2.2.4. Affinity propagation algorithm. Preference and damping are important parameters for AP that are used for determining the number of clusters. According to sklearn.cluster. AffinityPropagation packet, the preferences are not passed as arguments, they will be set to the median of the input similarities. Thus, we did not pass the preference parameter. We tested the damping parameter from 0.5 to 0.9 at intervals of 0.1. Table 8 lists the result. Dampling with 0.6 better than others.

3.2.2.5. Hierarchical agglomerative clustering. The number of clusters was tested from 2 to 10000 at intervals of 100 as shown in Fig. 6. The value increasing at this interval. the slope of each sampling is computed as shown in Fig. 7. A local maximum value was achieved at k = 2902 and from 2 to 4000 the average slope is higher than the slope from 2000 to 10000. Thus, we identify the number of clusters to be around 2902.

Table 10 presents the topic cluster, which includes the top five highest frequency paper-provided keywords in our collected dataset. Compare with k-mean and AP cluster, the Hierarchical agglomerative cluster result of the top five highest frequency paper provided keywords include many keywords. There are 663 keywords in social media related cluster, 143 in Internet, 445 in Knowledge management. 220 in Bibliometrics. 292 Information retrieval. Thus, it is not rational to list the top 10 degrees of keywords in the co-keyword network of each topic. Therefore, periodic sampling was applied, for social media related cluster, we listed the keyword that ranked 1, int(663/10), 2 × int(663/10),…9 × int(663/10) by degrees of keywords in the Co-keyword network.

3.2.2.6. Topic clustering algorithm selection. In this subsection, we discussed the selection process of the clustering model in indicator level. In addition, we checked the topic cluster of the top-5 highest-frequency keywords. We consider the top-5 because Essential Science Indicators (ESI) takes top 0.1% as the threshold value for global highly cited paper or subject evaluation. Here, we use the same threshold. We identify about 4000 clusters; therefore, the topic cluster of top-5 highest-frequency keywords are the most dominant clusters.

Table 11 lists the silhouette coefficient value of K-means clustering, DBSCAN, Affinity Propagation algorithm, and hierarchical agglomerative clustering. Affinity Propagation algorithm ranked first by the silhouette coefficient score. And k-means clustering ranked second. Later, we did a manual check on the topic cluster of top-5 highest-frequency keywords.

Table 9
Topic cluster of top 5 highest-frequency keywords (Affinity Propagation algorithm).

Rank	Keyword (frequency)	Keyword (The top 10 degrees of keywords in the Co-keyword network)
1	Internet (1043)	Internet (2979), family communication (15), computer technology (14), sociology of scientific knowledge (9), internet experience (9), homeless youth (8), email survey (8), rumoring theory (7), Kierkegaard (7), reason (7).
2	Knowledge management (976)	knowledge manage (2170), knowledge repositories(43), knowledge strategy (41), km (19), metaknowledge (19), knowledge act (18), legal research (16), center of excel (12), requirement analysis (12), knowledge creation process (12)
3	Social media (925)	social media (2680), conceptual art (12), models of inform (12), popper karl (12), marcxml (8), indexing system (7), tiara (7), wordpress (6), artist magazine (6), reference style (6).
4	Bibliometrics (859)	Bibliometric (2388), scientometric (33), scientific map (21), publication type (20), materials sci (20), siena (15). history of psychology (10), publication success (6), stochastic actor based model (6), pseudoscaphirhynchus (5)
5	Information retrieval (856)	information retriev al(1582), academic miss (8), enhanced working memories (7). cognitive information retrieval (7). Neanderthals (7), website design (6). price equation (5), technology simulated threat (5), nonconscious cognit (3), adolescent motherhood (2)

Table 10
The topic cluster of top 5 highest-frequency keywords (Hierarchical agglomerative**).**

Rank	Keyword (frequency)	Keyword (The degrees of keywords in the Co-keyword network)
1	Internet (1043)	internet (2979), public sector information (17), patient narratives (9), network camera (7), technical standard(6), online resource seek(5), stated choice method (5), telecommunication regulation and economics(4), parents perceptions (4) health issue (3)
2	Knowledge management (976)	knowledge management (2170), learnable (18), strategic leadership (9), universities faculty member (7), knowledge processing strategi (7), product model (5), pay for performance policy (5), dynamic organ (4), activity based management (4), dialogue system (3)
3	Social media (925)	social media (2680), political representation (15), content mining (9), human machine collaborative information process (7), organization public relations (6), news channel (5), facioscapulohumeral muscular dystrophy fshd (6), Wechat library (4), source (4), internet travel (4), brazilian popular music (2).
4	Bibliometrics (859)	Bibliometric (2388), botani (17), leadership approach (9), order of authorship (7), cultural safety (7), direct and diffuse solar radiation (5), bradfordzipf distribution (4). solid waste (4) sti studies (4), executive function (3)
5	Information retrieval (856)	information retriev al(1582), arabic text (17), evaluation time (9), multicomponent (7), linear transform (5), document order (5), lexical and syntactic relationship (4), local context analysis (4), indexing and document rank (4), cause effect relation (3)

For k-means clustering, Table 5 summarizes the topic cluster of the top 5 highest-frequency keywords with k-means. It could be seen that the top 10 degrees of keywords in the co-keyword network under each cluster are reasonable in light of research experiences. While, in Table 9 (Affinity propagation algorithm), the word homeless youth, Kierkegaard, and email survey were classified into Internet seems unexplainable. The materials sci and siena were classified into Bibliometrics also seems unreasonable.

In addition to the clustering coefficient, the readability of clustering results and the feasibility of the calculation process are also important goals of topic clustering. The algorithm with better divergence and reasonable CPU time is regarded as competent for clustering requirements (Kalepalli et al., 2020). Due to the high complexity of AP algorithm, the running time is much longer than that of K-means, especially for massive data. Even though k-means need the predefined number of clusters, but it provides a reasonable result and saves much more time. For other algorithms, the result of Table 6 shows that the LDA topic cluster approach would classify words with less relatedness into the same topic. Table 7 lists the silhouette coefficient of the cluster under different combinations of MinPts and ε parameter. It is hard to get a rational clustering count and silhouette coefficient. Therefore, DBSCAN cannot provide a good clustering result in our case. Thus, we think used K-means for further analysis is reasonable.

3.3. Authorship credit score computed by author sequence and author role

As Fig. 1 shows, for each researcher, per topic, we collected the citation count, published year, author role and sequence, and field-normalized journal impact of every paper that contain the words included in that topic. For author sequence, we used the harmonic authorship credit allocation model (formula provided in related work). Because harmonic counting method has been considered superior to the previously mentioned ones, mainly because of its ability to fit empirical data from medicine, psychology and chemistry robustly (Hagen, 2010; Liu & Fang, 2012; Osório, 2018). However, previous study claimed that the performance of an authorship credit allocation scheme seems to heavily depend on empirical datasets (Kim & Kim, 2015). Table 12 lists the authorship credit score for a paper with up to six authors. If a paper has two authors and receives 10 citation counts, then the first author gets 0.6667 of the citation count (10×0.6667) i.e., 6.667 citation count. The second author gets 0.3333 of the citation count, i.e., 3.333 citation count. In addition, author role is considered in this work. If the second author is the corresponding author, he can get the same citation count as the first author (i.e., 6.667 citation count, not 3.333).

Table 11
The silhouette coefficient value.

Methods	The number of topics	Silhouette coefficient
K-means clustering	4180	0.1482
DBSCAN	1106	-0.3488
Affinity Propagation algorithm	2891	0.1980
hierarchical agglomerative clustering	2902	0.1252

Table 12
Authorship credit score for papers with up to N = 5 authors.

Counting method	Authors number	Authorship Rank				
		1st	2nd	3rd	4th	5th
Harmonic	1	1				
	2	0.6667	0.3333			
	3	0.5455	0.2727	0.1818		
	4	0.4800	0.2400	0.1600	0.1200	
	5	0.4380	0.2190	0.1460	0.1095	0.0876

3.4. Indicator calculation

As shown in Fig. 1, academic ability was computed for each topic by considering the citation and quality of the papers (field-normalized journal impact; JIF, CiteScore, SJR, and SNIP). Furthermore, we summarized the academic ability of each author under each topic. The formula is as follows:

$$S_{\mathrm{NSAP}} = \sum_{n=1}^{\mathrm{n}=k} Topic_{n(Academic\_ability)} \tag{1}$$

where K is the number of topics and $Topic_{n(Academic\_ability)}$ comprises the citations of each article, publication number, published date, and quality of the papers under that topic. Scopus freely provides CiteScore, SNIP, and SIR (https://www.scopus.com/sources.uri). Clarivate Analytics Journal Citation Reports provide the newest JIF.

$$Topic_{n(Academic\_ability)\_Journal_{quality}} = \sum_{\mathrm{p}=1}^{\mathrm{p}=j} \left[\frac{Citation_p \times Journal_{quality}}{(Time_{\mathrm{now}} - Time_f) + 1} \times \frac{1/Author_{rank}}{1 + 1/2 + \ldots + 1/Author_{number}} \right]. \tag{2}$$

The first corresponding author shares the same rank as the first author. j is the number of papers that contain the target topic. Time_f is the year of publication, $\mathrm{Time}_{\mathrm{now}}$ is the current evaluation year, $Citation_p$ is the citation count of a special paper, and $Journal_{quality}$ is the journal JIF, CiteScore, SJR, or SNIP score of the journal in which paper P was published. $Author_{rank}$ is the author ranking of each paper and $Author_{number}$ is the number of authors for each paper. In addition, the corresponding author shares the same rank as the first author. In the WoS dataset, the "RP" field contains information on the corresponding author, including the name and address. Therefore, the NSAP indicator is given by:

$$S_{\mathrm{NSAP}} = \sum_{n=1}^{\mathrm{n}=k} \sum_{\mathrm{p}=1}^{\mathrm{p}=j} \left[\frac{Citation_p \times Journal_{quality}}{(Time_{\mathrm{now}} - Time_f) + 1} \times \frac{1/Author_{rank}}{1 + 1/2 + \ldots + 1/Author_{number}} \right] \tag{3}$$

where the first corresponding author shares the same rank as the first author. For the indicator evaluation, we compared our proposed method with the coauthor count, H-index, number of publications (document_count), author citation counts (raw citation counts for each author), and cited-by counts (total number of citing authors) of a special author.

4. Results

We displayed descriptive statistics for all indicators in our collected dataset. Later, the DSPM dataset was used as the gold standard to evaluate the performance in each indicator. An example was given to observe the details of scholar ability under each topic. We also analyzed the highest scores for scholars for the five popular topics in the Library and information science (LIS) subject. In addition, to validate the NSAP indicator in other research fields within the LIS discipline, the names of the Award of Merit recipients and Research in Information science awardees of the Association for Information Science and Technology (ASIS&T) are collected to evaluate the performance.

4.1. Descriptive statistics of NSAP and traditional scholar impact indicators

Table 13 presents the results of the descriptive statistics. The standard deviation reflects the dispersion degree of a dataset and the variance indicates the difference between the observed value and population mean. The author citation count refers to the number

Table 13
Descriptive statistics.

Indicators	Min	Max	Average	Standard deviation	Variance
NSAP_JIF	3.57E-03	7.28E+03	3.71E+01	1.44E+02	2.06E+04
NSAP_CiteScore	0.00E+00	1.33E+04	7.14E+01	2.68E+02	7.19E+04
NSAP_SJR	4.28E-03	3.79E+03	1.75E+01	7.56E+01	5.71E+03
NSAP_SNIP	0.00E+00	3.71E+03	2.27E+01	8.21E+01	6.75E+03
H-index	0.00E+00	1.18E+02	8.77E+00	1.01E+01	1.02E+02
Author citation count	0.00E+00	1.03E+05	8.11E+02	2.59E+03	6.71E+06
Coauthor count	0.00E+00	1.29E+04	5.85E+01	2.51E+02	6.30E+04
Cited-by counts	0.00E+00	8.98E+04	6.45E+02	1.96E+03	3.83E+06
document_count	0.00E+00	2.79E+03	3.76E+01	7.72E+01	5.96E+03

Note: JIF = journal impact factor of Journal Citation Report, MNCS = mean normalized citation score, SJR = SCImago Journal Rankings, SNIP = Source Normalized Impact per Paper.

Table 14
Rank correlation coefficient comparison.

	Coauthor count	H-index	Author citation counts	Cited-by counts	Document count
NSAP_JIF	0.1560	0.2868	0.3469	0.3460	0.2484
NSAP_CiteScore	0.1640	0.2940	0.3553	0.3545	0.2531
NSAP_SJR	0.1345	0.2790	0.3391	0.3392	0.2354
NSAP_SNIP	0.1202	0.2686	0.3267	0.3263	0.2313

Note: JIF = journal impact factor of Journal Citation Report, MNCS = mean normalized citation score, SJR = SCImago Journal Rankings, SNIP = Source Normalized Impact per Paper.

of raw citations that an author has obtained, while the cited-by author indicator represents how many times the researcher cited a special author. The document count is the number of publications that an author has published. As Table 13 presents, NSAP_SJR is a more even distribution than the NSAP score with the JIF, CiteScore, and SNIP indicators. The cited-by author indicator achieves the highest average count and highest standard deviation and variance, which indicates that the distribution of the cited-by counts are extremely uneven.

4.2. Performance evaluation

4.2.1. Indicator correlations

We ranked the authors based on the coauthor count, H-index, author citation count, cited-by counts, and document count. In addition, the authors ranked by the proposed indicator (NSAP) were used to calculate the correlation coefficient (i.e., the Pearson correlation coefficient) using each indicator. Table 14 lists the NSAP (JIF, CiteScore, SJR, and SNIP) ranking, which has a higher correlation with the author citation counts and cited-by count rank and lower correlation with the coauthor count and document count rank.

4.2.2. Indicator evaluation-DSPM dataset

We used the DSPM dataset, which was awarded annually between 1984–1993 and biennially afterward, as the gold standard. We ranked the scholars in the collected dataset as per each indicator. **A lower number represents a higher rank**. As most earlier articles do not have keywords or an abstract record, we only evaluated the DSPM awardees after 1993. As Table 15 demonstrates, the DSPM awardee ranked higher than the ranking assigned by the NSAP score. The score in red is the highest rank of each DSPM awardee. Table 17 indicates the accuracy of the DSPM awardees that are successfully retrieved in the top K results (Recall @ K).

It can be observed from Table 16 that the harmonic allocation model performs better than the factional allocation model with regards to the accuracy of the first 10, 100, and 500 results. By contrast, with regards to the accuracy of the first 1000 results, the recall tends to be similar to each other except for NSAP_SJR. Among the accuracy results of the harmonic allocation model, the NSAP_JIF turns out to achieve the best performance. As a result, the harmonic model is recognized as the most reasonable authorship credit allocation model combined with NSAP calculation.

After confirming the harmonic allocation model for NSAP, Table 17 lists the accuracy results of several variants of NSAP and some popular scholar indicators. These popular scholar indicators include coauthor count, H-index, author citation counts, cited-by counts, and document count. However, all these indicators perform poorly with regards to the accuracy of the first 10 and 100 ranked results. Among the first four variants of NSAP in Table 17, it can be seen that the recall is much higher when using JIF or CiteScore as the journal-level normalized indicator than when using SJR or SNIP. If the authorship credit allocation model is removed from NSAP, the recall result only shows a tiny increase with regards to the accuracy of the first 100 results, but decreases significantly with regards to the accuracy of the first 500 and 1000 results. Considering authorship credit allocation model or not seems to have very little effect on the performance of NSAP_SJR or NSAP_SNIP. If the publishing date is removed from NSAP, the recall results turn out to be obviously influenced, especially in terms of the performance with regards to accuracy of the first 10 ranked results. In addition, the accuracy decreases if the journal-level normalized indicator is eliminated from NSAP compared with those four variants of NSAP with JIF, Citescore, SJR, and SNIP as the journal-level normalized indicator.

Table 15
Indicator ranking comparison.

Year	Author	Scopus AU-ID	NSAP_ JIF	NSAP_ CiteScore	NSAP_ SJR	NSAP_ SNIP	Coauthor count	H-index	Author citation counts	Cited by counts	Document count
2019	Lutz Bornmann	8.85E+09	3	3	5	4	12432	1652	1642	2635	830
2017	Judit Bar-Ilan	5.72E+10	63	53	80	55	13323	3500	4460	4690	3808
2015	Mike Thelwall	5.54E+10	2	2	4	2	5323	737	832	1333	451
2013	Blaise Cronin	5.56E+10	808	888	1068	612	22315	4295	4275	4435	2571
2011	Persson	7E+09	2088	2183	2195	2662	26222	8965	7286	7183	14732
2009	Péter Vinkler	6.7E+09	446	407	550	439	30513	7139	9823	10823	10241
2009	Michel Zitt	7E+09	272	268	373	308	28929	10492	12646	13464	18566
2007	Katherine W. McCain	3.56E+10	2680	2587	2892	2062	25423	8323	6461	7401	11696
2005	Peter Ingwersen	7E+09	441	569	444	292	12011	7139	5774	6188	7836
2005	Howard D. White	5.63E+10	2437	2554	2687	2465	46740	25706	15089	14948	27667
2003	Loet Leydesdorff	7E+09	8	9	16	12	7886	499	594	875	553
2001	Ronald Rousseau	7.1E+09	101	101	120	91	7722	2921	2865	3257	1119
2001	Leo Egghe	5.72E+10	16578	16799	17536	17738	74104	67655	68568	68243	59081
1999	Wolfgang Glänzel	7E+09	58	63	79	57	7886	1039	1625	2373	1417
1999	Henk F. Moed	7E+09	33	29	45	43	17648	2278	2932	3270	5217
1995	Anthony F.J. Van Raan	7E+09	146	139	189	159	19601	1897	2536	3018	5150
1993	András Schubert	1.53E+10	457	471	524	572	19832	2740	2712	2854	3707

Note: JIF= journal impact factor of Journal Citation Report, MNCS= mean normalized citation score, SJR=SCImago Joumal Rankings, SNIP= Source Normalized Impact per Paper.

Table 16
Recall @ K computed by various authorship credit allocation models.

	Fractional				Harmonic			
	@10	@100	@500	@1000	@10	@100	@500	@1000
NSAP_JIF	0.1176	0.4118	0.5882	0.7647	0.1764	0.3529	0.7059	0.7647
NSAP_CiteScore	0.1176	0.4118	0.5882	0.7647	0.1764	0.3529	0.6471	0.7647
NSAP_SJR	0.1176	0.2941	0.5294	0.7647	0.1176	0.3529	0.5882	0.7059
NSAP_SNIP	0.1176	0.4118	0.5294	0.7647	0.1176	0.4118	0.6471	0.7647

*@10, @100, and @1000 are the accuracies of the first 10, 100, and 1000 ranked results, respectively.

Table 17
Recall @ K-The accuracies of DSPM awardees that are successfully retrieved in the top K results.

Recall	@10	@100	@500	@1000
NSAP_JIF	0.1764	0.3529	0.7059	0.7647
NSAP _CiteScore	0.1764	0.3529	0.6471	0.7647
NSAP_SJR	0.1176	0.3529	0.5882	0.7059
NSAP_SNIP	0.1176	0.4118	0.6471	0.7647
Coauthor count	0	0	0	0
H-index	0	0	0.0588	0.1176
Author citation counts	0	0	0.0588	0.1176
Cited-by counts	0	0	0	0.0588
Document count	0	0	0.0588	0.1764
NSAP_JIF_without_author credit allocation	0.1764	0.4118	0.5882	0.7059
NSAP_CiteScore_without_author credit allocation	0.1764	0.4118	0.5882	0.7059
NSAP_SJR_without_author credit allocation	0.1176	0.3529	0.5294	0.7059
NSAP_SNIP_without_author credit allocation	0.1764	0.4118	0.5294	0.7059
NSAP_JIF_without_published time	0	0.4706	0.7647	0.7647
NSAP_CiteScore_without_published time	0.0588	0.4706	0.7059	0.7647
NSAP_SJR_without_published time	0	0.3529	0.7059	0.7647
NSAP_SNIP_without_published time	0	0.4118	0.7647	0.8235
NSAP_without_field-normalized journal impact	0.1176	0.3529	0.5294	0.7059

Note: JIF = journal impact factor of Journal Citation Report, MNCS = mean normalized citation score, SJR = SCImago Journal Rankings, SNIP = Source Normalized Impact per Paper.
*@10, @100, and @1000 are the accuracies of the first 10, 100, and 1000 ranked results, respectively.

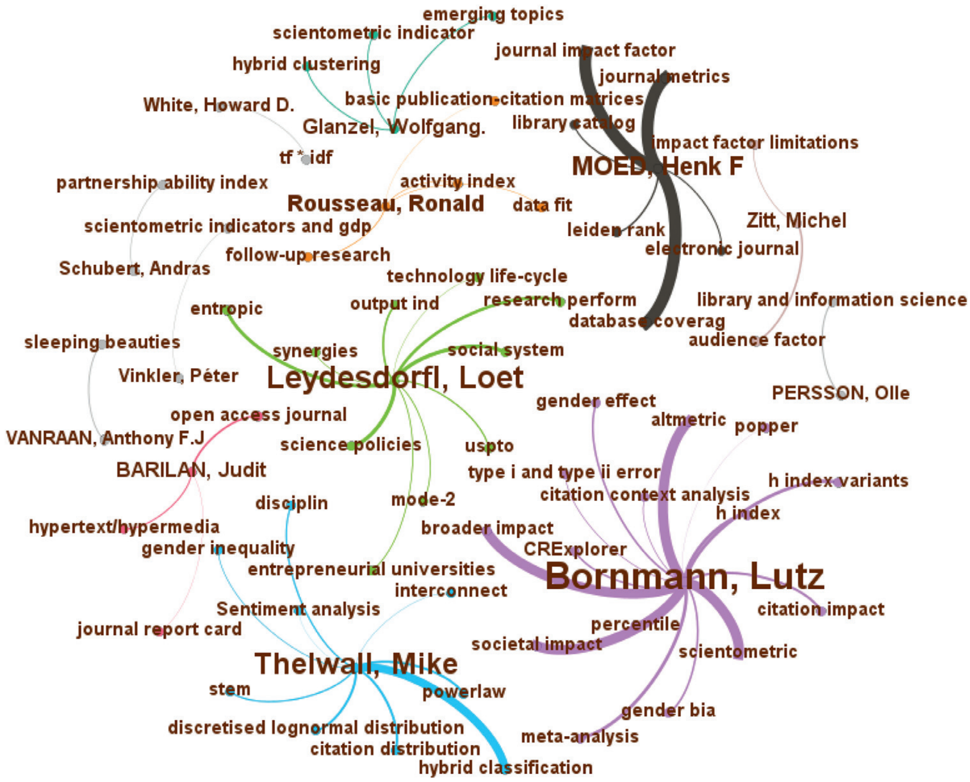

Fig. 8. Highest topic of DSPM awardee.

4.2.3. *Topic-based new indicator analysis-DSPM awardees*

For each topic, we record the highest NSAP score author and their related keyword under that topic. Fig. 8 illustrated the cases wherein the highest NSAP score authors are also the DSPM awardee in each topic. The weights of the edges between the author and topic term are the NSAP score.

We collected the nomination reason of each awardee; the nomination reason was proposed to summarize the awardees' achievements in the wide field of Informetrics and Scientometrics. Fig. 8 shows that Mike Thelwall is associated with many topic nodes. This is in line with the scientometric portrait proposed by Vellaichamy and Amsan (2016), which determined that Thelwall contributes to the research of citation analysis and sentiment analysis. Lutz Bornmann made and continues to make substantial contributions to a variety of important topics in bibliometrics (Daniel, 2019). The largest number of nodes is associated with Lutz Bornmann. Loet's outstanding articles are about the dynamics of innovation: from National Systems and mode-2 to a triple of university–industry–government relations, which are specially mentioned by Ronald Rousseau (2004). Fig. 8 indicates his achievements related to mode-2 and science policy. Ronald and Leo Egghe shared the award in the year 2001. Some of Ronald's research works are centered on the growth and obsolescence of literature, which could be seen from the node 'follow-up research'. Henk F. Moed won the 1999 Derek John de Solla Price Award for his special attention to the role of scientific journals, particularly the assessment of the journals' status, where we can see that the links between journal impact factor, journal metrics, database coverage, and Henk F. Moed are strengthened. Wolfgang Glänzel was an awardee at the same time. Wolfgang is a mathematician; specifically, a statistician, interested in stochastic models of citation processes and characteristics of skewed citation distributions, because of which the topic nodes 'scientometric indicator' and 'hybrid clustering' connect to him. Michel Zitt's professional background is in engineering. His research on field normalization issues belongs to what are arguably his most important contributions to the advancement of this field. The node 'impact factor limitations' is consistent with the award reason. van Raan named 'sleeping-beauties' in scientific literature. He is identified with the node 'sleeping beauties' in Fig 8. Thus, Fig 8 shows a high similarity between the reasons espoused by NSAP and those based on which the award is presented. Thus, the NSAP results match those of Price awardee identification.

We also picked certain popular topics in LIS. These popular topics are defined using the frequency of author keyword. Table 18 demonstrates the NSAP_JIF score under each topic in the collected dataset. The number in the bracket is the frequency of the keyword. A darker background indicates that the author is the DSPM awardee. We can see that the authors ranked 1 are varied between clusters, which indicates that scholars excel in different areas. Specifically, it turns out that Mike Thelwall contributes outstanding works on citation analysis as well as scientometrics. Bornmann, Lutz, as the most recent DSPM awardee, shows his powerful academic competence on more than one topic cluster. Garfield, Eugene, as the founder of the SCI index, demonstrates his competence on bibliometric works.

Table 18
Top 5 authors with highest NSAP scores under five popular topics.

Rank	Author ID	(Author name)	NSAP_JIF
Topic: Bibliometrics (859) related topic_Cluster			
1	8850037200	Bornmann, Lutz	415.72
2	22833445200	Abramo, Giovanni	361.24
3	7005088140	Garfield, Eugene	320.46
4	55396590500	Thelwall, Mike	195.20
5	56962739400	D'Angelo, Ciriaco Andrea	158.19
Topic: citation analysis (547) related topic_Cluster			
1	55396590500	Thelwall, Mike	357.64
2	14632830700	Waltman, Ludo	98.12
3	35229200000	Ding, Ying	85.87
4	8850037200	Bornmann, Lutz	78.71
5	55933111000	Kousha, Kayvan	74.86
Topic: social network (450) related topic_Cluster			
1	7003581501	Leonardi, Paul M.	135.72
2	19638629100	Baumgartner, Jody C.	108.82
3	57193690446	Hassan, Lobna	107.48
4	35239818900	Dwivedi, Yogesh K.	103.39
5	6602670816	Shiau, Wen-Lung	101.88
Topic: collaboration (433) related topic_Cluster			
1	35361989600	Hamari, Juho	236.88
2	56318870900	Sjoklint, Mimmi	118.44
3	35865769000	Abbasi, Alireza	112.40
4	23393982600	Ukkonen, Antti	78.96
5	29567519800	Pereira, Jorge Verissimo	68.42
Topic: scientometric (294) related topic_Cluster			
1	55396590500	Thelwall, Mike	199.34
2	8850037200	Bornmann, Lutz	124.75
3	56049501000	Martin-Martin, Alberto	100.46
4	23035818200	Orduna-Malea, Enrique	50.23
5	7004502597	Fairclough, Ruth	38.32

4.2.4. Awardees of the association for information science and technology honor

To validate the NSAP indicator, the names of the Award of Merit recipients and Research in Information science awardees of the Association for Information Science and Technology (ASIS&T) are collected. This award is a lifetime achievement award that recognizes sustained contributions to and/or achievements in the field of information science and/or the professions in which it is practiced.

The Award of Merit[1] of ASIS&T was established in 1964. To select the awardees in our dataset, Scopus author IDs were matched manually and 28 authors since 1993 were considered (listed in Appendix 1).

Similarly, Research in Information science awardees[2] of ASIS&T since 1993 were collected (listed in Appendix 2). This award recognizes an individual or team who have made an outstanding contribution to information science research. The NSAP performance in terms of Merit awardees and information science awardees is listed in Table 18.

First, for the identification of Merit awardees, it can be observed from Table 18 that with regards to the accuracy of first 10 and 100 ranked results, NSAP and popular scholar indicators perform poorly. However, NSAP shows good performance with regards to the accuracy of first 500 and 1000 ranked results, whereas popular scholar indicators still show poor performance.

Second, for the identification of information science awardees, it can be observed from Table 19 that the popular scholar indicators perform poorly as in the previous case, whereas the NSAP performance gradually improved with regards to the accuracy of the first 100, 500, and 1000 ranked results. One of the potential factors affecting the performance of NSAP is that information science awardees are rewarded for a systematic "program of research" in a single area at a level beyond the single study, but not

[1] https://www.asist.org/programs-services/awards-honors/award-of-merit/
[2] https://www.asist.org/programs-services/awards-honors/research-award/

Table 19
Recall @ K-The accuracies of merit awardees and information science awardees that are successfully retrieved in the top K results.

	Merit awardees				information science awardees			
Recall	@10	@100	@500	@1000	@10	@100	@500	@1000
NSAP_JIF	0	0	0.0714	0.2143	0	0.0385	0.1538	0.2692
NSAP _CiteScore	0	0	0.0357	0.1429	0	0.0385	0.0769	0.2692
NSAP_SJR	0	0	0.1071	0.1786	0	0.0769	0.1923	0.3077
NSAP_SNIP	0	0	0.1429	0.3929	0	0.1054	0.2692	0.4231
Coauthor count	0	0	0	0	0	0	0	0
H-index	0	0	0	0	0	0	0.0385	0.0385
Author citation counts	0	0	0	0	0	0	0	0.0385
Cited-by counts	0	0	0	0	0	0	0	0.0385
Document count	0	0	0	0	0	0	0	0.0385
NSAP_JIF_without_author credit allocation	0	0	0	0.1786	0	0.0385	0.1054	0.3077
NSAP_CiteScore_without_author credit allocation	0	0	0	0.1786	0	0.0385	0.1054	0.3077
NSAP_SJR_without_author credit allocation	0	0	0	0.1429	0	0.0385	0.1054	0.2308
NSAP_SNIP_without_author credit allocation	0	0	0.0714	0.1786	0	0.0385	0.2308	0.3077
NSAP_JIF_without_published time	0	0	0.2857	0.3929	0	0.0769	0.3077	0.5385
NSAP_CiteScore_without_published time	0	0	0.2143	0.3929	0	0.0769	0.2692	0.5
NSAP_SJR_without_published time	0	0	0.2857	0.4286	0	0.0769	0.3077	0.5385
NSAP_SNIP_without_published time	0	0	0.1071	0.1786	0	0.1054	0.5	0.5
NSAP_without_field-normalized journal impact	0	0	0.0357	0.1071	0	0.0385	0.1054	0.2308

Note: JIF = journal impact factor of Journal Citation Report, MNCS = mean normalized citation score, SJR = SCImago Journal Rankings, SNIP = Source Normalized Impact per Paper.
*@10, @100, and @1000 are the accuracies of the first 10, 100, and 1000 ranked results, respectively.

at the level of a lifetime's work. To improve the applicability of NSAP, the honor awarding time might need to be considered when identifying recipients in future work.

5. Discussion

Among the four variants of NSAP, four journal quality indicators (JIF, CiteScore, SJR, and SNIP) were used to produce the NSAP score. Table 17 and Table 19 reveals that regardless of which field-normalized journal quality is integrated into the NSAP indicator, the accuracies of the first 10, 100, and 1000 ranked results is better than those for the raw citation count-based indicators (H-index, author citation counts, cited-by counts). Furthermore, the design formula for JIF and CiteScore has a similar form. One of the underlying reasons causing the differences in the NSAP_JIF and NSAP_CiteScore values may be that CiteScore considers three years of data compared with JIF, which considers only two years. In addition, CiteScore considers not only articles and reviews but also editorials, letters, and news items in its equation, which JIF does not. CiteScore was calculated using Scopus data, which has more than double the titles covered by JIF with WoS data (da Silva & Memon, 2017).

The awardees of DSPM and ASIS&T were used as the gold standard to evaluate the validity of the NSAP indicator for calculating scholar performance assessment. The results demonstrated the outstanding topic-related scholar performance assessment properties of the NSAP indicator compared to raw citation count indicators. However, some high-ranked authors are not DSPM or ASIS&T winners, such as Dwivedi, Yogesh (35239818900), Gandomi, Amir (56788235100), and Sultan, Nabil Ahmed (16450198300). Authors Yogesh, Amir, and Nabil Ahmed ranked first, fourth, and fifth respectively as per NSAP_JIF score. Yogesh has fifteen publications in the International Journal of Information Management (which has the highest JIF in LIS). Amir ranked high because he once published a paper as the first author in the International Journal of Information Management and achieved 3430 citations. Sultan, Nabil Ahmed has seven publications (first author) in the International Journal of Information Management and has a total of 2,424 citations. However, given that none of these researchers is mainly focused on research in informetrics or information science, it is unlikely that they will be given the DSPM or ASIS&T award.

Furthermore, NSAP requires less computational overhead, making it convenient for various applications. All journal quality indicators could be freely obtained from WoS or Scopus. Keywords and raw citation counts are the basic metadata of publication records in the literature database. Scopus make the calculation process easier and more accurate by providing author name ambiguation using the Scopus author ID. No initial parameters are required for the calculation process. In this work, the results indicate that NSAP performs better than coauthor count, H-index, author citation counts, cited-by counts, and document count when used to identify DSPM awardees and ASIS&T honor awardees. In particular, for DSPM awardees identification, the award provides detailed scientific profiles for awardees about their contributions. NSAP is topic-dependent and effectively recognizes top scientists under specific research topics.

6. Conclusion

Raw citation-based scholar performance indicators lack cross-field comparability in practical evaluation applications and have difficulty determining the contributions of multi-author publications. This study proposed a new indicator, called the NSAP indicator, to assess scholar academic ability. This indicator considers the research topic; thus, we can compare the authors' academic ability under each topic. In addition, for each author, we considered the published date of each paper under the special topic, the author's role and sequence, and the journal quality.

The DSPM dataset was used as the gold standard to evaluate the validity of the NSAP indicator for calculating scholar performance assessment. The results demonstrated the outstanding topic-related scholar performance assessment properties of the NSAP indicator compared to raw citation count indicators. The results of comparing DSPM awardees and their highest NSAP topic indicates a high similarity with the award criteria. We also analyzed the highest-score scholars for the five most popular topics in our collected dataset. Further, a collection of the awardees of the ASIS&T is used to evaluate the performance for validating the NSAP indicator in other research fields within the LIS discipline.

The main limitation of this work is that the author's influence does not consider the topic-level weighted citation impact. Topic-level weighted citation impact takes into account the fact that a paper cited by a journal with a high JIF should have more weight than a low-impact journal. In the future, we will work on fine-grained scholar performance indicators by considering the prestige of the citing articles. Furthermore, only journal papers from LIS disciplines were considered in this study. Future research should focus on verifying the applicability of the NSAP in datasets of other disciplines and consider conference papers and monographs as well.

Author contributions

Qing Xie: Conceived and designed the analysis, Collected the data, Contributed data or analysis tools, Performed the analysis, Wrote the paper. **Xinyuan Zhang**: Conceived and designed the analysis, Collected the data, Contributed data or analysis tools, Performed the analysis, Wrote the paper. **Min Song**: Conceived and designed the analysis, Collected the data, Contributed data or analysis tools, Performed the analysis, Wrote the paper.

Acknowledgments

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2020S1A5B1104865)

Appendix 1

Table A1

Table A1
Award of merit of the association for information science and technology.

Year	Award of Merit Recipients	ID
2020	Diane H. Sonnenwald	6603988432
2019	Christine Borgman	7006568281
2018	Toni Carbo	6602083535
2017	Tom Wilson	7403496067
2016	Peter Ingwersen	7003444524
2015	Michael E.D. Koenig	7201819214
2014	Marjorie M.K. Hlava	6602970486
2013	Carol C. Kuhlthau	6602297186
2012	Michael Buckland	7003635510
2011	Gary Marchionini	7004087155
2010	Linda C. Smith	7410393566
2009	Carol Tenopir	7005106498
2008	Clifford Lynch	7202499263
2007	Donald H. Kraft	7102252611
2006	Blaise Cronin	55605719900
2005	Marcia Bates	7201434341
2004	Howard D. White	54790591200
2003	Nicholas J. Belkin	7102357181
2002	Karen Sparck Jones	6603497944
2001	Patrick G. Wilson	35479590200
2000	Donald R. Swanson	7201859819
1999	Jose Marie Griffiths	7401481776
1998	Henry Small	7003844054
1997	Dagobert Soergel	6701699377
1996	Jean Tague-Sutcliffe	6602478557
1995	Tefko Saracevic	56238291900
1994	Harold Borko	7003977709
1993	Robert M. Hayes	7402082398

Appendix 2

Table A2

Table A2
Research in information science award recipients of the association for information science and technology.

Year	Research in Information Science Award Recipients	ID
2020	Pertti Vakkari	6603768172
2019	Kevin Crowston	6701434689
2018	Judit Bar-Ilan	57202052068
2017	Caroline Haythornthwaite	6701453477
2016	Reijo Savolainen	7003329071
2014	Diane Kelly	35581749900
2013	Susan Herring	7005704556
2012	Kalervo Jarvelin	7003932928
2011	Susan Leigh Star	6701370611
2010	Christine Borgman	7006568281
2007	Ophir Frieder	7006557167
2006	Brenda Dervin	15756962000
2005	Carol Kuhlthau	6602297186
2004	Boyd Rayward	6602875492
2003	Peter Ingwersen	7003444524
2002	Carol Tenopir	7005106498
2001	Paul Kantor	57207563803
2000	W. Bruce Croft	7006788293
1999	Donald H. Kraft	7102252611
1999	David C. Blair	7202955593
1998	Marcia Bates	7201434341
1997	Nicholas J. Belkin	7102357181
1996	Gary Marchionini	7004087155
1995	Charles T. Meadow	6701413936
1994	Raya Fidel	6603863731
1993	Howard D. White	54790591200

References

Amjad, T., Daud, A., Che, D., & Akram, A. (2016). Muice: Mutual influence and citation exclusivity author rank. *Information Processing and Management, 52*(3), 374–386.
Assimakis, N., & Adam, M. (2010). A new author's productivity index: P-index. *Scientometrics, 85*(2), 415–427.
Bai, X., Liu, H., Zhang, F., Ning, Z., Kong, X., Lee, I., et al. (2017). An overview on evaluating and predicting scholarly article impact. *Information, 8*(3), 73–86.
Behrouzi, S., Sarmoor, Z. S., Hajsadeghi, K., & Kavousi, K. (2020). Predicting scientific research trends based on link prediction in keyword networks. *Journal of Informetrics, 14*(4), Article 101079.
Berk, R. A. (1986). *Performance assessment: Methods & applications*. Johns Hopkins University Press.
Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3*, 993–1022.
Bollen, J., Van de Sompel, H., Hagberg, A., & Chute, R. (2009). A principal component analysis of 39 scientific impact measures. *PloS one, 4*(6), Article e6022.
Bornmann, L., & Williams, R. (2013). How to calculate the practical significance of citation impact differences? An empirical example from evaluative institutional bibliometrics using adjusted predictions and marginal effects. *Journal of Informetrics, 7*(2), 562–574.
Bornmann, L., Marx, W., & Barth, A. (2013). The normalization of citation counts based on classification systems. *Publications, 1*(2), 78–86.
Braam, R. R., Moed, H. F., & Van Raan, A. F. (1991). Mapping of science by combined co-citation and word analysis. I. Structural aspects. *Journal of the American Society for information science, 42*(4), 233–251.
da Silva, & Memon, A. R. (2017). CiteScore: A cite for sore eyes, or a valuable, transparent metric? *Scientometrics, 111*(1), 553–556.
Daniel, H.-D. (2019). Lutz Bornmann: Recipient of the 2019 Derek John de Solla Price Medal. *Scientometrics, 121*(3), 1235–1238 Springer; Akadémiai KiadóDecember.
Douglass, R. (1992). A cluster-based approach to browsing large document collections. In *Proceedings of the Fifteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 318–329).
Duan, L., Xu, L., Guo, F., Lee, J., & Yan, B. (2007). A local-density based spatial clustering algorithm with noise. *Information systems, 32*(7), 978–986.
Egghe, L., Rousseau, R., & Van Hooydonk, G. (2000). Methods for accrediting publications to authors or countries: Consequences for evaluation studies. *Journal of the American Society for Information Science, 51*(2), 145–157.
Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics, 69*(1), 131–152.
Eom, S. (2008). All author cocitation analysis and first author cocitation analysis: A comparative empirical investigation. *Journal of Informetrics, 2*(1), 53–64.
Frey, B. J., & Dueck, D. (2007). Clustering by passing messages between data points. *Science, 315*(5814), 972–976.
Garfield, E., & Welljams-Dorof, A. (1992). Of Nobel class: A citation perspective on high impact research authors. *Theor Med Bioeth, 13*, 117–135. 10.1007/BF02163625.
Garfield, E. (1999). Journal impact factor: A brief review. *Canadian Medical Association Journal, 161*(8), 979–980.
Ghaemi, Z., & Farnaghi, M. (2019). A varied density-based clustering approach for event detection from heterogeneous twitter data. *ISPRS International Journal of Geo-Information, 8*(2), 82.
Grover, A., & Leskovec, J. (2016). node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 855–864).
Hagen, N. T. (2008). Harmonic allocation of authorship credit: Source-level correction of bibliometric bias assures accurate publication and citation analysis. *PLoS ONE, 3*(12), e4021.
Hagen, N. T. (2010). Harmonic publication and citation counting: Sharing authorship credit equitably—Not equally, geometrically or arithmetically. *Scientometrics, 84*(3), 785–793.
Hagen, N. T. (2013). Harmonic coauthor credit: A parsimonious quantification of the byline hierarchy. *Journal of Informetrics, 7*(4), 784–791.

Haunschild, R., Leydesdorff, L., Bornmann, L., Hellsten, I., & Marx, W. (2019). Does the public discuss other topics on climate change than researchers? A comparison of explorative networks based on author keywords and hashtags. *Journal of Informetrics, 13*(2), 695–707.

Hirsch, J. E. (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), 16569–16572.

Hodge, S. E., Greenberg, D. A., & Challice, C. E. (1981). Publication credit. *Science, 213*(4511), 950.

Hoeffel, C. (1998). *Journal impact factors. Allergy, 53*(12) 1225-1225.

Hu, J., & Zhang, Y. (2015). Research patterns and trends of Recommendation System in China using co-word analysis. *Information processing & management, 51*(4), 329–339.

Jiang, X., Sun, X., & Zhuge, H. (2012). Towards an effective and unbiased ranking of scientific literature through mutual reinforcement. Proceedings of the 21st ACM international conference on Information and knowledge management, 714-723.

Rosen-Zvi, M., Griffiths, T., Steyvers, M., & Smyth, P. (2012). The author-topic model for authors and documents. arXiv preprint arXiv:1207.4169.

Kalepalli, Y., Tasneem, S., Teja, P. D. P., & Manne, S. (2020). Effective comparison of LDA with LSA for topic modelling. In *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 1245–1250). IEEE.

Kastrati, Z., Imran, A. S., & Yayilgan, S. Y. (2019). The impact of deep learning on document classification using semantically rich representations. *Information Processing & Management, 56*(5), 1618–1632.

Khasseh, A. A., Soheili, F., Moghaddam, H. S., & Chelak, A. M. (2017). Intellectual structure of knowledge in iMetrics: A co-word analysis. *Information processing & management, 53*(3), 705–720.

Kim, J., & Diesner, J. (2014). A network-based approach to coauthorship credit allocation. *Scientometrics, 101*(1), 587–602.

Kim, J., & Kim, J. (2015). Rethinking the comparison of coauthorship credit allocation schemes. *Journal of Informetrics, 9*(3), 667–673.

Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents. In *International conference on machine learning* (pp. 1188–1196). PMLR.

Li, Y., Wu, C., Wang, X., & Luo, P. (2014). A network-based and multi-parameter model for finding influential authors. *Journal of Informetrics, 8*(3), 791–799.

Liu, X. Z., & Fang, H. (2012). Fairly sharing the credit of multi-authored papers and its application in the modification of h-index and g-index. *Scientometrics, 91*(1), 37–49. 10.1007/s11192-011-0571-y.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, 3111–3119.

Milojević, S., Sugimoto, C. R., Yan, E., & Ding, Y. (2011). The cognitive structure of library and information science: Analysis of article title words. *Journal of the American Society for Information Science and Technology, 62*(10), 1933–1953.

Moed, H., De Bruin, R., & Van Leeuwen, T. H. (1995). New bibliometric tools for the assessment of national research performance: Database description, overview of indicators and first applications. *Scientometrics, 33*(3), 381–422.

Moed, H. F. (2010). Measuring contextual citation impact of scientific journals. *Journal of Informetrics, 4*(3), 265–277.

Murtagh, F., & Contreras, P. (2017). Algorithms for hierarchical clustering: An overview, II. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 7*(6), e1219.

Oppenheim, C. (1998). Fractional counting of multiauthored publications. *Journal of the American Society for Information Science, 49*(5), 482.

Osório, A. (2018). On the impossibility of a perfect counting method to allocate the credits of multi-authored publications. *Scientometrics, 116*(3), 2161–2173.

Persson, O. (2001). All author citations versus first author citations. *Scientometrics, 50*(2), 339–344.

Price, D. D. (1981). Multiple authorship. *Science, 212*(4498), 986.

Radicchi, F., Fortunato, S., Markines, B., & Vespignani, A. (2009). Diffusion of scientific credits and the ranking of scientists. *Physical Review E, 80*(5), Article 056103.

Rahman, M. T., Mac Regenstein, J., Kassim, N. L. A., & Haque, N. (2017). The need to quantify authors' relative intellectual contributions in a multi-author paper. *Journal of Informetrics, 11*(1), 275–281.

Reddy, M., Makara, V., & Satish, R. U. V. N. (2017). Divisive hierarchical clustering with K-means and agglomerative hierarchical clustering. *International Journal of Computer Science Trends and Technology (IJCST), 5*(5), 5–11.

Rousseau, R. (2004). Loet Leydesdorff: Recipient of the 2003 Derek de Solla Price Award. *Scientometrics, 60*(3), 275–277.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics, 20*, 53–56.

Sayyadi, H., & Getoor, L. (2009). April). Futurerank: Ranking scientific articles by predicting their future pagerank. In *Proceedings of the 2009 SIAM International Conference on Data Mining* (pp. 533–544). Society for Industrial and Applied Mathematics.

Sembiring, R. W., Zain, J. M., & Embong, A. (2011). A comparative agglomerative hierarchical clustering method to cluster implemented course. arXiv preprint arXiv:1101.4270.

Shen, H. W., & Barabási, A. L. (2014). Collective credit allocation in science. *Proceedings of the National Academy of Sciences, 111*(34), 12325–12330.

Simoes, N., & Crespo, N. (2020). A flexible approach for measuring author-level publishing performance. *Scientometrics, 122*(1), 331–355.

Steinbach, M., Karypis, G., & Kumar, V. (2000). A comparison of document clustering techniques.

Strotmann, A., & Zhao, D. (2012). Author name disambiguation: What difference does it make in author-based citation analysis? *Journal of the American Society for Information Science and Technology, 63*(9), 1820–1833.

Thelwall, M., & Fairclough, R. (2015). The influence of time and discipline on the magnitude of correlations between citation counts and quality scores. *Journal of Informetrics, 9*(3), 529–541.

Trueba, F. J., & Guerrero, H. (2004). A robust formula to credit authors for their publications. *Scientometrics, 60*(2), 181–204.

Vellaichamy, A., & Amsan, E. (2016). Scientometric portrait of Mike Thelwall. *Library Philosophy and Practice*. (e-journal). 1487 http://digitalcommons.unl.edu/libphilprac/1487 .

Waltman, L. (2012). An empirical analysis of the use of alphabetical authorship in scientific publishing. *Journal of Informetrics, 6*(4), 700–711.

Wray, K. B. (2002). The epistemic significance of collaborative research. *Philosophy of Science, 69*(1), 150–168. 10.1086/338946.

Yan, E., Ding, Y., & Sugimoto, C. R. (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.

Zhang, Y., Lu, J., Liu, F., Liu, Q., Porter, A., Chen, H., & Zhang, G. (2018). Does deep learning help topic extraction? A kernel k-means clustering method with word embedding. *Journal of Informetrics, 12*(4), 1099–1117.

Zhang, F. (2017). Evaluating journal impact based on weighted citations. *Scientometrics, 113*(2), 1155–1169.

Zitt, M., Ramanana-Rahary, S., & Bassecoulard, E. (2005). Relativity of citation performance and excellence measures: From cross-field to cross-scale effects of field-normalisation. *Scientometrics, 63*(2), 373–401.