A weighted unified informetrics based on Scopus and WoS

Parul Khurana · Geetha Ganesan · Gulshan Kumar · Kiran Sharma

Received: date / Accepted: date

Abstract Numerous indexing databases keep track of the number of publications, citations, etc. in order to maintain the progress of science and individual. However, the choice of journals and articles varies among these indexing databases, hence the number of citations and $h$-index varies. There is no common platform exists that can provide a single count for the number of publications, citations, $h$-index, etc. To overcome this limitation, we have proposed a weighted unified informetrics, named “conflate”. The proposed system takes into account the input from multiple indexing databases and generates a single output. Here, we have used the data from Scopus and WoS to generate a conflate dataset. Further, a comparative analysis of conflate has been performed with Scopus and WoS at three levels: author, organization, and journal. Finally, a mapping is proposed between research publications and distributed ledger technology in order to provide a transparent and distributed view to its stakeholders.

Keywords Bibliometrics · Pay off matrix · Conflate · Distributed ledger technology

Parul Khurana
School of Computer Applications, Lovely Professional University, Phagwara, Punjab-144401, India
E-mail: parul.khurana@lpu.co.in

Geetha Ganesan
Advanced Computing Research Society, Porur, Chennai-600116, India
E-mail: gitaskumar@yahoo.com

Gulshan Kumar
School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab-144401, India
E-mail: gulshan3971@gmail.com

Kiran Sharma
School of Engineering and Technology, BML Munjal University, Gurugram, Haryana-122413, India
E-mail: kiran.sharma@bmu.edu.in
1 Introduction

Bibliometrics refers to the use of quantitative measures and statistical methods to analyze the impact of scientific contributions (Umate, Patil, Telrandhe, and Pathade, 2020). It helps in describing publication patterns, measuring the research quality and impact, and determining the influence of authors with different authors involved in publications (Daim, Rueda, Martin, and Gerdsrij 2006) [Archambault, Campbell, Gingras, and Larivière 2009] [Albert-Morant, Leal-Rodríguez, Fernández-Rodríguez, and Ariza-Montes 2018]. Scientific contributions work as a driving force for the continuous growth of science and society (Van Raan 2003) [Janowicz, Regalia, Hitzler, Mai, Delbecque, Fröhlich, Martiment, and Lazarus 2018]. These contributions receive citations which are considered as an important indicator of influence in the research publications process and are explored and studied widely in this digital world (Prigge 2008) [Sharma and Khurana 2020]. Citations provide a quantitative evaluation to measure research outputs, utilities, emergence, and extent of scientific communication to all the interested parties (Meho and Sugimoto 2009) [Hoffman, Ibáñez, and Simperl 2019]. Citation data is often impacted by the coverage of specific bibliographic databases because they collect the citations received by the publications indexed by them (Bar-Ilan, Levene, and Lin 2007). Such bibliographic databases offer scientific activities, number of publications, number of citations, and quality, importance, and impact of publications and their peer evaluation for bibliometric analysis (Martin 1996) [Abramo and D’Angelo 2011].

Bibliographic databases like Elsevier’s Scopus and Thomson Reuter’s WoS are being used worldwide to produce comparative statistics for their bibliometricians. Both are working continuously in this dynamic field so that required informetrics may be empowered with their databases (Bakkalbasi, Bauer, Glover, and Wang 2006) [Neuhaus and Daniel 2008] [Adriaanse and Rensleigh 2011] [Franceschini, Maisano, and Mastrogiacomo 2016]. Comparative statistics provided by bibliographic databases are used by funding agencies, government bodies, promotion committees, and other stakeholders to measure the quality and impact of authors (Meho and Yang 2006) [Martín-Martín, Orduna-Malea, Thelwall, and López-Cózar 2018]. Hence, bibliometric analysis has emerged as a powerful tool and partial system for ranking and accreditation agencies as well (Alvarez-Melgarejo and Torres-Barreto 2018) [Oliveira, de Barros, de Carvalho Pereira, Gomes, and da Costa 2018].

Different authors have shown the comparison of Scopus and WoS in various literatures for the coverage of journals (Mongeon and Paul-Hus 2016) [Liu, Huang, and Wang 2021], quality of content (Vafaeian, Noroozi Chakoli, and Hassanzadeh 2011), bibliometric statistics (De Groot and Raszewski 2012) [Bartol and Mackiewicz-Talarczyk 2015], subject categorization (Powell and Peterson 2017), features and capabilities (Thelwall 2018) [Martín-Martín, Orduna-Malea, Thelwall, and Delgado-López-Cózar 2019], citations counts, tracking and indexing (Jacso 2005) [Vaughan and Shaw 2008] [Li, Burnham, Lemley, and Britton 2010], author, university and country rankings
A weighted unified informetrics based on Scopus and WoS (López-Illescas, de Moya Anegón, and Moed, 2009), language coverage (Bartol, Budimir, Dekleva-Smrekar, Pusnik, and Juznic, 2014), longitudinal and cross-disciplinary comparisons (Norris and Oppenheim, 2007), content comprehensiveness and searching capabilities (Burnham, 2006; Aghaei Chadegani, Salehi, Yunus, Farhadi, Fooladi, Farhadi, and Ale Ebrahim, 2013). The prolific growth of both bibliographic databases have created new opportunities in terms of publications, citations and bibliometric indicators (Adriaanse and Rensleigh, 2013) as well.

Due to the availability of the number of bibliographic databases, interviewers and hiring agencies ask authors to provide publications, citations, and $h$ index count - bibliographic databases wise. This situation in India has raised a requirement of unified informetrics where an author can provide a single count for his/her bibliometric statistics across different bibliographic databases. Hence, we conducted this study to calculate a weighted unified informetics at three levels such as author, organization, and journal. We have used bibliographic databases such as Scopus and WoS due to their indexing age, availability of data, and authenticity. The objectives of our study are:

1. To propose a common platform that can provide a single article count, citation count, and $h$-index in the education field.
2. To check the statistical validity of the proposed platform in terms of the number of articles, the number of citations, and $h$-index at author, organization, and journal level.
3. Technological enhancement of the proposed platform by use of distributed ledger technology.

Further, the study is organized as follows: Section 2 describes data and methods including data selection, data filtration, data extraction, data analysis, and concept of pay-off matrix. Results are explained in Section 3. Section 4 describes the technological enhancement with the concept of distributed ledger technology. Finally, summary is given in Section 5.

2 Data and methods

2.1 Data selection and filtration

Data selection is performed at three levels:

- **Author level**: 400 faculty profiles out of 6316 profiles from various disciplines are accessed from “Monash University”, a public university in Melbourne, Australia (Monash, 2020). The choice of the data is mainly due to the openly available information of faculties especially ORCID ID, Scopus ID, and WoS ID. All 400 selected faculty profiles have an ORCID ID, Scopus ID, and WoS ID.

- **Organization level**: Top 100 Indian institutes out of 200 ranked institutions based on National Institutional Ranking Framework (NIRF) (MHRD, 2020) are accessed.
– **Journal level:** A random selection of 1000 journals listed in both Scopus and WoS is used.

Data filtration is performed on the basis of DOI (Digital Object Identifier). It provides a unique authentication to the publications [Simmonds 1999; Chandrakar 2006]. It’s desirable that articles carrying DOI numbers must be considered only for any kind of evaluation and calculation to establish scientific assignments [Gorraiz, Melero-Fuentes, Gumpenberger, and Valderrama-Zuriañ 2016]. Hence, while combining articles across multiple databases, we have considered the articles with DOI numbers only (see flowchart in Figure 1). To perform this study, data from Scopus and WoS was obtained with Python-based APIs [Michael-E-Rose 2019; Bacis 2019]. The primary reason for data selection from Scopus and WoS is arbitrary and the availability of data.

### 2.2 Data extraction and analysis

For data extraction from both Scopus and WoS, we require three inputs at three levels. An ORCID ID is required for authors’ information, organization name for university/institute access, and ISSN for journal information. For a given author, based on his/her ORCID ID, we retrieved the number of publications from both Scopus and WoS. Further, we filtered those publications based on the DOI and got the required data set for the given author for both Scopus and WoS. The next step is to look after the number of citations received by the given author for each publication. We examined each and every citation received on all publications for a given author and filtered only those citations that have DOI associated to them. The existence of DOI is necessary to condition in order to match the given record among multiple databases. The collection of similar records and respective citations based on DOI for a given author among Scopus and WoS will lead to a new filtered database named “conflate”. Algorithm 1 describes the steps for extracting article and citation details for the given ORCID ID. Similarly, the process is repeated at the organization and the journal levels. The conflate database of citations generated in algorithm 1 is used further for the analysis. First, for a given author, we filtered the *common* and the *unique* citations (publications records) from Scopus and WoS. Then, a weight is assigned to both common and unique citations (see section 2.3). These new weighted citations will be our conflate data for a given author. Algorithm 2 describes the process in sequence.
**Fig. 1** Flowchart demonstrate the process of visiting the author’s, organization’s and journal’s profile.
Algorithm 1: Generation of conflate based citations database

**Input:** ORCID ID, ISSN, or Organization Name  
**Output:** Conflated citations from N databases  

```plaintext
/* Check number of articles in N databases */
1 while N > 0 do  
    A[1], A[2], ..., A[N] ∈ DB[1], DB[2], ..., DB[N], ∀doi exists  
    where A[1] represents articles doi list for DB[1]
    /* Find doi of each citation */
2 for each doi in [A[1], A[2], ..., A[N]] do  
    N_c := number of citations
    for each N_c do  
      if doi exists then  
        CD[1] := doi  
        // CD is citer’s doi
      else  
        skip
3 Repeat step 2 for all A[1], A[2], ..., A[N] and get CD[1], CD[2], ..., CD[N]  
4 CD := CD[1] ∪ CD[2] ∪ ... ∪ CD[N]  
    where CD contains only those citations for a given ORCID ID whose doi’s exists (including common doi’s in all DB’s)
```

Algorithm 2: Generation of h index and related indicators

**Input:** Conflate based citations from N databases for a given ORCID ID, ISSN, or Organization Name  
**Output:** Updated h index and related indicators  
**Data:** List of citer’s doi’s, CD[1], A[N]  

```plaintext
1 common := CD[1] ∩ CD[2] ∩ ... ∩ CD[N]  
    // get the common doi’s in all DB
2 unique := CD − common  
    // only in individual DB’s  
3 Compute weightage of all doi’s  
    // Give full weightage, i.e.1 to doi’s in common set (in step 1)
    S1 = Number of doi’s in common  
    // Give \( \frac{1}{n} \) weightage to doi’s in unique set (in step 2)
    S2 = Number of doi’s in unique \( \times \frac{1}{n} \)
    // Final weighted citation  
    S = ceil(S1 + S2)  
4 C = unique(CD)  
    /* Find unique doi’s in CD */
5 Compute h-index and related indicators on C
```
2.3 Payoff matrix

For ranking the universities, various agencies like NIRF and Times Higher Education (THE) use databases of their choice like THE (2020) use Scopus database and NIRF gives equal weightage to all databases (MHRD, 2020). As all bibliographic databases carry their own special features and fields of information, we cannot decide that one indexing database is better than others (Papić, 2017; Rose and Kitchin, 2019; Yelne, Umate, and Patil, 2021; Khu-rana and Sharma, 2021). Hence, it is not appropriate to give more weightage to any specific database over others. However, a weighted measure can be used on such databases. On the basis of the above-discussed limitation, we have proposed the weight assignment as given in Table 1.

For a given publication (DOI exists in both Scopus and WoS), if a citation with DOI exists in both Scopus and WoS, then it would be considered as a common citation and would be assigned a weight 1. On the other hand, a unique citation, i.e. either in Scopus or WoS, the assigned weight would be 0.5 (in the case of N databases, the payoff will be 1/N). Table 2 demonstrates three examples with different scenarios.

- **Example 1**: Publication (P) has received 3 citations in Scopus and 0 in WoS. Hence, there is no common citation and we have left with 3 unique citations. So the proposed number of citations of P would be 2.
- **Example 2**: P has received 3 citations in Scopus and 3 in WoS. Let’s say there are 2 common citations and 2 unique citations (1 from each database). So the proposed number of citations of P would be 3.
- **Example 3**: When all citations are common citations. No payoff is assigned.

### Table 1
The payoff matrix demonstrates the weight for common and unique citations. Three possible cases are: (i) both databases have a common citation, (ii) either or both databases have unique citation, and (iii) no citation.

|       | Received citations | Not received citations |
|-------|--------------------|------------------------|
| Scopus | WoS                |                        |
| Received citations | (0.5, 0.5)        | (0.5, 0)               |
| Not received citations | (0, 0.5)         | (0, 0)                 |

### Table 2
Weight assignment to a publication P for N = 2 databases, Scopus (S) and WoS (W). Weight assigned to citations as common=1 and unique=0.5.

| Example | Citations | Weight = Common + Unique/N |
|---------|-----------|----------------------------|
|         | S         | W  | S+W | S | W  |
| 1       | 3         | 0  | 0   | 3 | 0  | 0 + 3/N = 1.5 \( \approx 2 \) |
| 2       | 3         | 3  | 2   | 1 | 1  | 3 + 2/N = \approx 3 |
| 3       | 3         | 3  | 3   | 0 | 0  | 3 + 0 = 3 |
3 Results

Here we have presented the comparative analysis of Scopus, WoS, and conflate at author’s, organization, and journal level.

3.1 At author’s level

Fig. 2 shows the comparison of conflate with Scopus and WoS on the basis of the number of articles, the number of citations, and $h$-index of 400 authors. Authors have been categorized into five disciplines (number of authors) based on their work domain: Social Sciences (66), Sciences (43), Humanities (20), Life Sciences (211), and Engineering (60). Scopus contains the large number of articles for Social Sciences, Sciences, Humanities, and Engineering whereas WoS shows for Life Sciences. The number of articles in conflate ranges in between Scopus and WoS count, except Life Sciences. A large number of citations are reported in Life Sciences in Scopus and the less number of citations are reported in Social Sciences in WoS. For Sciences, Engineering, and Life Sciences, conflate has reported the highest number of citations as compared to Scopus and WoS. For the rest of the disciplines, the number of citations reported by conflate is in between the range of Scopus and WoS. For the $h$-index of 400 authors, we found that conflate has reported the same $h$-index in Social Sciences and Sciences as reported by Scopus. For Humanities and Engineering, conflate has reported $h$-index in the range of Scopus and WoS. For Life Sciences, Scopus and WoS have reported the same $h$-index whereas conflate has reported one point higher of both.

Fig. 2: A comparative analysis between Scopus (colored in golden), WoS (colored in cyan), and conflate (colored in pink) for 400 authors as (a) the number of articles, (b) the number of citations, and (c) $h$-index. The analysis is performed for five disciplines: Humanities, Sciences, Social Sciences, Engineering, and Life Sciences. The standard deviation recorded for Scopus: articles (94.13), citations (4324.97) and $h$-index (14.82); for WoS: articles (105.43), citations (4263.24) and $h$-index (14.56); and for conflate: articles (94.06), citations (4599.34) and $h$-index (15.16).

Fig. 3 shows the comparative analysis of the number of articles, citations, and $h$-index between Scopus, WoS, and conflate. The best fit line (colored in red) shows less variation among the Scopus and conflate, and WoS and conflate. The overall slope is higher in WoS. In the comparative analysis of
A weighted unified informetrics based on Scopus and WoS

Fig. 3 Comparative analysis of 400 authors between Scopus and conflate (a-c) and WoS and conflate (d-f). The number of articles and citations in Scopus (33182, 1097446), WoS (31732, 1024808), and conflate (32376, 1130306). The maximum $h$-index in Scopus (91), WoS (95), and conflate (96). The red line represents the best fit line.

Scopus and conflate (see Fig. 3 (a-c)), it is observed that the average number of articles published by an author is 83, whereas in conflating it is 81. In Scopus, the average number of citations an author received is 2744 whereas in conflate it is 2826. The average $h$-index of an author in Scopus is 21 and 22 in conflate. Although the average number of articles calculated in Scopus is less than conflate; however, an average number of citations and average $h$-index are higher in conflate. Similarly, in the comparative analysis of WoS and conflate (see Fig. 3 (d-f)), the average number of articles published is 79 as compared to the average of 81 articles per author in conflate. The average number of citations published in WoS is 2562 as compared to 2826 in conflate. The average $h$-index per author in WoS is 20 whereas in conflating it is 22.

Table 3 represents the comparative analysis of Scopus, WoS, and conflate for 400 author’s articles, citations, and $h$-index among different disciplines.

| Categories | Authors - 400 | Articles | Citations | Average $h$-index |
|------------|--------------|----------|-----------|------------------|
| Life Sciences | 17798 | 18267 | 17971 | 22 |
| Social Sciences | 9521 | 3113 | 3457 | 15 |
| Engineering | 6741 | 5658 | 5940 | 23 |
| Sciences | 3397 | 3187 | 3340 | 24 |
| Humanities | 1720 | 1571 | 1688 | 22 |
3.2 At organization level

Here we analyzed the top 100 organizations in India and the categorization is done on the basis of their entity specification (count): Universities (69), IITs (16), NITs (8), and IISC & IISER (7). It is observed that the highest number of articles published among different databases are from IITs and the lowest number of articles are from NITs. Conflate reported that the number of articles published among different databases is varying between Scopus and WoS across all entities. In all entities, conflate reported the highest number of articles as compared to WoS and the lowest number of articles as compared to Scopus. Conflate reported the highest number of citations as compared to Scopus and WoS for all entities. In comparison with Scopus and WoS, Scopus has always reported more number of citations as compared to citations reported by WoS. h-index reported by conflate is also highest among both databases. IITs have received the highest h-index and NITs have received the lowest h-index among other entities. Conflate also reported that the results generated are always in between the range of Scopus and WoS. Among four entities, it can be observed that IITs have the highest h-index across multiple databases as shown in Fig. 4. Further, one can analyze the different disciplines of these organizations to keep track of the most popular discipline in terms of research publications.

Fig. 4 A comparative analysis between Scopus (colored in golden), WoS (colored in cyan), and conflate (colored in pink) for top 100 Indian institutes as (a) the number of articles, (b) the number of citations, and (c) h-index. The analysis is performed for four categories of institutes: NITs, Universities, IISC & IISER, and IITs. The standard deviation recorded for Scopus: articles (10459.32), citations (161121.19), and h-index (47.36); for WoS: articles (9637.33), citations (138670.18), and h-index (44.24); and for conflate: articles (9774.53), citations (197156.68), and h-index (51.75).

Fig. 5 shows the comparative analysis between Scopus, WoS, and conflate for top 100 Indian organizations. In the comparative analysis of Scopus and conflate (see Fig. 5(a-c)), it is observed that the average number of articles in Scopus is 9641 as compared to 8737 in conflate. The difference in the average number of articles states that all articles published in Scopus are not considered in conflate. The average number of citations recorded in Scopus is 113999 as compared to 134831 in conflate. The average h-index calculated in Scopus for these organizations is 91 which is quite lesser than the average h-index (100) calculated in conflate. Similarly, in the comparative analysis of WoS
Fig. 5 Comparative analysis of top 100 organizations between Scopus and conflate (a-c) and WoS and conflate (d-f). The number of articles and citations in Scopus (964093, 11399909), WoS (797158, 9337059), and conflate (873719, 13483112). The maximum h-index in Scopus (260), WoS (246), and conflate (289). The red line represents the best fit line.

and conflate (see Fig. 5 (d-f)), it is observed that the average number of articles in WoS is 7971 whereas in conflating it is 8737. Conflate also reported a significantly higher number of citations with an average score of 134831 as compared to 93371 in WoS. The average h-index in conflate is also 100 which is quite higher than the average h-index 82 reported by WoS. Table 4 represents the comparative analysis of Scopus, WoS, and conflate for articles, citations, and h-index for 100 organizations categorized into 4 main head organizations.

### Table 4 Comparative analysis of Scopus, WoS, and conflate for articles, citations and h-index for 100 organizations categorized in to 4 main head organizations.

| Organizations - 100 (Type Wise) | Articles | Citations | Average h-index |
|-------------------------------|----------|-----------|-----------------|
|                               | Scopus   | WoS       | Conflate        | Scopus   | WoS       | Conflate        |
| NITs                          | 30392    | 39059     | 47683           | 416302   | 318676    | 492760          |
| Universities                  | 50362    | 39059     | 47683           | 416302   | 318676    | 492760          |
| IISC & IISER                  | 77349    | 70063     | 73835           | 1252987  | 1107018   | 1542242         |
| IITs                          | 270495   | 236547    | 253042          | 3678723  | 2993534   | 4450159         |

3.3 Journal level

Here we analyzed 1000 journals and broadly divided into 5 disciplines (journal count), **Engineering (800), Social Sciences (119), Life Sciences (35), Sciences (27), and Humanities (19)**. The number of articles observed in **Sciences** is highest in Scopus with lowest in **Social Sciences**. For **Social Sciences**, conflate reported the highest number of articles among Scopus and WoS. For **Humanities, Engineering and Sciences**, conflate has reported a number of articles in
between Scopus and WoS. For *Life Sciences*, conflate has reported almost the same number of articles as compared to Scopus which is quite lesser than the WoS database. The number of citations reported by conflate is in between the range of Scopus and WoS for all disciplines where *sciences* is on top and *social sciences* is at the bottom. *h*-index reported by conflate for 1000 journals is the same as reported by Scopus for *Humanities, Sciences, and Life Sciences*. For *Social Sciences* and *Engineering*, it is in between the range of Scopus and WoS. Lowest *h*-index is reported by WoS for *Social Sciences* and highest by Scopus for *Life Sciences* as shown in Fig. 6. Fig. 7 shows the comparative

![Fig. 6](image1.png)

**Fig. 6** A comparative analysis between Scopus (colored in golden), WoS (colored in cyan), and conflate (colored in pink) of 1000 journals as (a) the number of articles, (b) the number of citations, and (c) *h*-index. The analysis is performed for five disciplines: Humanities, Sciences, Social Sciences, Engineering, and Life Sciences. The standard deviation recorded for Scopus: articles (1794.63), citations (59090.78), and *h*-index (44.40); for WoS: articles (1819.10), citations (45753.64), and *h*-index (40.06); and for conflate: articles (1677.79), citations (55723.79), and *h*-index (45.13).

![Fig. 7](image2.png)

**Fig. 7** Comparative analysis of 1000 journals between Scopus and conflate (a-c) and WoS and conflate (d-f). The number of articles and citations in Scopus (1528904, 30964292), WoS (1415093, 22570461), and conflate (1481823, 29276118). The maximum *h*-index in Scopus (408), WoS (344), and conflate (381). The red line represents the best fit line.
analysis between Scopus, WoS, and conflate for 1000 journals. In the comparative analysis of Scopus and conflate (see Fig. 7 (a-c)), it is observed that the average number of articles in Scopus is 1529 and in conflate is 1482. There is a slight hike in the average number of articles in Scopus. Similarly, the average number of citations in Scopus is 30964, and in conflate is 29276. The average h-index calculated in Scopus is 56 and in conflate is 53. Similarly, in the comparative analysis of WoS and conflate (see Fig. 7 (d-f)), it is observed that the average number of articles in WoS (1415) as compared to conflate (1482) shows significantly close values. The average number of citations in WoS is 22570 as compared to 29276 of conflate. Conflate clearly states that there is more scope of consideration of citations as compared to citations considered by WoS. The average h-index in WoS is 44 as compared to 53 in conflate. Conflate is clearly moving ahead in terms of the average number of citations and h-index calculation of WoS. Table 5 represents the comparative analysis of Scopus, WoS, and conflate for 1000 journal articles, citations, and h-index among different disciplines.

### Table 5 Comparative analysis of Scopus, WoS, and conflate for 1000 journals articles, citations and h-index for different disciplines.

| Journals - 1000 (Disciplines) | Articles | Scopus | WoS | Conflate | Scopus | WoS | Conflate | Average h-index |
|--------------------------------|----------|--------|-----|----------|--------|-----|----------|----------------|
| Life Sciences                  | 6061     | 6142   | 6089 | 114470   | 125220 | 113971 | 59         | 50             |
| Social Sciences                | 94755    | 86719  | 98316 | 1712570  | 1268836 | 1569405 | 42         | 30             |
| Engineering                    | 1298929  | 1179771| 1244492| 26468041 | 19170940| 24982597| 58         | 46             |
| Sciences                      | 56390    | 53458  | 54009 | 1061160  | 783108  | 1030698 | 56         | 49             |
| Humanities                    | 18769    | 27003  | 24957 | 577741   | 483357  | 562547 | 54         | 46             |

### 4 Technological Enhancement

Distributed ledger technology has found its applications in the field of education for verification of academic records (Aamir, Qureshi, Khan, and Huzaifa, 2020), sharing of student credentials (Mishra, Kalla, Singh, and Liyanage, 2020), adoption of smart learning environments (Ullah, Mugahed Al-Rahmi, Alzahrani, Alfarraj, and Aiblehaj, 2021) and in implementation of mobile-based higher education systems (Arndt and Guercio, 2020). This gives us an opportunity to work on the potential of this technology and enhance its features to its full capability, in the field of education. Making the use of distributed ledger technology in the research publication industry is a novel approach. There is a very good scope of observing challenges associated with new implementations (Upadhyay, 2020). Features like decentralization, persistency, anonymity, and auditability of records give more confidence to its stakeholders in a system presenting a scientific work of authors, organizations, and journals (McLean and Deane-Johns, 2016; Deshpande, Stewart, Lepetit, and Gunashekar, 2017). Hence, using DLT in the research publication industry can be considered as a viable choice to systematically achieve a sustained
system in the interest of its stakeholders. To achieve acceptability in DLT the consensus is used where every individual in a system must accept the happenings (Bamakan, Motavali, and Bondarti 2020). There are multiple studies on the performance and analysis of various consensus algorithms (Mingxiao, Xiaofeng, Zhe, Xiangwei, and Qijun 2017; Hao, Li, Dong, Fang, and Chen 2018; Yim, Yoo, Kwak, and Kim 2018; Bach, Mihaljevic, and Zagar 2018). To achieve consensus in unified informetrics, a concept of summation of common and unique citations may be used.

4.1 Mapping of publications and DLT

![Distributed Ledger Technology – Mapping with Research Publications](image)

**Fig. 8** Schematic representation of DLT mapping with research publications. An informetric ledger will contain the individual blocks for all the entities like authors, affiliations, and journals. A conflate ledger block will be created on the basis of articles referred to as records and citations referred to as a transaction for each individual entity. Block can be considered as the conflate of records and transactions whereas a ledger can be considered as the conflate of block, record, and transactions.

A distributed ledger data bank is believed and accessed autonomously by every contributor involved in a huge system (Maull, Godsiff, Mulligan, Brown, and Kewell 2017). The allotment is exclusive and records are held and autonomously constructed by each node (Collomb and Sok 2016). A block acts as a page of a register or record book. It is thus a stable collection of records which, once engraved, cannot be reformed or removed (Sunyaev 2020).
The blocks are supplemented to the chain in a direct-sequential order. Chain structure permanently time-stamp’s and stores exchange of value, preventing anyone from altering the ledger \cite{Naughton2016, Rauchs2018}. Each block’s record indicates a minimum of one transaction; however, many effective transactions can be characterized in a single block. Every transaction record (ledger entry) is connected to preceding transactions and is consistent for every contributing node \cite{Olnes2017}. Every ledger entry is retraceable through its complete antiquity and can be remodeled \cite{Liu2020}. Fig. 8 gives the insight details that how the terminologies in DLT may be mapped with academic research publications. To map DLT with academic research publications, we propose three ledgers as follows:

- **Author ledger:** In the author ledger, research contributions identified with DOI will act as a record, and citations received by such contributions will act as a Block. Transfer of citations in the author ledger will be considered as transactions. The uniqueness of the author ledger will be maintained by the ORCID ID of the authors.

- **Organization ledger:** In the organization ledger, research contributions identified with DOI will act as a record, and citations received by such contributions will act as a Block. Transfer of citations in the organization ledger will be considered as transactions. The uniqueness of the organization ledger will be maintained by the organization name.

- **Journal ledger:** In the journal ledger, research contributions identified with DOI will act as a record, and citations received by such contributions will act as a Block. Transfer of citations in the journal ledger will be considered as transactions. The uniqueness of the journal ledger will be maintained by journal ISSN.

### 4.2 Implementation process

We have implemented the distributed ledger in Python. It is the fully functional distributed ledger-based application \cite{Kansal2017}. It uses the concept of Python, HTML, and Javascript to build up the required framework where an individual can interact with the system through its web browser. The steps of implementation are:

1. Selection of profile as an author, organization, or journal.
2. Input ORCID or Organization name or ISSN.
3. Fetch the required publication and citation details from both bibliographic databases.
4. Do citation analysis.
5. Generate single publication and citation count of the selected profile.
6. Start DLT server on the available port, usually on 8000.
7. Start DLT based application on the available port, usually on 5000.
8. Browse the csv file generated for single publication and citation count of the selected profile.
9. Click on the Post button to post the request to the DLT server.
10. Click on the button “Request to mine” to mine the node in the ledger.
11. Click on the button “Resync” to resync with the blocks for updated data on the ledger.
12. Blocks will be displayed with the calculated $h$-index on the basis of single publication and citation count uploaded for the selected profile.

4.3 Implications of the proposed approach

In the current scenario, different bibliographic databases like Scopus, WoS, etc. have their own systems where an author and organization can find their scientific impact (Visser, van Eck, and Waltman 2021). This makes the process quite tedious, as one ends up with three different citation counts and three different $h$-index values for the same entities due to an individual calculation system of these bibliographic databases. The proposed system provides a single platform that fetches the required information of publications and citations from multiple bibliographic databases, performs analysis on the fetched data, and provides unified informetrics. The proposed system provides:

1. Cognitive and synthesized informetrics.
2. Expands the knowledge base of its stakeholders by comprising the information of multiple bibliographic databases.
3. Explores informetrics in a novel way and gives a clear assessment of research impact.
4. Supports the integration of $N$ bibliographic databases.
5. Presents in-depth analyses of the core components like publications, citations, self-citations, etc.
6. Supports better understanding of research impact of authors, organizations, and journals.
7. Facilitates its stakeholders for the establishment of a system providing a clear, authentic, and simulated environment for the research measurement of entities.

5 Summary

Scopus and WoS are considered as an important database which is being used worldwide. Both are traditional and the authentic source of accessing scientific work (Zhu and Liu 2020). Different universities, government organizations, recruiters, accreditation, and ranking agencies ask informetrics based on Scopus and WoS separately in their job applications and documentation. Authors are required to provide total publication count, citation count, $h$-index, etc. both from Scopus and WoS separately. This gives different informetrics for a single author. There is no common platform in our knowledge that can record or calculate single informetrics across multiple bibliographic databases. Therefore, a weighted unified informetrics system based on distributed ledger technology
named “conflate” has been discussed and proposed. The proposed solution provides a transparent and distributed view of the research contributors to their stakeholders. Calculated results also signify the efficiency of “conflate”. We have used Scopus and WoS for the implementation due to the availability of the data. The proposed implementation of the pay-off matrix also strengthens the overall framework of the proposed system.

The key findings of the work are: (i) It presents a unified method to maintain records associated with entities of author, organization, and journal. This method determines an absolute number of articles and citations for different entities. (ii) The mapping of multiple bibliographic databases for the calculation of $h$-index, and related informetric with the concept of pay-off matrix. (iii) The use of distributed ledger technology for the generation of ledger blocks for authors, organizations, and journals to safeguard their research contributions for any kind of manipulation and provides a robust platform for the presentation.

The presented work has some advantages as (i) The DOI-based data filtration helps us to identify the authenticity of received citations and publications (Gorraiz et al., 2016). (ii) Different stakeholders like government agencies, accreditation agencies, ranking organizations, and funding agencies can use the proposed system for the evaluation of the research contribution of individuals, organizations as well as journals. (iii) The introduction of distributed ledger technology in research publications will provide a verifiable research record of entities across different ledgers. This will introduce a tempering proof system for the integrity of records. (iv) The proposed system is a novel system introduced with the conflate of two traditional bibliographic databases like Scopus and WoS.

The major limitation of the study is the fact that we have considered the publications where DOI exists. In case WoS and Scopus do not have DOI numbers for the particular publications, we will not be able to consider the publication as authentic and the author will lose publications count and their citations count as well. Moreover, it could be a citation loss for low profile authors who have their work indexed only in Scopus or in WoS, refer to example 1 in Table. For such journals which are indexed only in Scopus or in WoS, also shows their limitations to other bibliographic databases. If an author publishes his work in a journal that is indexed in multiple bibliographic databases, there is a good chance of higher visibility of a scientific work to be read and cited worldwide. As new bibliographic databases may populate in the near future, the proposed system should support the integration of those databases into the existing system.

To conclude further, there are still several possible areas for further exploration and extension. Here are some interesting areas for possible future developments and research.

- **Different bibliographic databases:** We have studied the features of two bibliographic databases such as Scopus and WoS. Hence, the performed study is limited to two bibliographic databases. One can extend the study further
with the use of bibliographic databases like Microsoft Academics, Google Scholar (Martín-Martín, Thelwall, Orduna-Malea, and López-Cózar [2021]), OpenAIRE (Rettberg and Schmidt [2015]), DataCite (Brase [2009]), Mendeley (Reiswig [2010]) and Zenodo (Peters, Kraker, Lex, Gumpenberger, and Gorraiz [2017]). All these bibliographic databases may create conflated as per the model to calculate unified informetrics.

Different technological aspects: We have empowered citation analysis with distributed ledger technology. Hence, the performed study is limited to the implementation of only one technology in the research publishing industry. One can extend the study further with the use of “Gamification” and its gaming elements in the research publishing industry. The use of Gamification in the research publication industry can be helpful to increase the motivation and encouragement of its stakeholders for the extraction of unified informetrics for different entities (Kumar and Khurana [2012]).

Acknowledgements We would like to thank Dr. Ranbir Singh Batth and Mr. Sukhvir Singh for help and support. Both Scopus and WoS data has been downloaded from the Northwestern University, USA.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Aamir M, Qureshi R, Khan FA, Huzaifa M (2020) Blockchain based academic records verification in smart cities. Wireless Personal Communications pp 1–10

Abramo G, D’Angelo CA (2011) Evaluating research: from informed peer review to bibliometrics. Scientometrics 87(3):499–514

Adriaanse LS, Rensleigh C (2011) Comparing web of science, scopus and google scholar from an environmental sciences perspective. South African journal of libraries and information science 77(2):169–178

Adriaanse LS, Rensleigh C (2013) Web of science, scopus and google scholar: A content comprehensiveness comparison. The Electronic Library

Aghaei Chadegani A, Salehi H, Yunus M, Farhadi H, Fooladi M, Farhadi M, Ale Ebrahim N (2013) A comparison between two main academic literature collections: Web of science and scopus databases. Asian social science 9(5):18–26

Albort-Morant G, Leal-Rodríguez AL, Fernández-Rodríguez V, Ariza-Montes A (2018) Assessing the origins, evolution and prospects of the literature on dynamic capabilities: A bibliometric analysis. European Research on Management and Business Economics 24(1):42–52
Alvarez-Melgarejo M, Torres-Barreto ML (2018) Resources and capabilities from their very outset: a bibliometric comparison between Scopus and the web of science. Rev Eur Stud 10:1
Archambault É, Campbell D, Gingras Y, Lariviè re V (2009) Comparing bibliometric statistics obtained from the web of science and Scopus. Journal of the American society for information science and technology 60(7):1320–1326
Arndt T, Guercio A (2020) Blockchain-based transcripts for mobile higher-education. International Journal of Information and Education Technology 10(2):84
Bach LM, Mihaljevic B, Zagar M (2018) Comparative analysis of blockchain consensus algorithms. In: 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), IEEE, pp 1545–1550
Bacis E (2019) enricobacis/wos. URL https://github.com/enricobacis/wos
Bakkalbasi N, Bauer K, Glover J, Wang L (2006) Three options for citation tracking: Google scholar, Scopus and web of science. Biomedical digital libraries 3(1):1–8
Bamakan SMH, Motavali A, Bondarti AB (2020) A survey of blockchain consensus algorithms performance evaluation criteria. Expert Systems with Applications p 113385
Bar-Ilan J, Levene M, Lin A (2007) Some measures for comparing citation databases. Journal of Informetrics 1(1):26–34
Bartol T, Mackiewicz-Talarczyk M (2015) Bibliometric analysis of publishing trends in fiber crops in Google scholar, Scopus, and web of science. Journal of Natural Fibers 12(6):531–541
Bartol T, Budimir G, Dekleva-Smrekar D, Pusnik M, Juznic P (2014) Assessment of research fields in Scopus and web of science in the view of national research evaluation in Slovenia. Scientometrics 98(2):1491–1504
Brase J (2009) Datacite—A global registration agency for research data. In: 2009 fourth international conference on cooperation and promotion of information resources in science and technology, IEEE, pp 257–261
Burnham JF (2006) Scopus database: a review. Biomedical digital libraries 3(1):1–8
Chandrakar R (2006) Digital object identifier system: an overview. The Electronic Library
Collomb A, Sok K (2016) Blockchain/distributed ledger technology (DLT): What impact on the financial sector? Digiworld Economic Journal 3(103)
Daim TU, Rueda G, Martin H, Gerdts P (2006) Forecasting emerging technologies: Use of bibliometrics and patent analysis. Technological forecasting and social change 73(8):981–1012
De Groote SL, Raszewski R (2012) Coverage of Google scholar, Scopus, and web of science: A case study of the h-index in nursing. Nursing outlook 60(6):391–400
Deshpande A, Stewart K, Lepetit L, Gunashekar S (2017) Distributed ledger technologies/blockchain: Challenges, opportunities and the prospects for
standards. Overview report The British Standards Institution (BSI) 40:40
Franceschini F, Maisano D, Mastrogiacomo L (2016) Empirical analysis and
classification of database errors in scopus and web of science. Journal of
Informetrics 10(4):933–953
Gorraiz J, Melero-Fuentes D, Gumpenberger C, Valderrama-Zurián JC (2016)
Availability of digital object identifiers (dois) in web of science and scopus.
Journal of informetrics 10(1):98–109
Hao Y, Li Y, Dong X, Fang L, Chen P (2018) Performance analysis of con-
sensus algorithm in private blockchain. In: 2018 IEEE Intelligent Vehicles
Symposium (IV), IEEE, pp 280–285
Hoffman MR, Ibáñez LD, Simperl E (2019) Scholarly publishing on the
blockchain—from smart papers to smart informetrics. Data Science 2(1-2):291–310
Jacso P (2005) As we may search—comparison of major features of the web
of science, scopus, and google scholar citation-based and citation-enhanced
databases. Current science 89(9):1537–1547
Janowicz K, Regalia B, Hitzler P, Mai G, Delbecque S, Fröhlich M, Marti-
nett P, Lazarus T (2018) On the prospects of blockchain and distributed
ledger technologies for open science and academic publishing. Semantic web
9(5):545–555
Kansal S (2017) satwikkansal/python_blockchain_app. URL https://
github.com/satwikkansal/python_blockchain_app/tree/ibm_blockchain_post
Khurana P, Sharma K (2021) Impact of h-index on authors ranking: An
improvement to the h-index for lower-ranked author. arXiv preprint
arXiv:210508281
Kumar B, Khurana P (2012) Gamification in education-learn computer pro-
gramming with fun. International Journal of Computers and Distributed
Systems 2(1):46–53
Li J, Burnham JF, Lemley T, Britton RM (2010) Citation analysis: Compar-
ison of web of science®, scopus™, scifinder®, and google scholar. Journal
of electronic resources in medical libraries 7(3):196–217
Liu W, Huang M, Wang H (2021) Same journal but different numbers of pub-
lished records indexed in scopus and web of science core collection: causes,
consequences, and solutions. Scientometrics 126(5):4541–4550
Liu X, Farahani B, Firouzi F (2020) Distributed ledger technology. In: Inte-
ligent Internet of Things, Springer, pp 393–431
López-Illescas C, de Moya Anejón F, Moed HF (2009) Comparing bibliometric
country-by-country rankings derived from the web of science and scopus:
The effect of poorly cited journals in oncology. Journal of information science
35(2):244–256
Martín B (1996) The use of multiple indicators in the assessment of basic
research. Scientometrics 36(3):343–362
Martín-Martín A, Orduna-Malea E, Thelwall M, López-Cózar ED (2018)
Google scholar, web of science, and scopus: A systematic comparison of
citations in 252 subject categories. Journal of informetrics 12(4):1160–1177
Martín-Martín A, Orduna-Malea E, Thelwall M, Delgado-López-Cózar E (2019) Google scholar, web of science, and scopus: which is best for me? Impact of Social Sciences Blog

Martín-Martín A, Thelwall M, Orduna-Malea E, López-Cózar ED (2021) Google scholar, microsoft academic, scopus, dimensions, web of science, and open citations’ coc: a multidisciplinary comparison of coverage via citations. Scientometrics 126(1):871–906

Maull R, Godsiff P, Mulligan C, Brown A, Kewell B (2017) Distributed ledger technology: Applications and implications. Strategic Change 26(5):481–489

McLean S, Deane-Johns S (2016) Demystifying blockchain and distributed ledger technology—hype or hero. Computer Law Review International 17(4):97–102

Meho LI, Sugimoto CR (2009) Assessing the scholarly impact of information studies: A tale of two citation databases—scopus and web of science. Journal of the American Society for information science and technology 60(12):2499–2508

Meho LI, Yang K (2006) A new era in citation and bibliometric analyses: Web of science, scopus, and google scholar. arXiv preprint cs/0612132

MHRD (2020) National institutional ranking framework. URL https://www.nirfindia.org/nirfpdfcdn/2020/framework/Overall.pdf

Michael-E-Rose (2019) pybliometrics-dev/pybliometrics. URL https://github.com/pybliometrics-dev/pybliometrics

Mingxiao D, Xiaofeng M, Zhe Z, Xiangwei W, Qijun C (2017) A review on consensus algorithm of blockchain. In: 2017 IEEE international conference on systems, man, and cybernetics (SMC), IEEE, pp 2567–2572

Mishra RA, Kalla A, Singh NA, Liyanage M (2020) Implementation and analysis of blockchain based dapp for secure sharing of students’ credentials. In: 2020 IEEE 17th Annual Consumer Communications & Networking Conference (CCNC), IEEE, pp 1–2

Monash (2020) Monash. URL https://www.monash.edu/

Mongeon P, Paul-Hus A (2016) The journal coverage of web of science and scopus: a comparative analysis. Scientometrics 106(1):213–228

Naughton J (2016) Is blockchain the most important it invention of our age. The Guardian 24

Neuhaus C, Daniel HD (2008) Data sources for performing citation analysis: an overview. Journal of documentation

Norris M, Oppenheim C (2007) Comparing alternatives to the web of science for coverage of the social sciences’ literature. Journal of informetrics 1(2):161–169

Oliveira AS, de Barros MD, de Carvalho Pereira F, Gomes CFS, da Costa HG (2018) Prospective scenarios: A literature review on the scopus database. Futures 100:20–33

Olnes S, Ubacht J, Janssen M (2017) Blockchain in government: Benefits and implications of distributed ledger technology for information sharing

Papić A (2017) Informetrics: The development, conditions and perspectives. In: 2017 40th International Convention on Information and Communication
Technology, Electronics and Microelectronics (MIPRO), IEEE, pp 700–704
Peters I, Kraker P, Lex E, Gumpenberger C, Gorraiz JI (2017) Zenodo in the spotlight of traditional and new metrics. Frontiers in Research Metrics and Analytics 2:13
Powell KR, Peterson SR (2017) Coverage and quality: A comparison of web of science and scopus databases for reporting faculty nursing publication metrics. Nursing outlook 65(5):572–578
Pringle J (2008) Trends in the use of isi citation databases for evaluation. Learned Publishing 21(2):85–91
Rauchs M, Glidden A, Gordon B, Pieters GC, Recanatini M, Rostand F, Vagner K, Zhang BZ (2018) Distributed ledger technology systems: A conceptual framework. Available at SSRN 3230013
Reiswig J (2010) Mendeley. Journal of the Medical Library Association: JMLA 98(2):193
Rettberg N, Schmidt B (2015) Openaire. College & Research Libraries News 76(6):306–310
Rose ME, Kitchin JR (2019) pybliometrics: Scriptable bibliometrics using a python interface to scopus. SoftwareX 10:100263
Sharma K, Khurana P (2020) Growth and dynamics of econophysics: A bibliometric and network analysis. arXiv preprint arXiv:201101881
Simmonds AW (1999) The digital object identifier (doi). Publishing research quarterly 15(2):10–13
Sunyaev A (2020) Distributed ledger technology. In: Internet Computing, Springer, pp 265–299
THE (2020) About the rankings. URL https://www.timeshighereducation.com/world-university-rankings/about-the-times-higher-education-world-university-rankings
Thelwall M (2018) Dimensions: A competitor to scopus and the web of science? Journal of informetrics 12(2):430–435
Ullah N, Mugahed Al-Rahmi W, Alzahrani AI, Alfarraj O, Alblehai FM (2021) Blockchain technology adoption in smart learning environments. Sustainability 13(4):1801
Umate R, Patil M, Telrandhe S, Pathade A (2020) Bibliometric analysis of publications from web of science affiliated to health sciences university from 2017-2019. Dermatology 41:1756–1764
Upadhyay N (2020) Demystifying blockchain: A critical analysis of challenges, applications and opportunities. International Journal of Information Management 54:102120
Vafeaian A, Noroozi Chakoli A, Hassanzadeh M (2011) Comparative evaluation of content quality: Scopus and isi (web of science). Library and Information Science Research 1(2)
Van Raan A (2003) The use of bibliometric analysis in research performance assessment and monitoring of interdisciplinary scientific developments. TATuP-Zeitschrift für Technikfolgenabschätzung in Theorie und Praxis 12(1):20–29
Vaughan L, Shaw D (2008) A new look at evidence of scholarly citation in citation indexes and from web sources. Scientometrics 74(2):317–330
Visser M, van Eck NJ, Waltman L (2021) Large-scale comparison of bibliographic data sources: Scopus, web of science, dimensions, crossref, and microsoft academic. Quantitative Science Studies 2(1):20–41
Yelne S, Umate MR, Patil M (2021) Assessment of scientific productivity of health sciences university in central india using bibliometric analysis of published materials from scopus and web of science databases between 2017 to 2019. Annals of the Romanian Society for Cell Biology pp 7173–7189
Yim J, Yoo HK, Kwak Jy, Kim SM (2018) Blockchain and consensus algorithm. Electronics and telecommunications trends 33(1):45–56
Zhu J, Liu W (2020) A tale of two databases: The use of web of science and scopus in academic papers. Scientometrics pp 1–15