Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 07 Issue 01
2024

Research Article
https://doi.org/10.56979/701/2024

Identification of Scientific Researchers at the Early Stage of Field of Study Trends

Lubna Zafar[1], Nayyer Masood[2], Fazle Hadi[3], and Sheeraz Ahmed[4*]

[1]University of Poonch Rawalakot, 12350, Azad Kashmir, Pakistan.
[2]Capital University of Science and Technology, Islamabad, 44000, Pakistan.
[3]Higher Education Department, Khyber Pakhtunkhwa, 19110, Pakistan.
[4]Iqra National University, Peshawar, 25000, Pakistan.
[*]Corresponding Author: Sheeraz Ahmed. Email: sheeraz.ahmad@inu.edu.pk

Received: February 11, 2024 **Accepted:** May 21, 2024 **Published:** June 01, 2024

Abstract: Classifying researchers at the emerging phase of a Field of Study (FoS) trend is of crucial. This process will reveal the early influential authors and guage the popularity of a particular FoS trend. Researchers might not only be active in emerging FoS trends relevant to their fields, but they might also find it highly helpful to be kept informed about the progresses of important new research areas. Companies and institutional funding agencies are also required to be frequently informed on changes to the scientific landscape, so that they can make initial choices about their important funds. The scientific community has produced numerous studies on the detection and analysis of FoS trends. These studies focus on multiple issues like, (i) birth and establishment of an FoS trend, (ii) number of publications and researchers in an FoS trend, (iii) communities of researchers being formed around an FoS trend, (vii) grouping of different FoS trends, etc. This study aims to identify authors active during the early stages of an FoS trend in the field of Computer Science. It utilizes scientific articles published between 1950 and 2018 within the Computer Science domain, sourced from the Microsoft Academic Graph (MAG) dataset. We have proposed an approach to detect influential researchers who were involved at the emerging stage of an FoS trend known as trend setters and the authors who followed it afterwards known as trend followers. The influential authors (trend setters) achieved high citation count and significance in a particular FoS. In our proposed approach, firstly, we have calculated the debut year of an FoS. Then, we have computed the FoS publication count, its author count and FoS trend by using Filed of Study Multigraph (FoM) with degree centrality measure. Afterwards, we applied Rogers' innovation diffusion theory for the detection of trend setters and followers. Lastly, we have compared our list of researchers (trend setters) with two existing lists of well-known Computer Science researchers. The following are the lists; (i) top 10 influential authors identified by [1] (ii) An existing list of Computer Science researchers with an H-index of 40 or higher (available at www.cs.ucla.edu/~palsberg/h-number.html) is utilized. The experimental results demonstrate that our proposed method successfully identifies many of the influential researchers featured on this list. In some instances, exact matches were found in relation to the FoS, confirming their status as trendsetters.

Keywords: Computer Science; Field of Study(FoS); Trend Setters; Emerging FoS.

1. Introduction

The research environment changes and evolves continuously and as a result new research fields emerge while some other fade out. The new research fields that emerge generally form a research field of study (FoS) trend. A Field of Study (FoS) trend refers to a research area that is popular during a particular time period or can be defined as an FoS that is growing in importance and influence over time [2]. For example, it can be said that the Artificial Intelligence is an FoS that currently has trend in Computer Science research area. Trend in a research area grabs the attention of large number of researchers. Moreover, it generally has applicability in different domains. For example, p2p network and semantic search are two

FoS that emerged around 2002, however, semantic search is still a trendy FoS as we see many publications in this FoS, on the other side, p2p network is not that trendy and number of publications and applications are quite less relatively. One way to supplement this argument is through google N-gram viewer [3] that shows the frequency of use of different terms in published literature.

Staying informed with the popular FoS is one method to keep up with the scientific landscape. It's critical to be aware of prior, current, and emerging FoS trends not only for a new researcher but also for the established or experienced ones. For example, a researcher may wish to conduct study in an FoS that has not received much attention. It can also be useful to a businessman attempting to analyse the risks of starting a new venture. Identifying FoS trends is crucial for determining the potential success or impact of a researcher's chosen fields of interest. It is essential for researchers, academic publishers, journal editors, institutional funding organizations, and other key stakeholders to recognize emerging FoS patterns in the research landscape. Similarly, identifying researchers during the early stages of a FoS trend is important since it will show which significant individuals helped or launched the trend's popularity.

There are three main stages for an FoS to become a trend:(a) embryonic, (b) early and (c) recognized. In embryonic stage of an FoS, a concept or an idea did not emerge, yet. Although it is already taking shape, a FoS has not yet been distinctly named and acknowledged by the scholarly community. To examine the problems and perspectives related to the emergence of the new FoS, academics from diverse domains are publishing and beginning new collaborations. In early stage of an FoS, now it has been recently emerged and few researchers starting publications and will agree on certain concepts. Afterward, an FoS becomes mature and enters in its recognized stage and several researchers actively publishing their results [4]. For example, figure 1 shows the embryonic, early and recognized stage of an FoS. The embryonic stage of “Semantic Search” FoS is 2003, it was still a concept where numerous researchers from fields such as the World Wide Web, Information Retrieval, Semantic Web, and Search Engines were joining forces. After the 2003, the FoS emerges, getting its identity, and enters in the early stage, and a group of researchers started publishing in this FoS. After few years, the FoS reaches its recognized stage with an increasing number of publications per year.

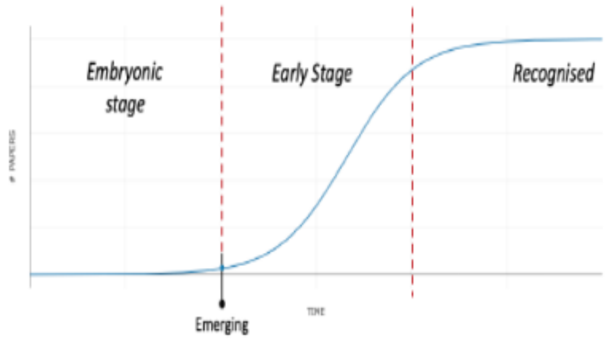

Figure 1. Semantic Search: FoS lifecycle representation [4]

By identifying FoS trends, we can pinpoint the researchers who were active at the early stages of an FoS trend, known as "trend setters," as well as the authors who later followed, referred to as "trend followers". Classifying authors into these two categories will help researchers to identify the influential authors in a specific FoS. Studying work of “trend-setters” of an FoS guides a researcher that how an FoS was originally conceived and proposed, the later review on that FoS will guide the stages it has gone through. For example, E.F Codd’s work [5] on Relational Data Model (2601 citations) or Tim Berners Lees work [6] on Semantic Web (2190 citations) gives real insight into these areas. That is why their work is still being cited heavily even today.

Classifying researchers at the early phase of an FoS is of importance as it will define who the noteworthy authors that started were or in growth the popularity of a particular FoS trend. For example, the Association of Computational Machinery (ACM) program distinguishes and regards researchers for their accomplishments in the Computer Science and Information Technology fields. The detection of researchers that are recognized as “trend setters” might assistance in defining the researchers to cogitate for such honors. According to state-of-art approaches, as soon as an innovative scientific area of research emerges, it drives over two key stages. In the preliminary phase, a group of researchers come to an agreement on few elementary concepts, construct a theoretical background and instigate to form a new scientific discipline. The research field then enters an established phase, during which a large number of researchers begin working on it, producing, and publishing results [4].

An approach highlighted by [1] also emphasized the initial phase, referred to as the embryonic phase, during which a Field of Study (FoS) lacks clear definition and labeling by the research community. However, this phase is now becoming evident, with researchers from diverse disciplines collaborating and conducting studies to elucidate the paradigms and issues associated with the early stages of new FoS. The emergence of new FoS in its early phases can offer significant benefits to all stakeholders in the research community. Academic editors and publishers can utilize this information to recommend the latest and most compelling content. Researchers can not only engage with emerging trends in new FoS relevant to their fields but also find it highly advantageous to stay abreast of the development of crucial new research areas. Similarly, companies and institutional funding agencies must stay updated on the evolving research landscape to make well-informed decisions about their funding priorities.

Our study aims to develop an approach for detecting "trend setters" and "trend followers" by identifying: (i) the debut year of a Field of Study (FoS), (ii) the FoS trend in papers, and (iii) the FoS trend among authors at an early stage through the construction of a multigraph using the degree centrality measure. Additionally, we focus on determining the researchers who published during the early stage of an FoS. This approach extends the work of [1], where influential authors of an FoS are identified in the embryonic stage, whereas our approach detects trend setters during the early stage of an FoS after its inception.

The rest of the paper is structured as follows: Section 2 presents a review of the literature on the emergence and evolution of Fields of Study (FoS), as well as techniques for detecting them. It also provides a detailed analysis of these techniques in scientific trends, emphasizing any issues and research gaps. Section 3 elaborates on the dataset and the proposed methodology. Section 4 covers the experimental setup, results, evaluation, and comparison with existing state-of-the-art approaches. Finally, Section 5 concludes the research by discussing limitations, outlining future work, and providing closing remarks.

2. Literature Review and Background Study

The importance of identifying and analyzing FoS evolution and its trends for a piece of work as well as for researchers is discussed in this section. In this section, we explore different methods for identifying the emergence of Fields of Study (FoS), techniques for tracking FoS evolution, approaches for detecting researchers at different FoS stages, and the challenges and issues underscoring the significance of this research. Additionally, we examine the limitations and research gaps present in these studies.

The static element of a FoS, or its identification within a group of documents, is the focus of FoS detection. Evolution of the FoS, on the other hand, is concerned with the dynamic nature of FoS, or how they change over time. [1] focused on two objectives in particular: First Story Tracking and Detection. The First Story Detection (FSD) task is used to detect the first story find previously unnoticed rising ideas. This task, in particular, keeps an eye on the incoming document flow to see if any new subjects have arrived. A competent FSD system, for example, should be able to recognise early Semantic Web papers from 2001, as well as Deep Learning and Cloud Computing papers from the mid-decade (2000-2009). This task, on the other hand, can only recognise people after they've already appeared, rather than anticipating or forecasting them.

Tracking, on the other hand, searches for fresh articles that address issues that have already been covered. When analysing incoming articles, the analysing system should be aware of the topics covered in the document collection and be able to categorise them appropriately. This implies it will extract the themes from each new document and arrange them with similar documents in the collection. The system may now do a statistical analysis to track the current state and evolution of each issue. Twenty years ago, some of the technology employed to aid in this endeavour was rather advanced. Cutting-edge algorithms for extracting topics from texts, such as PLSA [7], are now available, as well as a variety of similarity metrics for grouping articles based on their topics. These two tasks can be used by users to keep track of and analyze concerns as they arise. Although these are two distinct tasks, some solutions combine them to assess the existing state of each topic while simultaneously discovering new ones and organizing a large number of incoming articles.

[8] introduced a methodology for tracing the evolution of themes within a corpus over time, potentially uncovering novel motifs previously unseen in history. Various strategies in the literature aim to achieve this, such as investigating custom metrics based on the aggregate number of relevant papers [9-10-11]or the quantity of authors involved [12]. Other approaches utilize co-word analysis [13-14], hybrid methodologies [9], or citation analysis [10-11] to discern trends in document citations. Alternatively, some employ citation analysis to identify emerging themes. Lastly, a third approach constructs science maps through overlay mapping techniques, relying on expert analysis to identify new topics [15-16]

The burst detection approach for detecting emerging subjects recognises rapid changes in word usage, according to [17][18]is credited with inventing the burst detection approach for spotting emergent subjects, which detects rapid changes in word usage. This technique has stirred a good number of approaches for identifying research trends [19][10][20-23]. Citespace II, Sci2 and Network Workbench, and all include burst detection as part of their bigger tool sets.

Augur [1] is a revolutionary way for identifying study volunteers early on. Augur looks at the diachronic linkages across fields of study and can spot clusters of subjects with dynamics linked to the formation of new disciplines. A novel community discovery algorithm, the Advanced Clique Percolation Method (ACPM), was devised expressly for this objective, is also featured. From 2000 through 2011, Augur was compared to a gold standard of 1,408 new themes. Kleinberg's method involves analysing a stream of documents for bits that behave "bursty," that is, when they occur in a rapid burst of activity. This method uses probabilistic automation, with numerous phases dependent on how frequently each term is used. There are as many automata as there are words. When the frequency of the word they are connected with changes significantly, like at the beginning or conclusion of the burst period, the automata flip states.

Using a comprehensive knowledge tree that classifies research publications based on the topics covered in the Computer Science Ontology, a study [24] developed a framework for identifying, analyzing, and forecasting research topics. The authors first demonstrated how to use a domain ontology to annotate a scientific knowledge graph comprising research papers and their metadata with a collection of research themes. Then, based on this knowledge structure, numerous strategies for analysing research from various perspectives were discussed. Finally, based on this paradigm, two methodologies for anticipating research trends were described.

A scientific knowledge graph is a structured graph that was established and developed by the authors [25] with the purpose of transforming scientific knowledge. The authors used a variety of approaches used in literature like machine learning and Natural Language Processing NLP, as well as a workflow to combine their findings. Furthermore, knowledge from multiple scientific publications was combined into a single knowledge graph with the goal of representing detailed knowledge of the scientific literature in the Semantic Web domain.

This technique can be used to discover FoS and concepts that have gained traction and have sparked heated debate for some time. It must be integrated into a pipeline that first pre-processes the texts because it does burst analysis on every word, including stop words. Jo et al. [10] devised a method that incorporated phrase distributions including n-grams and the citation graph distribution for publications containing the term in question. If a term is related to a topic, for example, the authors expect that documents that contain that term will have a stronger relationship than documents chosen at random. The approach is successful, according to their findings, and can even detect new developing subjects. However, because the citation network of a phrase takes time to build to become firmly connected, their approach has a temporal lag.

By looking at how citation patterns have changed throughout time, [26-31] can discover the birth of a new region. These methods are based on the premise that bringing two previously unconnected or poorly connected locations together will result in a better outcome could signal the formation of a new subject that can build on earlier Takes. Because the authors focused on a small area of optics, manual analysis was possible; nevertheless, we could argue that if the domain was increased, such a strategy would not be scalable. [26] use co-citation networks in their research to group bibliographically related texts or to share a list of publications that have been cited. This strategy, however, has the same faults as Takeda and Kajikawa [30]. Furthermore, no statistical metric for analysing the introduction of a new topic is provided because these approaches are evaluated by a human expert.

Shibata et al. [11] on the other hand, used topological methodologies to identify the rise of new topics without the help of expert specialists. The citation network was separated into clusters by the authors and assigned the most representative word to each cluster. By looking at the age of the cluster, the method detects impending topics. This strategy, according to the authors, has a time lag, which we agree with. This is a problem that any citation and co-citation network-based strategy encounters. They are underrepresented in such networks because new articles can take up to two years to be mentioned.

Shibata et al. urge that these algorithms be supplemented with data from other sources, such as venues, to detect the introduction of new subjects [11]. Clarivate Analytics employs a different approach, relying on citations rather than topological network analysis. Clarivate Analytics has published a report called Research Fronts since 2013, which highlights a variety of important research fronts, including emerging and hot ones. According to "Research Fronts 2017", "A research front is made up of a core of highly cited articles that are linked to the citing journals that often co-reference the core," this report lists 100 hot research fronts as well as 43 new ones. They grouped the total number of research fronts (9,690) into ten macro-areas todiscover the most promising research fronts. The top ten research fronts for each of these ten organisations are then chosen based on their highly mentioned publications' average year i.e., core publications. Following that, the identified core articles, associated nations, and institutions are shown. Instead, they seek for research fronts that are increasing in fields where in the last two years, notable publications have been published (2015 to onwards). Human experts next analyse and interpret the evolving research fronts in order to catch recent trends and estimate their importance. This strategy has two significant flaws. This method, like others according to citation analysis, there is a time lag between the emergence of a research topic (emergent research fronts) and its identification. We can detect the emergence of a new issue two years later in the worst-case scenario, even with a two-year time lag. The second issue is the method's potential for low recall. Despite the lack of statistical data, in this report precision and recall were used to identify hot and developing research fronts, the method can be used to detect a problem like this. Because their primary papers garnered inadequate citations in the past two years, many fascinating subjects may go undiscovered.

To find out how the number of writers affects the emergence of new ideas, researchers have looked into co-authorship networks and the number of authors. A model that integrates three distinct emergence signals was proposed by Guo et al [12]. They came to their conclusions based on the frequency of keywords, the expanding number of authors, and the interdisciplinarity of the sources mentioned. The Rao-Stirling diversity index, which is calculated on a year-to-year basis [15] is used to calculate the final indicator. Bursts of keywords appear before the introduction of new themes, followed by rapid increases in the number of authors cited, as well as the interdisciplinarity of the references cited, according to the researchers.

Bettencourt et al. looked analysed co-authorship networks to determine if there were any trends that could be linked to the formation of new research fields [9]. Three main patterns were discovered: (i) the average number of nodes grows, showing that the network that surrounds such nodes is growing denser; (ii) the average path length ins two nodes stays the same or shrinks, suggesting that the network's width is changing; and (iii) the largest component has a growing number of edges. These all developments point to a tightening of the co-authorship network. As a result, forming a new research group is seen as a precursor to the development of new research topics. The authors developed a method for determining the genesis of themes by studying the expansion of conference networks. They started by creating a progressive conference network using co-word analysis, with nodes representing links and conferences signifying connections indicating proximity based on keywords extracted from published papers. They then look for conferences that are becoming more and more similar to one another, collapsing over one another as a sign of new topics emerging. [8] devised a mechanism for monitoring subjects' progress over time [8]. The approach takes two successive slices of the corpus after splitting it into discrete time windows, using LDA, extract the topics, and then examines how these subjects changed over time.

The primary concept is that by comparing topics created over a short period of time, one can discern how subjects develop and how their birth and death are captured. A strategy for predicting the emergence of a new subject was devised by Morinaga and Yamanishi utilizing the probabilistic Finite Mixture Model [14]. The authors dynamically learned the structure of the subjects from the papers in each year using this technique. The changes in the collected components were then examined by researchers to determine if any new subjects had developed. On the other hand, nobody has ever put their studies to the test. It is challenging to predict the creation of new subject fields using scholarly articles.

One of the first overlay mapping methodologies was developed by [32], who mapped the "backbone of science." They started by classifying the data into temporal frames, then looking for phrase clusters and linking them to research areas for each window, such as year. By monitoring the clusters for two years in a row, they were able to match similar themes across time and discover new clusters connected to new topics. Similarly, Leydesdorff et al. developed overlay maps to assist policymakers in locating research bodies that cross traditional academic boundaries [15]. These overlay mapping technologies are fascinating because they enable users to visually analyse locations in a global research environment where the number of publications is rapidly increasing. They can only provide a coarse-grained perspective since they neglect intricate linkages between research subjects. According to Rafols et al. ,they should be used in conjunction with other maps that provide more detailed viewpoints [33].

As we have observed that in previous state-of-art approaches researchers proposed techniques for the detection, evolution and the development of FoS over time. Their focus is on embryonic, well established and recognized FoS, where a few and an active number of authors are involved with a number of publications.

The detailed analysis shows that the scientific community has presented various studies on;

a. birth and establishment of an FoS trend
b. number of publications and researchers in an FoS trend
c. communities of researchers being formed around an FoS trend
d. lifespan of an FoS trend
e. grouping of different FoS trends etc

On the other hand, a thorough analysis on the state-of-the-art techniques is still lacking:

a. significance of following an FoS trend in Computer Science field
b. researchers who are involved at the emerging stage of an FoS trend

The gap leads to the following research question:

RQ1: How can we differentiate between trend setters and trend followers?

In the section that follows, we go into greater depth on how to fill in these research gaps by outlining the proposed approach.

3. Research Methodology

Classifying researchers at the early phase of an FoS is of importance as it will define who the significant authors that started were or in growth the popularity of a particular FoS trend. We have proposed an approach to detect influential researchers who were involved at the early stage of an FoS trend known as trend setters and the authors who followed it afterwards known as trend followers. The influential authors (trend setters) achieved high citation count and significance in a particular FoS. In our proposed approach, firstly, we have considered the debut year of an FoS as per approach of [1]. We selected the “Semantic Search” FoS because it has been discussed in our reference paper. From the debut year, our approach determines the trend setters through following steps:

1. We selected all the authors who published in the birth year of selected FoS. In this work, we have selected “Semantic Search” with birth year 2003 as it has been discussed in our base paper [1].
2. Then, we computed the publication count of these authors for next five years in Semantic Search, their citation count for the papers on Semantic Search and the degree centrality of FoM [34] for their papers on Semantic Search for next five years. We sorted three lists in descending order.
3. Afterwards, we applied Rogers Information of Diffusion Theories (IDT) [35] on three lists generated above. As per the Rogers IDT, top 2.5% authors are taken as trend setters and rest as different types of trend followers.
4. Lastly, we have compared our lists of researchers (trend setters) with two existing lists that contain highly recognized Computer Science scientists. The lists are as follows; (i) top 10 influential authors identified by [1] and (ii) an existing list of Computer Science scientists with H-index of 40 or higher (www.cs.ucla.edu/ palsberg/h-number.html). Figure 2 describes the modules of our proposed approach.

Figure 2. Proposed Approach

3.1. Dataset Collection

The dataset used in this study was obtained from Microsoft Academic [37] and is referred to as the Microsoft Academic Graph (MAG) dataset. It provides details about various academic publications, study fields, and the relationships between academic articles. The academic articles include books, journal articles, and conference papers. As can be seen in table 1 below, the information about these publications includes id, title, abstract, authors.name, venue, year, keywords, FoS, n citation, references, doc type, publisher and doi.

Various academic fields, including Physics, Computer Science, Engineering, Chemistry, and many others, are represented in MAG's academic papers. Without examining the document's text or abstract, the FoS of each study allows for the separation of statistics regarding overall data and data particular to Computer Science. Table 2 shows the MAG dataset statistics about multidiscipline and Computer Science.

The research domain of a specific scientific article is identified by Fields of Study (FoS) in the Microsoft Academic Graph (MAG). For example, a study that compares various machine learning techniques like Support Vector Machine (SVM), Naive Bayes, etc., would be categorized under FoS such as "Machine Learning" or "Artificial Intelligence" [2].

Table 1. MAG articles schema

Field Name	Description	Example
Id	**MAG ID**	00000707-26a7-491e-85b2-31063816253a
Title	**paper title**	The Research and Application of Resource Dissemination Based on Credibility and UCON
authors.name	**author name**	Fengying Wang, Fei Wang
Venue	**paper venue**	Computational intelligence and security
Year	**Published year**	2007
Keywords	**keywords**	['mirrors', 'technological innovation', 'certificate authority', 'image databases', 'computational intelligence', 'trust management', 'contracts', 'fuzzy set theory', 'usage control', 'access control authorization fuzzy set theory image databases mirrors contracts

		computational intelligence security fuzzy systems technological innovation', 'access control models', 'membership function', 'authorization', 'access control', 'security', 'fuzzy systems', 'digital right management']
FOS	**fields study of**	[Membership function', 'Computer Science', 'Knowledge management', 'Artificial intelligence', 'Information security', 'Access control', 'Computational intelligence', 'Data mining', 'Authorization', 'Fuzzy set', 'Computer security', 'Certificate authority"]
n_citation	**number of citation**	50

Table 2. MAG dataset count of multidiscipline and Computer Science entities

Entity	Total count	Computer secience count
Papers	228,956,810	1,354,603
Auhtors	21,969,837	2,24,591
Conferences	4,414	1,277
Fields of study (FoS)	50,007	9,800

As depicted in figure 3, each paper within MAG is assigned a unique identifier and is connected to one or more relevant FoS at different levels of the MAG hierarchy, ranging from level-0 to level-3.

A portion of the MAG hierarchy, from level 0 to level 3, is depicted in the figure above. FoS at a more general level, such as engineering, computer science, etc., are included in Level-0. The figure 4 illustrates that the lower levels have more focused FoS.

Figure illustrates an example of mapping by assigning a paper from the field of Computer Science to various levels of FoS, from level 3 to level 0. An FoS may have more than one parent FoS because the structure of the FoS in MAG is often that of a directed acyclic graph.

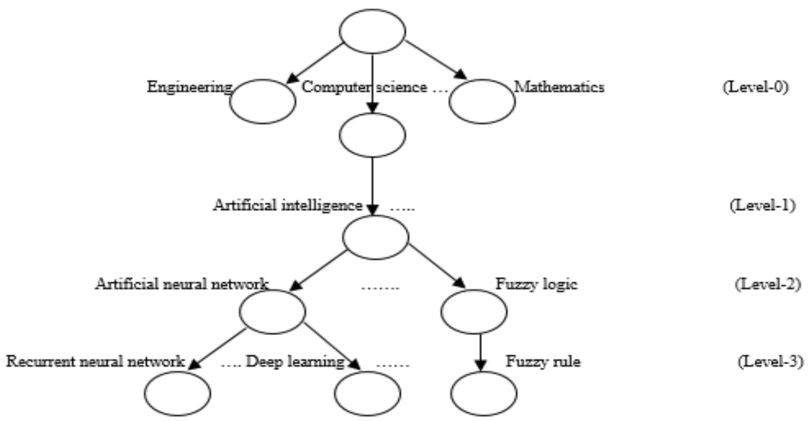

Figure 3. MAG different levels

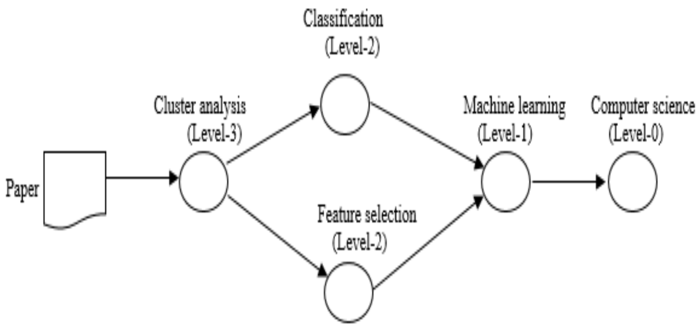

Figure 4. An example of Computer Science FoS levels

For instance, classification (level-2) and feature selection (level-3) are subsets of machine learning (level-1) and computer science (level-0), while cluster analysis (level-3) is a subset of both.

3.2. Preprocessing

As was previously mentioned, the MAG dataset includes publications from several fields. We choose the computer science research papers that were published between 1950 and 2018 for our investigation. Despite the fact that the MAG includes the articles that have been presented at conferences and in journals, however, because important findings are typically published initially in conferences, we have solely taken conference articles into consideration [2]. In the pertinent section, the preprocessing of data sets related to the study issue is covered in detail.

3.3. FoS Debut

In [1], authors identified Fields of Study (FoS) within Computer Science that emerged during the period from 2000 to 2009, as presented in table 3. The most straightforward method for detecting the inception of an FoS is to determine the year when the FoS label was first used as a keyword in a paper. For instance, the term "cloud computing" made its initial appearance in 2006. However, relying solely on the year when the label first surfaced can be deceptive. An FoS label might initially appear in a few articles with limited significance and then gain popularity in subsequent years with a completely different interpretation.

Table 3. Selected debutant FoS and year [1]

FOS	Year of Debut
Service Discovery	2000
Ontology Engineering	2000
Ontology alignment	2005
Service-oriented architecture	2003
Smart power grids	2005
Sentiment analysis	2005
Semantic search	2003
Linked data	2004
Semantic web technology	2001
Vehicular ad-hoc networks	2004
Mobile ad-hoc networks	2001
p2p network	2002
Location based services	2001
Service oriented computing	2003
Ambient intelligence	2002
Social tagging	2006
Community detection	2006
Cloud computing	2006
User-generated content	2006
Information retrieval technology	2008
Web 2.0	2006
Ambient assisted living	2006
Internet of things	2009

This phenomenon is exemplified by "linked data," initially employed in the database context to denote interconnected pieces of data, before being embraced by the semantic web as a distinct method for disseminating data using the Resource Description Framework (RDF) format [1]. This label misuse can create significant noise. To handle this issue, authors choose as debut year of an FoS the first year in which it reaches at least 5 publications. At the same time, they named the previous five years of debut year as embryonic duration and from this duration they identified influential authors; those whose work ultimately gave birth to this new FoS in the debut year. In this way, they can be more certain that a new label is already recognized by multiple researchers.

Table 4. Semantic search publication count and author count from 2003-2007

FoS	Year	Publication Count	Author Count
Semantic search	2003	232	545
Semantic search	2004	313	756
Semantic search	2005	421	959
Semantic search	2006	482	1151
Semantic search	2007	609	1413

3.4. Different Counts for Selected FoS

As mentioned above, we are working on the Semantic Search (SS) FoS, following our base paper. We selected all those authors who worked in SS in the debut year, that is 2003. Table 4 below shows the number of authors and number of publications in SS in five years starting from birth year. As shown in the table above, there were 545 authors who worked in SS in 2003 (the birth year) and our proposal is that the trend setters for the SS are among these 545 authors.

We selected all papers of these authors involving SS FoS for five years (2003-2007) and also the citation counts of those papers. Table 5 shows some of the authors for SS with their respective paper count and citation count. After having these two lists, we computed the third list and that is the degree centrality measure of FoM constructed for each author against his work on SS for the years 2003-2007. For this purpose, we collected the papers of individual authors working in SS FoS during 2003-2007. For each author, we prepared co-occurrence data for the SS FoS.

Table 5. Semantic search authors publication count and citation count from 2003-2007

Sr. No	Researcher	Publication Count	Citation Count
1.	Dieter Fensel	97	375
2.	Dan Suciu	62	323
3.	Justin Zobel	44	164
4.	James Allan	42	111
5.	W. Bruce Croft	29	109
6.	Dragomir R. Radev	28	67
7.	Alon Halevy	27	74
8.	Katia Sycara	26	60
9.	James Hendler	25	54
10.	Clement Yu	24	76
11.	Wolfgang Nejdl	23	48
12.	Victor Vianu	22	62
13.	Amit Sheth	20	45
14.	Andre Esteva	20	35
15.	Tom Gillespie	19	37
16.	Richard Christie	18	30
17.	Wenpeng Yin	17	32
18.	William W. Cohen	15	32
19.	Yuanzhang Li	15	34
20.	Berthier Ribeiro-Neto	13	22

Table 6. Semantic search co-occurrences with other FoS and its degree from 2003-2007.

FoS	Semantic Search co-occurrence with other FoS	FoS Trend-Degree				
		2003	2004	2005	2006	2007
Semantic Search	Content-based retrieval	4	-	-	-	-

Semantic Search	Computational semantics	3	2	4	4	3
Semantic Search	Semantic equivalence	3	2	-	3	3
Semantic Search	Social semantic web	3	5	3	3	4
Semantic Search	Semantic computing	-	3	-	5	5
Semantic Search	Digital libraries	-	4	4	4	3
Semantic Search	Intelligent agents	-	-	6	-	5
Semantic Search	Explicit semantic analysis	-	-	3	3	4
Semantic Search	Similarity heuristic	-	-	4	5	4
Semantic Search	Support vector machine	-	-	-	3	6
Semantic Search	Information retrieval systems	-	-	-	4	5

Table 6 shows the SS co-occurrences and its degree with other FoS during a specific time period. This data has been compiled for one author publications during 2003-2007. As can be seen from the table that SS appeared with content-based retrieval, computational semantics, semantic equivalence and social semantic web in 2003. Likewise, with computational semantics, semantic equivalence, social semantic web, semantic computing and with digital libraries in 2004 and other years.

From this data we prepared FoS multigraph (FoM) for a particular author as discussed in [34]. From this FoM we computed the degree centrality of the SS FoS. In this way, we prepared the degree centrality for all authors working in SS FoS.

Table 7. Authors SS degree from 2003-2007

Sr. No.	Researcher	FoS Degree	Sr. No.	Researcher	FoS Degree
1	Dieter Fensel	323	11.	Wolfgang Nejdl	63
2	Dan Suciu	276	12.	Victor Vianu	54
3	Justin Zobel	113	13.	Amit Sheth	61
4	James Allan	101	14.	Andre Esteva	55
5	W. Bruce Croft	82	15.	Tom Gillespie	51
6	Dragomir R.Rade v	50	16.	Richard Christie	49
7	Alon Halevy	97	17.	Wenpeng Yin	39
8	Katia Sycara	71	18.	William W. Cohen	35
9	James Hendler	87	19.	Yuanzhang Li	31
10	Clement Yu	88	20.	Berthier Ribeiro-Neto	29

Table 7 above shows the values of degree centrality of authors working in SS FoS. Table 7 contains degree centrality of the authors working in the SS FoS in year 2003 (debut year) and we compiled this table for their work between year 2003-2007. So far, we have prepared three lists containing publication count, citation count and degree centrality of 545 authors working in SS FoS in year 2003 (table 5 & 7). We are going to use these lists to find out the trend setters for the field of Semantic Search as explained in the next section.

3.5. Emerging FoS and Rate of Adoption

After the detection of FoS debut year, its publication count, author count and FoS trend (FoS trend-degree). Now, it is possible to identify the researchers involved at the early stage of an FoS trend and who followed FoS trend afterwards. An FoS in its debut year seems appears only in the papers of this time period and not in papers (back years), it can be now the early stage of an FoS. The authors involved at this stage of FoS are the trend setters or innovators and others are trend followers. We use Rogers IDT[35]to detect the trend setters and followers from the early stage of an FoS trend.

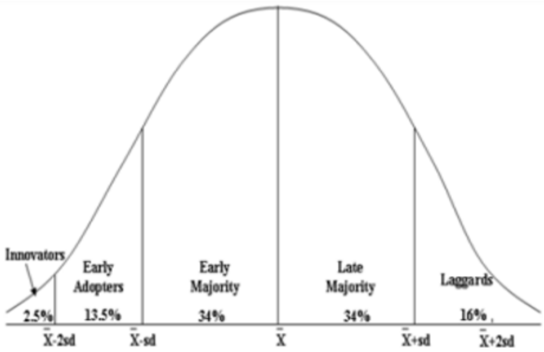

Figure 5..Rogers Innovation Diffusion Theory adopter categories [35]

The above figure 5 shows the trend setters and followers categories as depicted in [35]. We applied Rogers IDT on the three lists that we prepared for 545 authors who worked on SS FoS in the debut year, that is, 2003. We have presented the trend setters, the five adopter classes and the estimated fraction of authors encompassed to each are positioned on the adopter dispersal. The part to the leftward of the mean time x of adoption minus two standard deviations 2sd comprises the initial 2.5 percentage of the researchers intricate in the emergence of a trend the innovators or trend setters. The subsequent 13.5 percentage of researchers who adopt/accept the new trend are encompassed in the part among the mean minus one standard deviation sd and the mean minus two standard deviations; they are labeled early adopters.

The succeeding 34 percentage of the researchers are adopters, known as early majority, are comprised in the part among the mean time of adoption and minus 1sd. Amongst the mean and 1sd to the right of the mean are positioned the subsequent 34 percentage of authors to accept/adopt the trend, the late majority. The preceding 16 percentage of authors are known as laggards [35].

3.5.1. Trend Setters/Innovators

According to Rogers [35], innovativeness refers to the propensity of an individual or a group to embrace new concepts ahead of their peers within an organization. Innovativeness plays a key role in comprehending the significant behaviors within the innovation-decision process and guides the classification of adopters based on their level of innovativeness. Innovators or trend-setters demonstrate a willingness to explore novel ideas and concepts, acting as gatekeepers who introduce new concepts from external sources into a system. They are adept at navigating through high levels of ambiguity regarding new innovations compared to followers. For instance, they are typically the first to adopt a new idea or concept within their environment and are less influenced by the opinions of other followers within the organization. Innovators embrace novel ideas even when there is little familiarity with the innovation within the organization. They represent the initial 2.5% of individuals within an organization who adopt an innovation.

3.5.2. Trend Follower Categories

Trend followers validate a new idea by embracing it, believing it has reached a point of safety for adoption. They are individuals within a system who prefer to wait until the majority of their peers have embraced the innovation. Conversely, there are individuals in a system who are typically the last to adopt an innovation, often those who stand to benefit the most from it. Due to limited resources and a lack of awareness or knowledge about innovations, they prioritize ensuring the effectiveness of an innovation before adoption. Consequently, they tend to base their decision on whether the innovation has been successfully adopted by other members of the system in the past. Due to all these characteristics, some followers innovation-decision period is relatively long [35]. Here, we are considering early adopters, early majority, late majority and laggards in trend follower categories. We have detected trend setters who are involved at the early stage of an FoS (only in debut year) as 2.5% of authors and trend followers who followed the FoS trend after debut year by applying Rogers [35].

Table 8. Semantic search trend setters and followers in 2003

FoS	Debut Year	#papers	#authors	Trend Setters	Trend Followers
Semantic Search	2003	232	545	13.625	531.375

As table 8 above shows trend setters and followers distributions. After the detection of researchers as trend setters and followers, we have calculated the trend setters publication count, FoS publication count and author FoS trend by using FoM with degree centrality measure.

4. Results and Discussions

This section presents the results of our research question, as we have detected the individuals involved at the early stage of an FoS in the above section. This is challenging to evaluate trend setters at the early stages of an FoS. We have compared our list of researchers (trend setters) with two existing lists that contain highly recognized Computer Science scientists. The lists are as follows; (i) top 10 influential authors identified by [1] and (ii)an existing list of Computer Science scientists with Hindex of 40 or higher(www.cs.ucla.edu/ palsberg/h-number.html). The H-index is defined as a measure to compute the scientific output of a researcher, where h is the number of publications with citation count higher or equal to h [36].

Table 9 shows a comparison of the "Semantic Search" FoS trend setters at the early stage identified by our approach with top 10 influential authors identified by [1].The table shows that Dan Suciu, Justin Zobel, Dieter Fensel, W. Bruce Croft, Clement Yu, Dragomir R. Radev, James Allan and Victor Vianu have the exact match with influential authors [1]. These authors worked and published at the embryonic and early stage of "Semantic Search" FoS.

Table 9 above shows comparison of influential authors in Semantic Search FoS identified in [1] and those established by our approach. The table highlights following aspects:

- Seven out of top ten influential authors are common in both approaches, however, there is difference in the rankings of such authors as highlighted in the table .
- Three of the authors that are not in the top ten lie within top twenty trend setters as proposed by our approach.
- The strength of approach by [1] is that they identify the influential authors from the five years prior to the birth year of an FoS, whereas we identify the trend setters from the work done in next five years of the birth year of FoS. In spite of this difference, majority of authors are common in both approaches.
- In order to evaluate that which of the two approaches identifies better trend setters, we evaluated the major authors working in the Semantic Search FoS from 2003- 2007.

Table 9. Top-left, we show trend setters and on top-right, the top 10 influential authors of the semantic search FoS

Influential Authors	Ranking by [1]	Ranking by Proposed Approach
W.Bruce Croft	1	5
Dieter Fensel	2	1
Dan Suciu	3	2
William W.Cohen	4	18
Berthier Ribeiro-Neto	5	20
Clement T.Yu	6	10
James Zobel	8	3

Dragomir R.Radev	9	6
VictorVianu	10	12
Alon Halvey	-	7
Katia Sycara	-	8
James Hendler	-	9

Table 10. Researchers appears in various lists, their publication count, citation count, FoS degree in semantic search FoS from 2003-2007

Rank	Researcher	Publication count	Citation Count	FoS Degree	Influential Author	H-Index	Trend Setter
1.	Dieter Fensel	97	375	323	yes	yes	yes
2.	Dan Suciu	62	323	276	yes	yes	yes
3.	Justin Zobel	44	164	113	yes	yes	yes
4.	James Allan	42	111	101	yes	yes	yes
5.	W. Bruce Croft	29	109	82	yes	yes	yes
6.	Dragomir R. Radev	28	67	50	yes	yes	yes
7.	Alon Halevy	27	74	97	no	yes	yes
8.	Katia Sycara	26	60	71	no	yes	yes
9.	James Hendler	25	54	87	no	yes	yes
10.	Clement Yu	24	76	88	yes	yes	yes
11.	Wolfgang Nejdl	23	48	63	no	yes	yes
12.	Victor Vianu	22	62	54	yes	yes	yes
13.	Amit Sheth	20	45	61	no	yes	yes
14.	Andre Esteva	20	35	55	no	yes	yes
15.	Tom Gillespie	19	37	51	no	yes	yes
16.	Richard Christie	18	30	49	no	yes	yes
17.	Wenpeng Yin	17	32	39	no	yes	yes
18.	William W. Cohen	15	32	35	yes	yes	yes
19	Yuanzhang Li	15	34	31	no	yes	yes
20.	Berthier Ribeiro-Neto	13	22	29	yes	yes	yes

Table 10 shows researchers identified by our approach in the early stage of an FoS trend, that is, trend setters for the FoS of Semantic Search. As shown in the table, the authors selected by our approach have more work in the concerned FoS, whereas, the authors at serial 7, 8 and 9, had relatively less work in the later years as compared to other authors which have been identified by our approach.

Moreover, all of the top twenty authors identified by our approach are in the list of authors having high h-index [36]. So in the nutshell, we can say that the approach of [1] identifies influential authors before the birth of an FoS, however, our approach identifies trend setters in the early years after the birth. Most of the authors are common in both lists, however, those identified by our approach proved more influential in the future. This is the edge that we have over our base approach.

5. Conclusion and Recommendations

This study holds significance for various stakeholders within the research environment, including researchers, subject experts, policymakers, academic publishers, journal editors, institutional funding bodies, and other relevant parties. It offers insights into past, present, and emerging trends in Fields of Study (FoS). Researchers and subject experts can swiftly discern FoS trends within their disciplines and observe the pioneers or contributors who initiated or popularized specific trends. This information can assist policymakers, academic publishers, journal editors, institutional funding bodies, and other stakeholders in allocating research resources more effectively to specific FoS and subject fields with greater confidence.

In this research, our focus was on identifying authors who served as trendsetters and followers during the early stages of FoS trends in the field of Computer Science, utilizing the Microsoft Academic Graph (MAG) encompassing research papers from 1950 to 2018. Each paper in MAG is associated with a list of FoS. Our approach involved determining the debut year of an FoS, followed by calculating the FoS publication count, author count, and FoS trend using FoM with a degree centrality measure. We then applied Rogers' theory to detect trendsetters and followers. Finally, we compared our identified trendsetters with two existing lists of highly recognized Computer Science scientists: influential authors and those with an H-index of 40 or higher. The experimental results indicated that our approach successfully identified many influential researchers as listed in these rankings, with some cases showing exact matches between their recognition and their role as trendsetters in specific FoS.

5.1. Future Work

For future studies, we intend to explore additional features such as author collaborations and publication venues. This exploration may provide insights into other dynamics associated with new FoS trends, such as the pace of collaboration among prominent researchers or the popularity of new FoS in scientific venues. Moreover, we aim to utilize more up-to-date scientific datasets to further advance our understanding in this area.

References

1. Salatino AA, Osborne F, Motta E. AUGUR: forecasting the emergence of new research topics. InProceedings of the 18th ACM/IEEE on joint conference on digital libraries 2018 May 23 (pp. 303-312).
2. Effendy S, Yap RH. Analysing trends in computer science research: A preliminary study using the microsoft academic graph. InProceedings of the 26th international conference on world wide web companion 2017 Apr 3 (pp. 1245-1250).
3. Sparavigna AC, Marazzato R. Using Google Ngram viewer for scientific referencing and history of science. arXiv preprint arXiv:1512.01364. 2015 Dec 4.
4. Salatino AA, Osborne F, Motta E. How are topics born? Understanding the research dynamics preceding the emergence of new areas. PeerJ Computer Science. 2017 Jun 19;3:e119.
5. Codd EF. Extending the database relational model to capture more meaning. ACM Transactions on Database Systems (TODS). 1979 Dec 1;4(4):397-434.
6. Berners-Lee T, Cailliau R, Luotonen A, Nielsen HF, Secret A. The world-wide web. Communications of the ACM. 1994 Aug 1;37(8):76-82.
7. Hofmann T. Probabilistic latent semantic indexing. InProceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval 1999 Aug 1 (pp. 50-57).
8. Di Caro L, Guerzoni M, Nuccio M, Siragusa G. A bimodal network approach to model topic dynamics. arXiv preprint arXiv:1709.09373. 2017 Sep 27.
9. Bettencourt LM, Kaiser DI, Kaur J. Scientific discovery and topological transitions in collaboration networks. Journal of Informetrics. 2009 Jul 1;3(3):210-21.
10. Jo Y, Lagoze C, Giles CL. Detecting research topics via the correlation between graphs and texts. InProceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining 2007 Aug 12 (pp. 370-379).
11. Shibata N, Kajikawa Y, Takeda Y, Matsushima K. Detecting emerging research fronts based on topological measures in citation networks of scientific publications. Technovation. 2008 Nov 1;28(11):758-75.
12. Guo H, Weingart S, Börner K. Mixed-indicators model for identifying emerging research areas. Scientometrics. 2011 Oct 1;89(1):421-35.
13. Furukawa T, Mori K, Arino K, Hayashi K, Shirakawa N. Identifying the evolutionary process of emerging technologies: A chronological network analysis of World Wide Web conference sessions. Technological Forecasting and Social Change. 2015 Feb 1;91:280-94.
14. Morinaga S, Yamanishi K. Tracking dynamics of topic trends using a finite mixture model. InProceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining 2004 Aug 22 (pp. 811-816).
15. Leydesdorff L, Rafols I, Chen C. Interactive overlays of journals and the measurement of interdisciplinarity on the basis of aggregated journal–journal citations. Journal of the American society for Information science and Technology. 2013 Dec;64(12):2573-86.
16. Boyack KW, Klavans R, Börner K. Mapping the backbone of science. Scientometrics. 2005 Aug;64:351-74.
17. Small H, Boyack KW, Klavans R. Identifying emerging topics in science and technology. Research policy. 2014 Oct 1;43(8):1450-67.
18. Kleinberg J. Bursty and hierarchical structure in streams. InProceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining 2002 Jul 23 (pp. 91-101).
19. Decker SL, Aleman-Meza B, Cameron D, Arpinar IB. Detection of bursty and emerging trends towards identification of researchers at the early stage of trends.
20. Mane KK, Börner K. Mapping topics and topic bursts in PNAS. Proceedings of the national academy of sciences. 2004 Apr 6;101(suppl_1):5287-90.
21. Ke W, Borner K, Viswanath L. Major information visualization authors, papers and topics in the acm library. InIEEE symposium on information visualization 2004 Oct 10 (pp. r1-r1). IEEE.
22. Boyack KW, Mane K, Borner K. Mapping medline papers, genes, and proteins related to melanoma research. InProceedings. Eighth International Conference on Information Visualisation, 2004. IV 2004. 2004 Jul 16 (pp. 965-971). IEEE.
23. He D, Zhu X, Parker DS. How does research evolve? Pattern mining for research meme cycles. In2011 IEEE 11th International Conference on Data Mining 2011 Dec 11 (pp. 1068-1073). IEEE.
24. Dessì D, Osborne F, Recupero DR, Buscaldi D, Motta E. Generating knowledge graphs by employing natural language processing and machine learning techniques within the scholarly domain. Future Generation Computer Systems. 2021 Mar 1;116:253-64.
25. CiteSpace II CC. Detecting and visualizing emerging trends and transient patterns in scientific literature., 2006, 57. DOI: https://doi. org/10.1002/asi.;20317:359-77.

26. Morris SA, Yen G, Wu Z, Asnake B. Time line visualization of research fronts. Journal of the American society for information science and technology. 2003 Mar;54(5):413-22.
27. Small H, Upham P. Citation structure of an emerging research area on the verge of application. Scientometrics. 2009 May;79:365-75.
28. Ding Y. Community detection: Topological vs. topical. Journal of Informetrics. 2011 Oct 1;5(4):498-514.
29. Morris SA. Manifestation of emerging specialties in journal literature: A growth model of papers, references, exemplars, bibliographic coupling, cocitation, and clustering coefficient distribution. Journal of the American Society for Information Science and Technology. 2005 Oct;56(12):1250-73.
30. Astrom F. Changes in the LIS research front: Time-sliced cocitation analyses of LIS journal articles, 1990–2004. Journal of the American Society for Information Science and Technology. 2007 May;58(7):947-57.
31. Rao CR. Diversity: Its measurement, decomposition, apportionment and analysis. Sankhyā: The Indian Journal of Statistics, Series A. 1982 Feb 1:1-22.
32. Rafols I, Porter AL, Leydesdorff L. Science overlay maps: A new tool for research policy and library management. Journal of the American Society for information Science and Technology. 2010 Sep;61(9):1871-87.
33. Zafar L, Masood N. Impact of Field of Study Trend on Scientific Articles. IEEE Access. 2020 Jul6;8:128295-307.
34. Miller RL. Rogers' innovation diffusion theory (1962, 1995). InInformation seeking behavior and technology adoption: Theories and trends 2015 (pp. 261-274). IGI Global.
35. Holden C. Random samples: Data point-impact factor. Science. 2005;309(8):1181c.
36. Sinha A, Shen Z, Song Y, Ma H, Eide D, Hsu BJ, Wang K. An overview of microsoft academic service (mas) and applications. InProceedings of the 24th international conference on world wide web 2015 May 18 (pp. 243-246).
37. Allan J, Carbonell JG, Doddington G, Yamron J, Yang Y. Topic detection and tracking pilot study final report. 1998.
38. Ho JC, Saw EC, Lu LY, Liu JS. Technological barriers and research trends in fuel cell technologies: A citation network analysis. Technological Forecasting and Social Change. 2014 Feb 1;82:66-79.
39. Stirling A. A general framework for analysing diversity in science, technology and society. Journal of the Royal Society interface. 2007 Aug 22;4(15):707-19.
40. 39. Salatino AA, Mannocci A, Osborne F. Detection, analysis, and prediction of research topics with scientific knowledge graphs. Predicting the Dynamics of Research Impact. 2021:225-52.
41. Takeda Y, Kajikawa Y. Optics: A bibliometric approach to detect emerging research domains and intellectual bases. Scientometrics. 2009 Mar;78:543-58.