We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Scientometrics of Scientometrics: Mapping Historical Footprint and Emerging Technologies in Scientometrics

Meen Chul Kim and Yongjun Zhu

Additional information is available at the end of the chapter

Abstract

Scientometrics is the study of quantitative aspects of science, technology, and innovation. This chapter identifies thematic patterns and emerging trends of the published literature in scientometrics using a variety of tools and techniques, including CiteSpace, VOSviewer, and dynamic topic modeling. Using 8098 bibliographic records of published scientometrics research, we explored domain-level citation paths, subject category assignment, keyword co-occurrence, topic models, and document co-citation network to map and characterize the intellectual landscapes of scientometrics. Findings reveal that the domain is multi-disciplinary in that a wide range of disciplines contribute to the growth of literature, but only partially interdisciplinary as some works heavily cites from similar domains. Early literature was interested in measuring the impact of a science and evaluating research performance and productivity. Modeling scientometrics laws and indicators is also of greatest interest. Later work explored applications of scientometrics to a variety of domains such as material sciences, medicine, environmental sciences, and social media analytics. Impact measure and science mapping are among the topics receiving consistent attention.

Keywords: scientometrics, science mapping, domain analysis, visual analytics, intellectual structure, emerging technologies

1. Introduction

Scientometrics is the quantitative study of science. It aims to analyze and evaluate science, technology, and innovation. Major research includes measuring the impact of authors, publications, journals, institutes, and countries as referenced to sets of scientific publications such as

articles and patents. It also aims to understand the behavior of scientific citations as a mean of scholarly communication and map intellectual landscapes of a science. Other effort focuses on the production of indicators for use in the evaluation of performance and productivity [1]. In practice, there is a significant overlap between scientometrics and other neighboring domains such as bibliometrics, informetrics, webometrics, and cybermetrics. Bibliometrics, one of the canonical research domains in library and information science, studies quantitative aspects of written publications. Informetrics is the study of quantitative aspects of information [2], regarded as an umbrella domain overarching the rest of them. Björneborn and Ingwersen [3] describe the relationships between these domains as abstracted in **Figure 1**.

Driven by a variety of research communities, the volume of published literature in these domains has exponentially grown. Given the increasing publications and the scientific diversity in disciplines, a systematic investigation of the intellectual structure is in need to identify not only emerging trends and new developments but also historic areas of innovation and current challenges. The motivation of the present chapter lies in our intention to identify the intellectual structure of scientometrics in a systematic manner. Toward that end, we explore epistemological characteristics, thematic patterns, and emerging trends of the field, using scientometrics approaches. In particular, we operationalize scientometrics as encompassing closely related domains such as informetrics, bibliometrics, cybermetrics, and webometrics. In the rest of this manuscript, we use the term "scientometrics" inclusively. The present chapter aims to trace the evolution and applications of scientific knowledge in scientometrics. Thus, we also operationalize emerging trends and recent developments uncovered throughout the present chapter as "emerging technologies" in scientometrics.

The contributions of the present chapter include followings. First, it helps the scientometrics community to be more self-explanatory as it has a detailed publication-based profile. Secondly, researchers in the field can benefit from this systematic domain analysis by
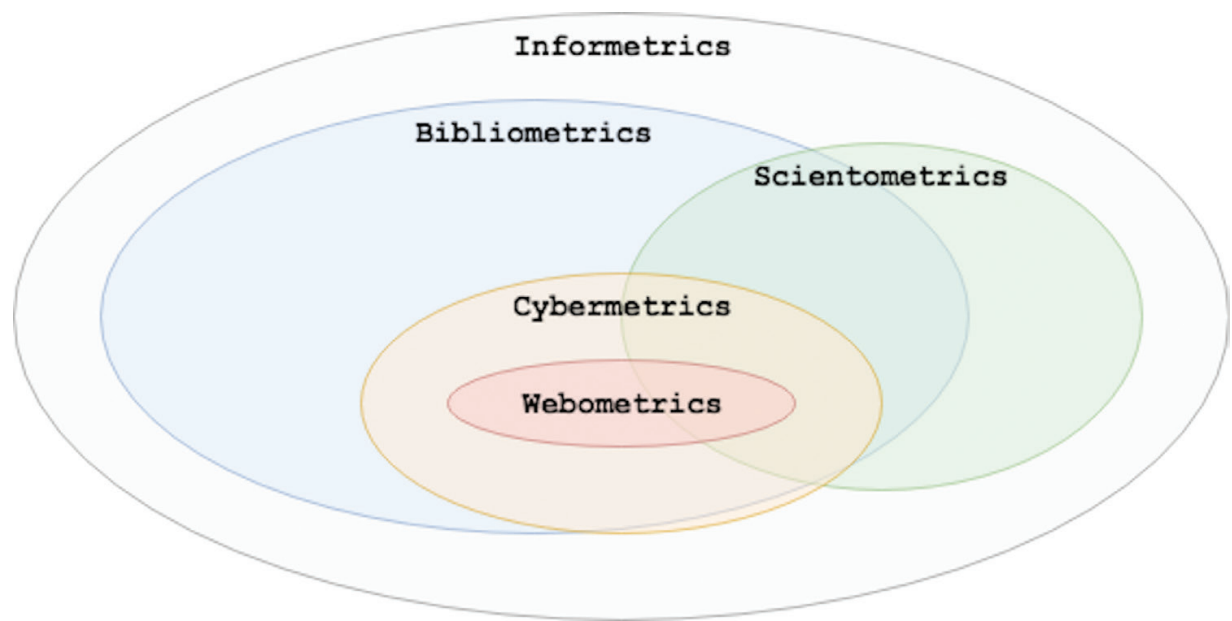
Figure 1. Relationships between metrics sciences re-cited from [3].

identifying emerging technologies, better positioning their research, and expanding research territories. Finally, it guides those interested in the field to learn about historic footprint and current issues.

The rest of the chapter is organized as follows. We introduce the methodology of the study. Then, the intellectual landscapes of scientometrics is described. We conclude this chapter with discussion into findings, implications, and limitations.

2. Methodology

This section details our data collection method and analytical approaches. **Figure 2** pipelines the research procedure.

2.1. Data collection

The present chapter explores the intellectual structure of published literature in scientometrics. Considering the aforementioned operationalization of scientometrics, we conducted a topic search on the web of science (WoS). The search query consisted of seven terms as follows: Bibliometric* OR scientometric* OR informetric* OR webometric* OR altmetric* OR cybermetric* OR entitymetric*. The wildcard character "*" captures any relevant variations of a term such as bibliometrics and bibliometric analysis. A bibliographic record is considered as relevant if any of the terms appear in its title, abstract, or keywords. As of December 31, 2017, the query returned 8098 bibliographic records written in English between 1990 and 2017. The subscription of the authors' institutes covered from 1980s at the time of querying, but in many
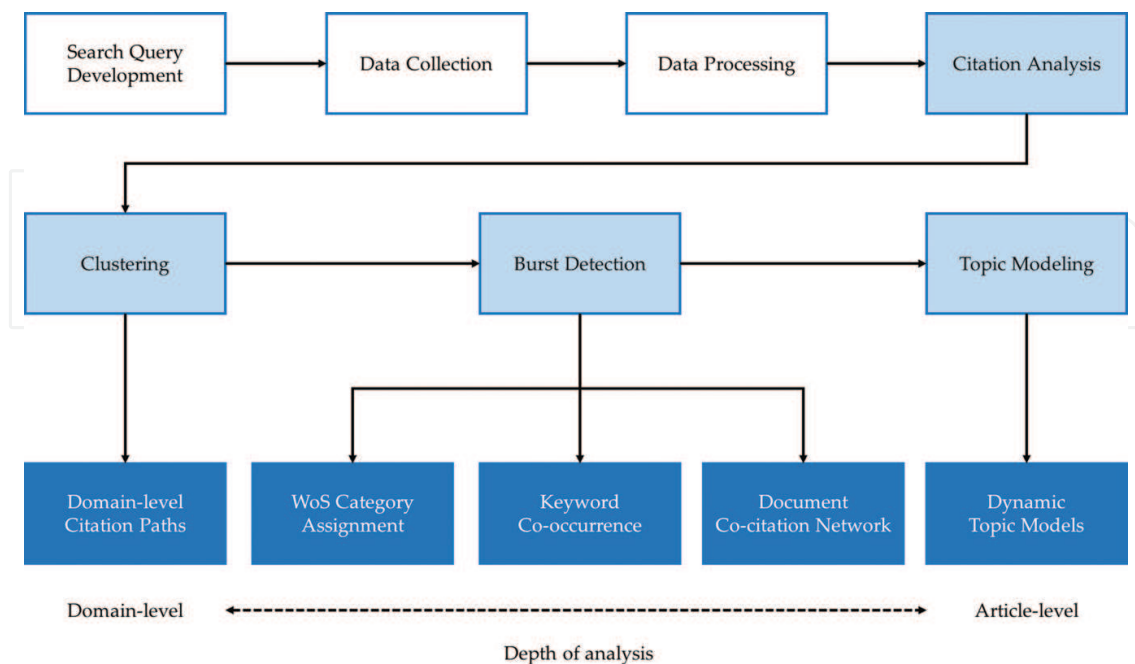
Figure 2. Research procedure.

cases text fields were omitted. Thus, we excluded data before 1990. The brief statistics of the retrieved data set is described in **Table 1**.

Figure 3 renders the record distribution over time in our data collection. As illustrated, there has been exponentially increasing interest in scientometrics from the community.

Table 2 describes the contributing terms to the data retrieval and corresponding number of records to each term. As shown, the literature has used "bibliometric*" the most frequently.

2.2. Investigating the intellectual structure in scientometrics

Scientometrics depicts the intellectual landscapes of a science with a variety of bibliographic units such as authors, keywords, texts, and citations and networks of those entities. The present chapter systematically mapped historical footprint and emerging technologies from published research in scientometrics. In particular, we investigated citation paths at a disciplinary level, co-occurrence of WoS categories and keywords, and networks of co-cited references. Network clustering and topic modeling were also used to find homogeneous sets of literature and coherent streams of research. In so doing, we captured emerging trends, recent developments, and current challenges in the domain. Especially, we employed a top-down approach in analyzing data going from macro-level to micro-level. It had us add richer interpretations as we gradually moved on to lower-level units of analysis such as journal-level citation paths, subject categories, keywords, titles and abstracts to cited references. To this end, this chapter is mainly guided by

Duration	Total	Articles	Procs.	Reviews	Authors	Keywords	Refs.
1990–2017	8098	7013	413	672	23,791	98,493	328,096

Table 1. Data statistics.
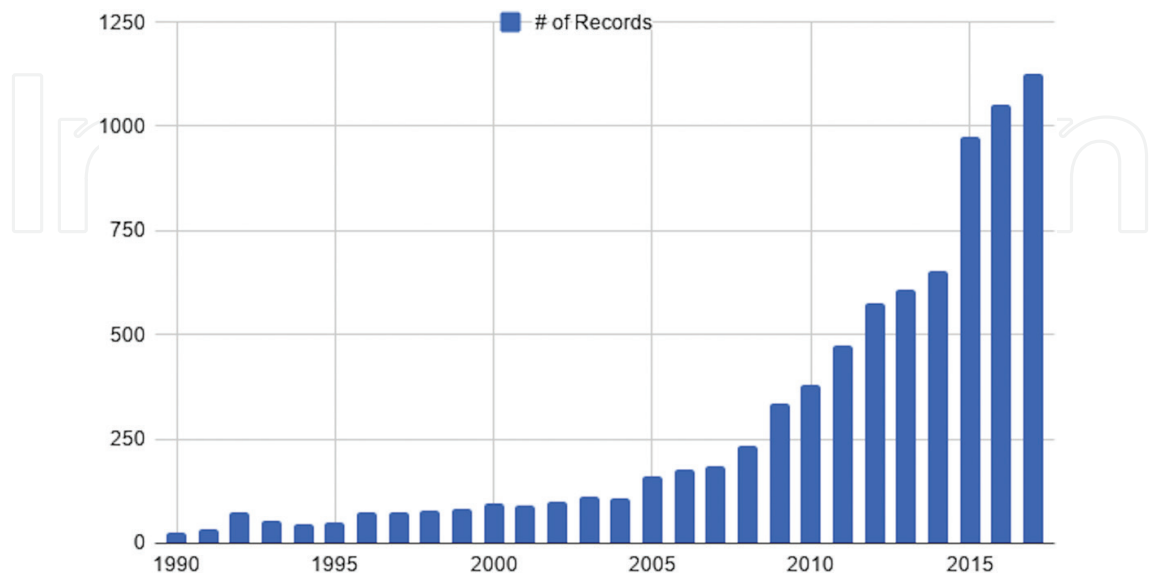
Figure 3. The distribution of records over time.

Term	Duration	Total	Articles	Procs.	Reviews
bibliometric*	1990–2017	6352	5449	313	590
scientometric*	1990–2017	1779	1577	93	109
informetric*	1990–2017	382	334	28	20
webometric*	1997–2017	288	254	25	9
altmetric*	2012–2017	261	237	7	17
cybermetric*	1999–2015	28	27	1	—
entitymetric*	2013–2015	3	3	—	—

Table 2. Querying terms (the wildcard character "*" captures any relevant variations of a term).

two suites of software, namely CiteSpace [4–6] and VOSviewer [7]. The input is a collection of bibliographic records relevant to a topic of interest. Given the records, the toolkits detect and render thematic patterns and emerging trends in science as networked in a variety of bibliographic units. As argued by preceding papers [8, 9], this chapter's approaches have several methodological merits over a conventional domain analysis. First, a much more inclusive range of topically relevant literature can be examined. Second, an inquiring individual does not need prior expertise to analyze a domain of interest. Finally, this kind of survey can be conducted as frequently as in need given the fast growth of a science. The underlying techniques and findings of the present chapter could be more clearly delivered as we introduce followings:

- Network reduction: In network analysis, investigating the entire nodes and edges between them is computationally challenging. It may not intuitively communicate the topological structure to the audience as well for it is visually overwhelming with many links. To handle this, we select up to 100 frequently occurring entities such as keywords and cited references within a one-year time slice.

- Clustering: Clustering is unsupervised learning which uncover latent groups of entities sharing homogeneous characteristics. We employ a network clustering technique called smart local moving [10] to capture thematically similar clusters on a document co-citation network.

- Burst detection: Proposed by [11], burst detection models the burstiness of features which rise sharply in frequency. An entity has bursting activities when it intensively appears during a specific span of time. We can overcome the limitation coming from considering cumulative, snapshot metrics as impact measures.

- Cluster labeling: CiteSpace labels clusters with extracted terms from titles and abstracts of citing articles. There are three algorithms to serve cluster labeling: (1) latent semantic analysis (LSA), (2) log-likelihood ratio (LLR), and (3) mutual information (MI). LSA captures unknown semantic relationships over all the documents while LLR and MI reflect a unique aspect of a cluster [5].

- Topic modeling: Topic modeling is unsupervised machine learning which aims to discover latent semantic structure occurring in a text body. We employ dynamic topic modeling

(DTM) which is a generative technique extended from Latent Dirichlet Allocation (LDA). DTM captures the evolution of latent topics in a collection of documents whereas it was oblivious to the preceding model [12].

3. Results

3.1. Domain-level research patterns

Citation paths at a disciplinary level are depicted in the visual representation called a dual-map overlay [6] (see **Figure 4**). The left regions represent where the collected literature publishes while the right regions render where it cites from. Citing literature and cited literature are also called research frontier and knowledge base respectively. The base map consists of the journal/conference-level citation relationships among over 10,000 venues. Major clusters are labeled by terms chosen from the titles of venues in corresponding clusters. First, all of the terms' log-likelihood ratios are calculated based on their frequency in clusters. The use of LLR achieves to represent those terms' uniqueness in clusters. Then, top three terms are selected to tag clusters, based on their LLR values in descending order. Citation trajectories are colored based on the citing regions. The width of the paths is proportional to the z-score-scaled citation frequency.

Table 3 describes these trajectories in descending order of the third column, namely Z-score. The color of each row is corresponding to the path. Findings indicate that scientometrics has been largely driven by social sciences and medicine as represented by "psychology, education, health" and "medicine, medical, clinical" respectively at the first column. Literature from social sciences heavily cites from "psychology, education, social", "systems, computing, computer", "health, nursing, medicine", "economics, economic, political", and "molecular, biology, genetics", yielding five citation paths. Research frontiers from medicine are based on "health, nursing, medicine" and "molecular, biology, genetics", having two additional trajectories. These observations
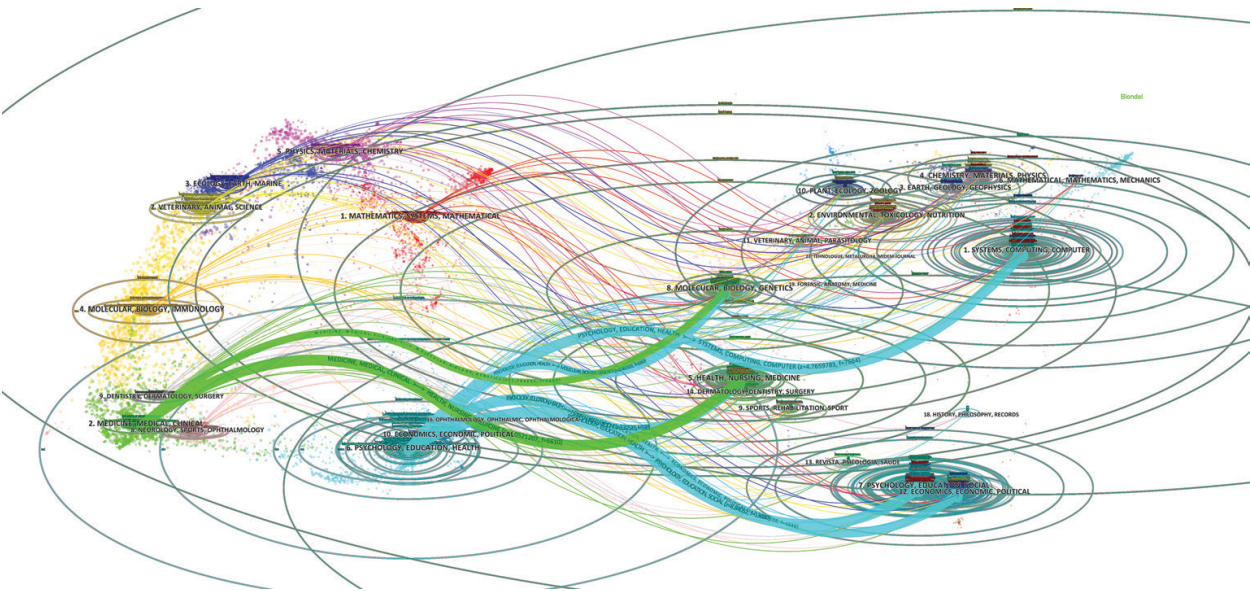

Figure 4. Citation paths at a disciplinary level.

Research frontier	Knowledge base	Z-score
Psychology, education, health	Psychology, education, social	8.841
Psychology, education, health	Systems, computing, computer	4.766
Medicine, medical, clinical	Health, nursing, medicine	4.052
Psychology, education, health	Health, nursing, medicine	3.313
Psychology, education, health	Economics, economic, political	2.724
Psychology, education, health	Molecular, biology, genetics	2.461
Medicine, medical, clinical	Molecular, biology, genetics	1.984

Table 3. Domain-level citation trends.

WoS category	Year	Frequency	Density
Information science & library science	1990	3880	138.571
Computer science	1990	3260	116.429
Computer science, interdisciplinary applications	1990	2284	81.571
Computer science, information systems	1990	925	33.036
Business & economics	1992	653	25.115
Management	1992	374	14.385
Engineering	1992	292	11.231
Public administration	1992	199	7.654
Planning & development	1992	179	6.885
Education & educational research	1992	165	6.346
Social sciences – other topics	1992	160	6.154
Science & technology – other topics	1993	462	18.480
Multidisciplinary sciences	1993	348	13.920
Business	1994	242	10.083
Neurosciences & neurology	1996	159	7.227
Environmental sciences & ecology	1997	261	12.429
General & internal medicine	1999	145	7.632
Surgery	2000	162	9.000
Public, environmental & occupational health	2003	201	13.400
Environmental sciences	2006	189	15.750

Table 4. Top 20 frequently assigned WoS categories.

show scientometrics is multidisciplinary and partially interdisciplinary; Multidisciplinary since scientometrics research has been published in multiple disciplines; Partially interdisciplinary for literature published in "psychology, education, health" has a variety of intellectual bases while "medicine, medical, clinical" largely cites from neighboring domains.

We considered WoS category assignment to literature as another important indicator representing domain-level thematic concentration. The top 20 frequently assigned WoS categories to the records are described in **Table 4**. It shows the year it was first assigned, and the density of how many times per year a specific category has been given, from its first year. The table is sorted in ascending order of the year. Results show that three categories have been assigned more than 2000 times – "information science & library science" (n = 3880), "computer science" (n = 3260), and "computer science, interdisciplinary applications" (n = 2284). These categories were first assigned from the beginning in the data set, demonstrating the greatest densities. The most frequently assigned category to be added to the top four list is "computer science, information systems." This category also demonstrates a relatively high density (33.036), given its first year of assignment was 1990. This finding suggests that literature under these four categories has had the largest influence on the emergence and development of scientific knowledge in scientometrics. In turn, research with scientific foci in social sciences, engineering, medical & health sciences, and environmental sciences brought along a multidisciplinary grasp to the domain.

3.2. Trending keywords

Given by authors and indexers, keywords reflect representative concepts underlying published literature. The top 20 frequently occurring keywords in the data set are described in **Table 5**. It shows the year it first appeared, and the density of how many times on average a specific keyword has appeared, from its first year. Findings indicate that in the beginning, "bibliometrics" and "scientometrics" focused on employing "citation analysis" to examine the "impact" of a "science". We assume that "journal" and "publication" were considered as units of analysis. Another effort focused on evaluating research "performance" and "productivity" and examining the "pattern" of scientific "collaboration." The other stream of research had interest in devising a "bibliometric indicator" such as journal "impact factor", which led to the recent development of the widely accepted author-level metric "h-index."

Keyword	Year	Frequency	Density
Science	1991	1613	59.741
Bibliometric analysis	1991	871	32.259
Journal	1991	815	30.185
Citation	1991	803	29.741
Bibliometrics	1992	1914	73.615
Impact	1992	969	37.269
Citation analysis	1992	814	31.308
Publication	1992	700	26.923
Scientometrics	1992	646	24.846
Indicator	1992	596	22.923
Performance	1992	348	13.385

Keyword	Year	Frequency	Density
Productivity	1992	270	10.385
Collaboration	1993	353	14.120
Bibliometric indicator	1993	290	11.600
Pattern	1993	273	10.920
Network	1994	357	14.875
Impact factor	1996	527	23.955
Index	2002	324	20.250
h-index	2007	386	35.091
Scopus	2008	280	28.000

Table 5. Top 20 frequently occurring keywords.

Figure 5 displays the keyword co-occurrence in the data set. We used a technique called a density visualization guided by VOSviewer. The font size of a keyword is proportional to its occurrence frequency. The more frequently a pair of keywords co-occurs, the closer the pair is located to the red spots. The visualization resulted in 484 keywords which occurred more than or equal to 18 times. As depicted, "bibliometrics" frequently co-occurred with "impact" which is consistent with the finding above. It also determined that devising an "impact factor" for "journal ranking" was among the important themes in scientometrics.

Table 6 lists 20 keywords which have surged during a specific duration of time. The investigation of keyword bursts adds temporal contexts in understanding historic footprint and emerging technologies in scientometrics which were oblivious to the snapshot metrics. The keywords were sorted in ascending order of the beginning years of bursts. "physics" is one of the keywords with the longest bursts, ending in 2010. It also has the second strongest bursts when not including "science." It indicates applications of scientometrics to physics and/or knowledge transfer from physics to scientometrics had intensively been conducted from the early years. The widely accepted author-level metric, namely h-index, was also derived from physics. The second longest bursts from 1992 is led by "law", also demonstrating a relatively high value of bursts. It shows the identification of laws existing in scientometrics phenomena was among the important initiatives. "publication output" is the keyword with the third longest and strongest bursts. It is argued that the evaluation of research performance and productivity was one of the key themes in the domain. The strongest burst episode from 1992 is associated with "indicator." In consideration with other keywords such as "stationary distribution", "model", and "informetric distribution", we argue modeling an indicator of impact measure was of greatest interest in scientometrics.

3.3. Temporal topic models

We analyzed another text fields, namely titles and abstracts since more informational points of content can be examined than only exploring keywords. We aimed to uncover the evolution of latent topics in the records over time. Toward that end, we removed stop words from
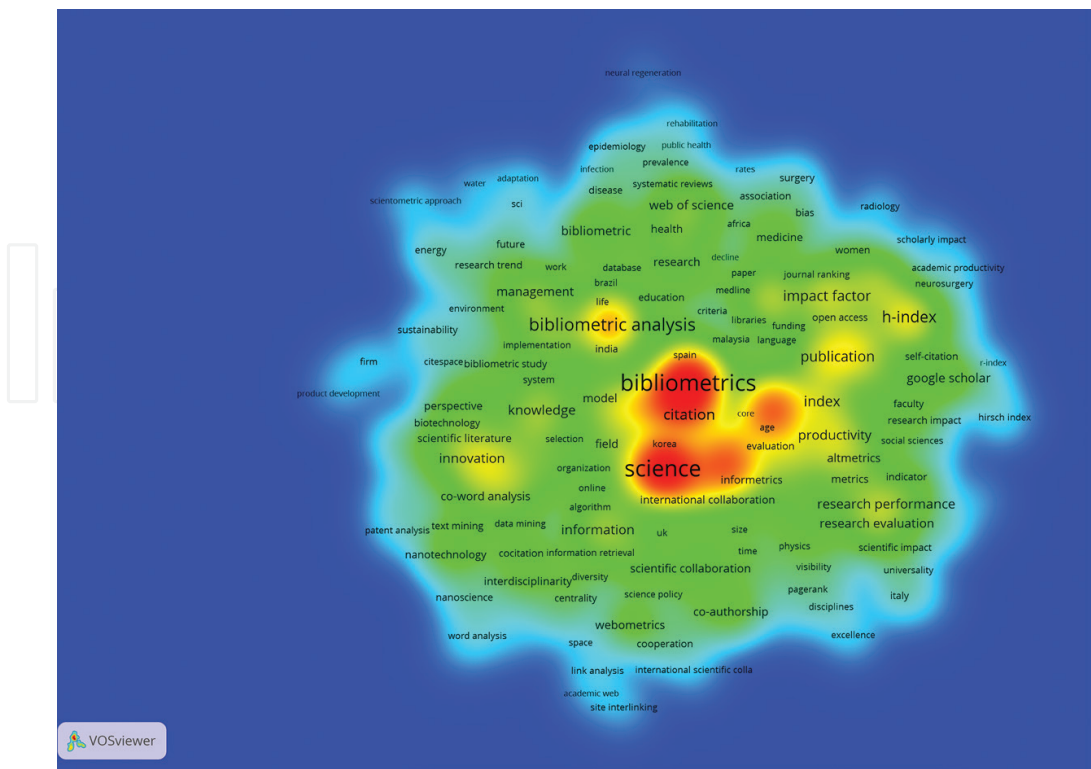
Figure 5. Keyword co-occurrence network (n = 484).

text, using a list of stop words in Python NLTK. The text was lowercased, tokenized, and de-accented. Then, we lemmatized the tokens and extracted noun phrases by bigram indexing. Text pre-processing and topic modeling were driven by gensim, a robust text mining toolkit in Python. **Table 7** describes 20 topics and 10 corresponding terms per topic. The terms were sorted in descending order of the average probabilities over the 28 years. Results show that most of the terms having high probabilities are unigram-formed.

Figure 6 illustrates the topical trends from 1990 till 2017 using a visualization technique called a bump chart. The topics are sorted in descending order of normalized probability distributions in the beginning year. We further discuss nine prominent topics, Topics 9, 17, 7, 4, 1, 5, 11, 16, and 0, due to their relatively high probability distributions. We categorized these topics into four trends: (1) rising, (2) rising-falling, (3) falling, and (4) static.

1. Rising topics: Topics 9, 17, 7, and 1 are consistently rising. Topic 9 we labeled "applications of scientometrics to material sciences" has received the greatest attention over time. Topic 17 which has sharply increased is named "publication-based scholarly communication." Topics 7 and 1 have been always in the top topic list and recently received increasing attention. We labeled them "evaluation of funded research" and "applications of scientometrics to medical education" respectively. Findings indicate that applications of scientometrics to domains other than biomedical sciences are of increasing concerns in the scientific community.

2. Rising-falling topics: Topics 4, 16, and 0 repeat rising and falling. Topic 4 can be named "literature-based research in healthcare." Topics 16 and 0 can be understood as "applications

of scientometrics to biomedicine" and "literature-based research in medicine" respectively. Knowledge discovery in healthcare and biomedical sciences has been among the greatest interest in scientometrics. We assume that this stream of research has ups and downs based on the change of scientific foci.

3. Falling topics: Topic 5 has fallen. We labeled it "history and philosophy of scientometrics." It is obvious that a study of theory and practice tends to be prominent in early years of a science. As staging into the maturation, this kind of topic naturally moves way from interest. It has also decreased in scientometrics.

4. Static topics: Topic 11 has been statically distributed over time. Based on the extracted terms, Topic 11 is interpreted as "mapping intellectual structure using citation and network analysis." This is one of the canonical research themes in scientometrics receiving consistent attention from the beginning of the domain.

Keyword	Burst	Begin	End	1990 - 2017
science	30.062	1990	2001	
british science	3.371	1990	1997	
stationary distribution	6.931	1992	2003	
model	10.688	1992	2002	
fact	4.116	1992	1993	
citation impact	4.850	1992	1997	
library	10.410	1992	2004	
physics	13.560	1992	2010	
co-citation analysis	3.430	1992	1993	
informetric distribution	6.299	1992	2006	
relative citation impact	3.430	1992	1993	
newest version	4.116	1992	1993	
indicator	17.159	1992	2000	
country	10.903	1992	2001	
law	10.395	1992	2007	
publication output	11.459	1992	2006	
figure	4.116	1992	1993	
cooperation	4.189	1993	2005	
combined co-citation	3.607	1994	2005	
journal	4.427	1994	1996	

Table 6. Top 20 keywords with the greatest intensive burstiness.

Topic 0	Topic 1	Topic 2	Topic 3
article	psychology	publication	productivity
journal	education	cancer	faculty
author	nursing	document	publication
article published	Brazilian	drug	index
number	research	research	gender
literature	study	descriptor	result
study	psychiatry	Korean	study
research	theses	Latin American	conclusion
medicine	school	literature	woman
publication	aids	drug	year
Topic 4	**Topic 5**	**Topic 6**	**Topic 7**
health	science	research	research
research	history	country	evaluation
publication	scientometrics	science	impact
public health	book	collaboration	funding
literature	reception	publication	assessment
medicine	removal	output	policy
method	philosophy	physics	researcher
result	nature	university	project
disease	colleague	study	scientist
health care	sport	productivity	work
Topic 8	**Topic 9**	**Topic 10**	**Topic 11**
performance	technology	research	structure
indicator	literature	field	analysis
research	patent	analysis	map
bibliometric indicator	nanotechnology	information	network
quality	serial	study	mapping
evaluation	indexing	science	citation
group	application	development	data
measure	development	data	cluster
data	material	paper	database
peer review	core	knowledge	method
Topic 12	**Topic 13**	**Topic 14**	**Topic 15**
study	distribution	information	paper
population	model	web	research
method	data	library	publication
country	index	link	literature

| disease | two | use | country |
data	one	online	journal
research	theory	library information	period
result	paper	search	number
health	number	internet	sci
water	function	study	study
capacity	system	subject	bibliometric analysis
Topic 16	**Topic 17**	**Topic 18**	**Topic 19**
research	communication	journal	ecology
rehabilitation	bibliometrics	citation	species
stem cell	scholarly communication	analysis	geography
neuroscience	dss	impact	climate change
credit	publishing	study	city
guideline	science	impact factor	conservation
paper	library information	paper	knowledge
study	media	reference	biodiversity
transplantation	theory	science	tourism
article	impact	author	study

Table 7. 20 generated topics.
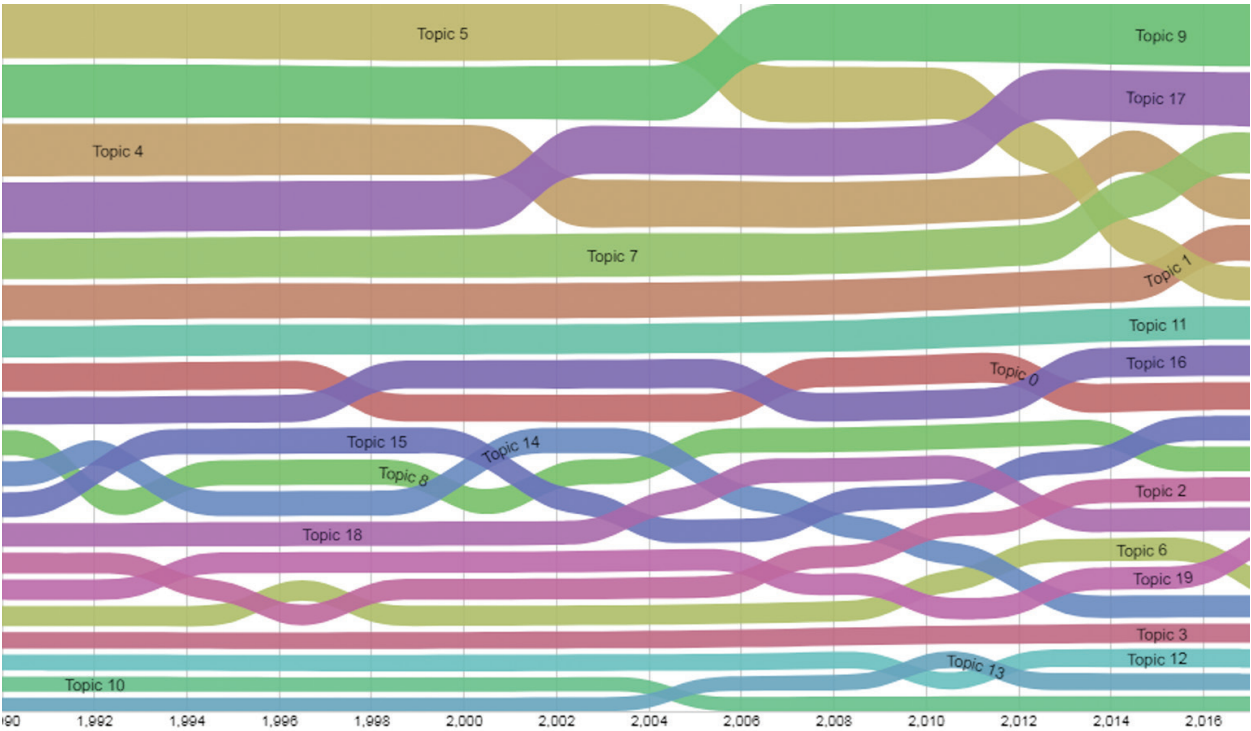
Figure 6. Topical trends.

### 3.4. Document co-citation network

Previous section utilized titles and abstracts to investigate topical trends without any bound context. This section examined those fields in a context of document-level co-citation relationship. **Figure 7** visualizes the document co-citation network in the data set. Each node is a cited reference extracted from the reference sections of the records and the size of the node is proportional to its cumulative frequency of received citations. Nodes with inner circles in
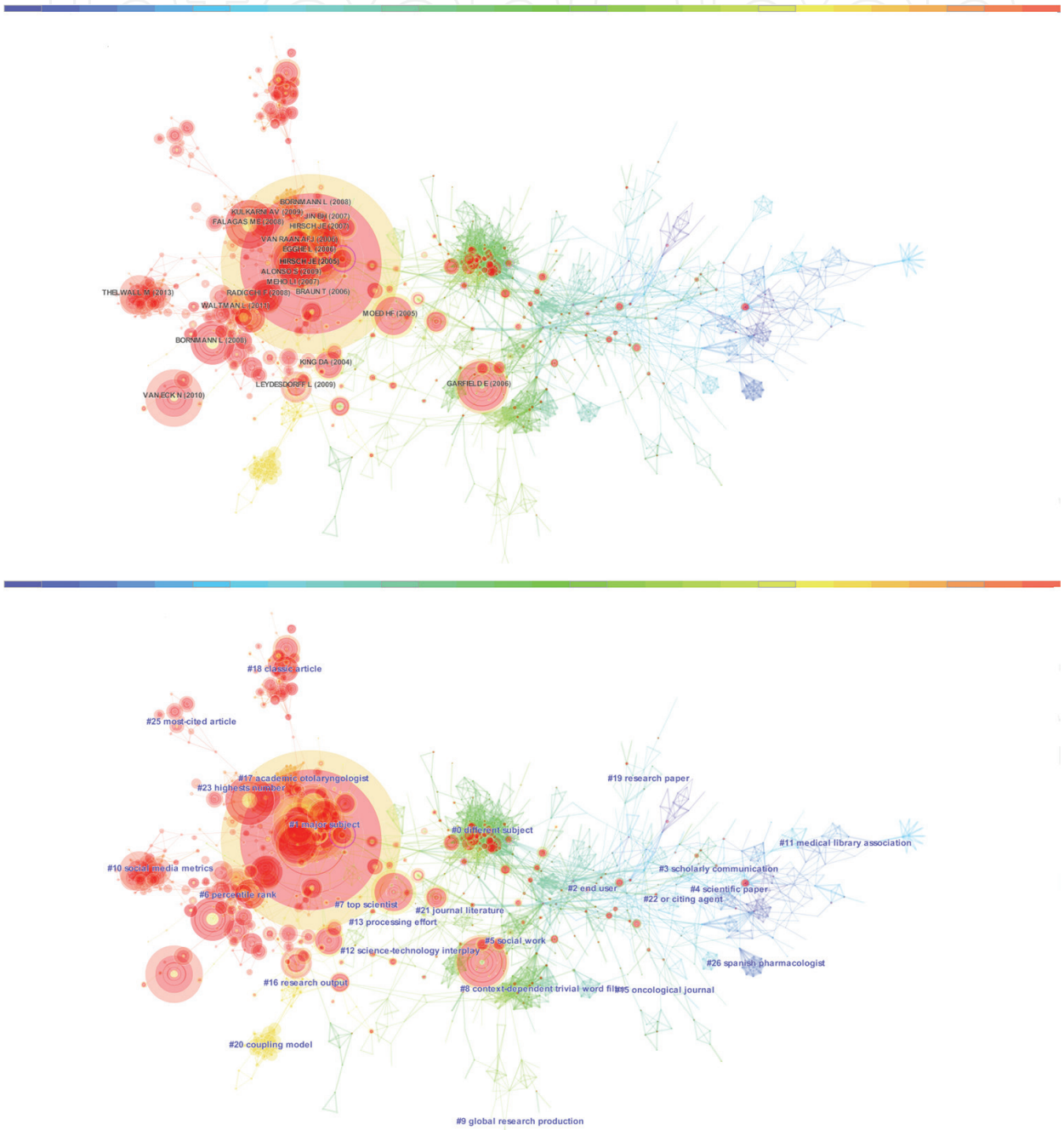
Figure 7. Document co-citation networks with truncated labels of first authors' names and published years (upward) and cluster labels (downward) (n = 1856, e = 6127).

red represent articles with citation bursts. We labeled the most highly cited 20 articles in black following a truncated form of <LAST NAME> < ABBREVIATED FIRST NAME> (<YEAR>) so as to only display first authors' names and published years (see the upward in **Figure 7**). They are cited more than or equal to 95 times locally, meaning in the data set. The color legend at the top of the display indicates links and citations in cooler colors happen more closely to 1990 whereas hotter ones occur in closer years to 2017. Based on the color scheme, we can keep track of the evolution of the document network. Findings show that most of the landmark articles were published relatively recently. Cumulative citations and citation bursts also intensively happened with these articles. Next, we conducted clustering and labeled the clusters in blue, using LLR (see the downward in **Figure 7**). Clusters are numbered in such a way that higher rankings are given to the clusters containing more references. In order to add richer contexts in interpreting the clustering results, we generated another visualization called a timeline visualization (see **Figure 8**).

In **Figure 8**, we re-grouped all the nodes on multiple lines so that the cluster memberships can be more accessibly identified. As depicted in the figure, emerging trends can further be captured by examining Clusters 1, 6, 10, 16, 17, 18 given cluster sizes, recency, cumulative citations, and citation bursts. **Table 8** summarizes these clusters in terms of cluster size, three types of labels, and mean year of citees, i.e. cluster age. Of the selected clusters, Cluster 1 is the largest and oldest. In consideration with Cluster 6, results show that impact measure is still among the important themes in scientometrics. The third largest and newest group of literature is Cluster 10. It indicates practical applications of social media analytics to scientometrics is receiving the most recent attention. Other emerging topics include international collaboration (Cluster 16) and applications to medicine (Cluster 17) and environmental sciences and policy (Cluster 18).
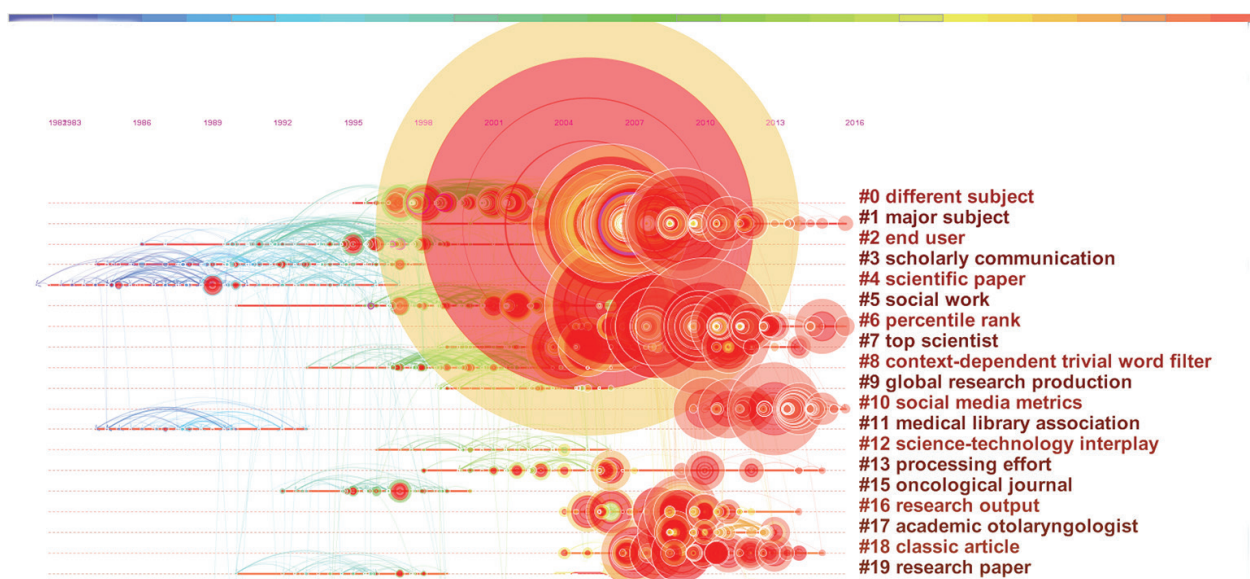
Figure 8. Timeline visualization with LLR cluster labels.

Cluster	Size	Age	Labels		
			LSA	**LLR**	**MI**
1	142	2007	h-Index	Major subject	Productivity incentive
6	74	2010	References	Percentile rank	Average citation
10	47	2013	Papers	Social media metrics	Practical application
16	34	2008	China	Processing effort	Worldwide research productivity
17	32	2011	Documents	Academic otolaryngologist	Peer-reviewed ophthalmology
18	30	2009	Water	Classic article	National policy intervention

Table 8. Cluster summary.

4. Discussion

4.1. Epistemological characteristics

The domain-level investigation revealed the following characteristics of published research in scientometrics. First, scientometrics research is multidisciplinary. Multiple disciplines such as "psychology, education, health" and "medicine, medical, clinical" are engaged in advancing knowledge in the domain. In particular, computer and information sciences had the largest influence on the emergence and development of scientific knowledge. The assignment of WoS categories also evidenced the multidisciplinarity of scientometrics as a variety of domains such as social sciences, engineering, medical and health sciences, and environmental sciences have contributed to the growth of the field. Second, scientometrics is not yet fully interdisciplinary as shown in the finding that research frontiers from "medicine, medical, clinical" largely cites from similar domains. Examining domain-level citation patterns in consideration with the WoS category assignment obtained a solid overview of the publication profile of the field. It revealed the growth of the domain by visualizing the distribution of citation trajectories at a disciplinary level, adding richer contexts with examining the distribution of WoS category assignment. Finally, most of the landmark articles were published relatively recently, namely after 2004 in spite of the long history of the domain. We argue that the domain's maturation is still ongoing.

4.2. Historic footprint and emerging technologies

The analysis of keywords, topic models, and document clusters identified the following thematic patterns in scientometrics research. In the beginning some researchers focused on employing citation analysis to measure the impact of a science. Another effort focused on the evaluation of performance and productivity of research, employing scientometrics approaches. The identification of patterns in scientific collaboration was also among the important themes. The other effort had interest in modeling scientometrics laws and proposing scientometric indicators and impact measures. Recently, applications of scientometrics

approaches to a variety of domains such as material sciences, medicine, and environmental sciences have received increasing attention. In reverse, practical applications of social media analytics to scientometrics is also receiving the most recent interest. Impact measure and science mapping are among the canonical research themes receiving consistent attention from the beginning of the domain.

5. Conclusion

The present chapter aimed to explore epistemological characteristics, historic areas of innovation, and emerging trends in scientometrics. We achieved this by investigating domain-level citation paths, WoS category assignment, keyword co-occurrence, temporal topic models, and document clusters. The findings indicate the domain of scientometrics is multidisciplinary and partially interdisciplinary. Social sciences and biomedicine have published to the field, but not yet cited from each other. We argue that the maturation of scientometrics as a scientific field is still ongoing. Next, early studies tried to measure a science's impact and performance and productivity of published research. Successive effort investigated laws and indicators in scientometrics and explored scientific collaboration. Recent literature is paying attention to topics such as applying scientometrics approaches to different domains and bringing social media analytics in scientometrics.

The approaches of the present study provide advantages in investigating intellectual structure of a science as follows. First, we tried to make our data collection inclusive by investigating closely neighboring domains. Conventional studies of domain analysis often cover only a fraction of published literature. Our method provides a systematic way to explore the broader coverage of a scientific discipline. Second, we investigated the domain from a multi-faceted point of view. Domain-level citation trajectories, subject category assignment, networks of subject categories and keywords, bursting keywords, topic models, and document co-citation networks were identified in this study. Sub-sections in Results triangulated each other, adding richer interpretations from macro units of analysis to micro ones. Finally, the analytical procedure and tools employed in the present work enabled us to explore time-aware research trends in the domain. In addition, one can conduct this kind of domain analysis of his or her concern as frequently as needed without prior knowledge or experience. Thus, the proposed approaches have a relatively higher reproducibility and lower cost for conducting studies at a larger scale, especially as in the era of mass publication.

There are several limitations in our work. First, the topic search we conducted on WoS may have missed relevant records. It is acknowledged that the vocabulary mismatch presents a challenge for keyword-based search. We may be able to overcome this drawback by employing citation indexing or iterative search query development as an alternative strategy in order to capture a much broader context. Second, WoS as our source of data may have under-represented conference proceedings. It is also recognized as an issue for disciplines such as social sciences and arts and humanities [13]. At the time of data retrieval, the authors' institutes only subscribed to the core collection of WoS. Thus, it was inevitable not to miss some

relevant records accordingly. Additional sources such as Scopus are recommended for future refinements of this type of analysis. In addition, some findings or sub-sections in Results may seem too general to characterize emerging technologies in scientometrics when considered independently from the entire context. We argue that that is not because of the limitation of our approaches and tools but due to the characteristics of bibliographic records. That means textual fields that can be used only include titles, abstracts, and keywords which are often abstract to be inclusive. To overcome this, we employed not only frequency-based metrics such as citation counts and latent semantic analysis but also burst detection and probability-oriented techniques such as LLR, MI, and DTM. Then, we tried to triangulate the findings from each sub-section, adding richer interpretations as moving between different units of analysis. We argue that our approaches be more strengthened if we can have access to more informational sources such as full text. Finally, we selected 100 highly cited references to generate the intellectual landscapes. Although this data reduction is in part intuitive, we can strengthen our approach by choosing cited articles based on more refined indicators such as h-index or g-index. It may be worth conducting a separate study of the theoretical implications of using a variety of conceivable selection criteria. We also plan to apply the present chapter's approaches to much more comprehensive records that cover a various type of publication materials.

Conflict of interest

There are no conflicts of interest.

Author details

Meen Chul Kim[1]* and Yongjun Zhu[2]

*Address all correspondence to: meenchul.kim@drexel.edu

1 Drexel University, Philadelphia, PA, USA

2 Sungkyunkwan University, Seoul, South Korea

References

[1] Leydesdorff L, Milojević S. Scientometrics. In: International Encyclopedia of the Social & Behavioral Sciences. 2nd ed. Oxford, UK: Elsevier; 2015

[2] Wolfram D. Applied Informetrics for Information Retrieval Research. Westport, CT: Libraries Unlimited; 2003

[3] Björneborn L, Ingwersen P. Toward a basic framework for webometrics. Journal of the American Society for Information Science and Technology. 2004;**55**(14):1216-1227

[4] Chen C. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. Journal of the American Society for Information Science and Technology. 2006;**57**(3):359-377

[5] Chen C, Ibekwe-SanJuan F, Hou J. The structure and dynamics of co-citation clusters: A multiple-perspective co-citation analysis. Journal of the American Society for Information Science and Technology. 2010;**61**(7):1386-1409

[6] Chen C, Leydesdorff L. Patterns of connections and movements in dual-map overlays: A new method of publication portfolio analysis. Journal of the American Society for Information Science and Technology. 2014;**65**(2):334-351

[7] Van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics. 2010;**84**(2):523-538

[8] Kim MC, Zhu Y, Chen C. How are they different? A quantitative domain comparison of information visualization and data visualization (2000-2014). Scientometrics. 2016;**107**(1):123-165

[9] Zhu Y, Kim MC, Chen C. An investigation of the intellectual structure of opinion mining research. Information Research. 2017;**22**(1): paper 739. http://www.informationr.net/ir/22-1/paper739.html

[10] Waltman L, Eck V, NJ. A smart local moving algorithm for large-scale modularity-based community detection. European Physical Journal B. 2013;**86**(11):471

[11] Kleinberg J. Bursty and hierarchical structure in streams. Data Mining and Knowledge Discovery. 2003;**7**(4):373-397

[12] Blei DM, Lafferty J D. Dynamic topic models. In: Proceedings of the 23rd International Conference on Machine Learning. pp. 113-120

[13] Mongeon P, Paul-Hus A. The journal coverage of web of science and Scopus: A comparative analysis. Scientometrics. 2016;**106**(1):213-228