



# Research on the Characteristics and Trends of MOOC Based on CiteSpace

Xinyuan Wang<sup>1,2,a</sup>, Zhenyang Zhang<sup>2,b,\*</sup>

<sup>1</sup>*School of Economics and Management, Hulunbuir University, Hulunbuir 021000, China*

<sup>2</sup>*Graduate School of Management of Technology, Pukyong National University, Busan 48547, Korea*

<sup>a</sup>*wangxinyuan@pukyong.ac.kr*, <sup>b</sup>*zyzhang@pukyong.ac.kr*

*Corresponding Author: Zhenyang Zhang*

## ABSTRACT

This study uses MOOC as the theme for conducting a bibliometric and visual analysis of literature. The period was limited to 2018-2022, and a total of 1223 eligible articles were screened in the WOS core database, and bibliometric analysis was conducted using CiteSpace. With the continued impact of COVID-19, the number of MOOC users is increasing. The types of MOOC are moving toward specialized courses, and course providers gradually focus on developing the quality of information literacy. Moreover, the teacher status is also gradually important. Combined with the analysis results, this study could provide directional references for further research on MOOC.

**Keywords:** MOOC; Evaluation; Bibliometric; CiteSpace; Visual Analysis

## 1. INTRODUCTION

MOOC (mass open online course) was born with the improvement of information technology and the change in teaching concepts. MOOC teaching resources are distributed through the Internet, including videos, text materials, and test questions. The courses are open to massive people, and learners usually only need to register for the course platform for free to join. The characteristics of MOOCs led to an initial concentration of courses in general education courses, and the providers of such courses were also generally higher education institutions.

Starting with the spread of COVID-19, MOOC courses have become more widespread and more popular. With the help of online platforms, higher education institutions have built or launched more professional courses with their advantages, which can better cultivate talents, develop science and technology, and serve society. The MOOC boom originated in the United States, represented by the success of Coursera, Udacity, and edX, making 2012 the "year of the MOOC". In this context, we believe it is necessary to use bibliometric and visualization methods to analyze MOOC-related literature, find the current research status and new trends, and provide a reference basis for future research directions.

## 2. RESEARCH DESIGN

### 2.1. The source of data

Dave Cormier and Bryan Alexander first introduced the MOOC concept in 2008 [3]. The early success cases started in 2012 and originated in the United States. Therefore, this study selects the English literature in the Web of Science (WOS) Core Collection from January 1, 2008, to March 25, 2022.

### 2.2. Research Methods and Tool

This study adopts a bibliometric approach to study the external characteristics of literature. The bibliometric approach can demonstrate the association between various knowledge units in academic resource research, such as authors, keywords, and countries, by drawing a knowledge map. It transforms a large amount of data, information, and knowledge into a more intuitive and visual form to improve discovery, understanding, analysis, and overall concept formation. It is an essential tool for evaluating the current status of research in a specific field and predicting research trends [4].

There are kinds of software that can be applied for the bibliometric. This study uses CiteSpace, a tool developed by Prof. Chaomei Chen specifically for

Scientometrics and visual analysis to complete the knowledge map [1]. Because the developer is fluent in Chinese and English, CiteSpace is loved by scholars worldwide. The software version is abundant, and scholars could use the corresponding version for different needs. This study involves merging many synonymous keywords, so we choose to use 5.6.R5. This study uses CiteSpace to analyze the literature from the perspective of the number of articles issued, time, country, keywords, and hot trends.

### 2.3. Research procedures

The initial search revealed many MOOC-related articles with many similar terms. In order to keep the research direction consistent, the search considered similar terms and added the term "evaluation" and "evaluate" to narrow the search. The final search formula was ((TS=("massive open online course\*")) OR TS=("Online course\*")) OR TS=("Online course\*") OR TS=("open Online course\*") OR TS=("Mooc\*") AND (TS=(evaluate) or TS=(evaluation)), where "\*" is a wildcard including the noun plural case.

The search type was limited to the article, excluding proceedings papers, early access, review articles, meeting abstracts, editorial materials, and editorial materials to ensure the scientific nature. We manually eliminated articles with poor relevance and finally selected 1223 articles for analysis.

In order to further ensure the quality of the analysis output data, the study organized the retrieval results of data clutter. First, literature de-duplication was achieved

by the Remove Duplicates function in CiteSpace. Second, synonymous keywords are merged; for example, "MOOC" is the abbreviation of "Mass Open Online." After reading the literature and comparing keywords, this study also asked the following keywords to be merged with "MOOC," including "online teaching," "open online course," "online education," "distance learning," "online learning." Thirdly, the words that are not significant and have a low occurrence rate in the literature, such as "suggestions" and "countermeasures," were deleted.

## 3. RESULTS AND DISCUSSION

### 3.1. General Trends and Features

CiteSpace collation and literature reading found that although MOOC was introduced in 2008, the early literature was small. The early topics focused on elaborating basic concepts such as teaching concepts and teaching models. There were relatively few studies on the evaluation of MOOC, as shown in Figure 1. This situation has changed since 2017. Because of the impact of COVID-19, the number of MOOC courses has proliferated, the size of participants has become more extensive, and scholars have put the evaluation of MOOC courses on the agenda. We can see the surge in numbers in 2021 on the graph. The number of publications in 2022 is only 54, but the total number could be large based on the trend line. This indicates that research on MOOC is becoming more and more widespread and valued.

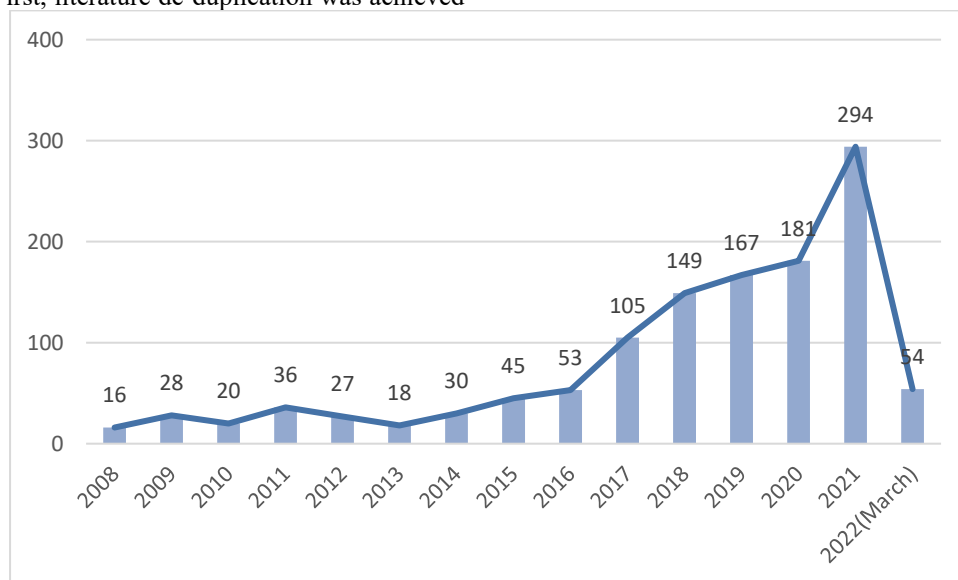


Figure 1: Number of publications from 2008 to 2022 by year

### 3.2. Highly cited literature

The number of citations is an important indicator to evaluate the academic value of literature and the academic status of its authors. According to the WOS

Core Database search, Margaryan et al. (2005) published the most cited article. This article analyzed a random selection of 76 MOOCs and divided the period into xMOOCs and cMOOCs, finding that most MOOCs had suitable course materials and presentations but low

instructional design quality. This article, which has been cited until now, also disguised the need for more outstanding quality evaluation of MOOC courses and

aligned with the following research hotspots. The results of such an analysis validate the scientific validity and accuracy of CiteSpace.

**Table 1** Top 10 highly cited Articles based on WOS Core Collection

Times Cited	Authors	Article Title	Publication Year
361	Margaryan, A; Bianco, M; Littlejohn, A	Instructional quality of Massive Open Online Courses (MOOCs)	2015
245	Kuo, YC; Walker, AE; Schröder, KEE; Belland, BR	Interaction, Internet self-efficacy, and self-regulated learning as predictors of student satisfaction in online education courses	2014
199	Agudo-Peregrina, AF; Iglesias-Pradas, S; Conde-Gonzalez, MA; Hernandez-Garcia, A	Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning	2014
189	Chen, CM; Wu, CH	Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance	2015
187	Monahan, T; McArdle, G; Bertolotto, M	Virtual reality for collaborative e-learning	2008
144	Dziuban, C; Graham, CR; Moskal, PD; Norberg, A; Sicilia, N	Blended learning: the new normal and emerging technologies	2018
137	Schwendimann, BA; Rodriguez-Triana, MJ; Vozniuk, A; Prieto, LP; Boroujeni, MS; Holzer, A; Gillet, D; Dillenbourg, P	Perceiving Learning at a Glance: A Systematic Literature Review of Learning Dashboard Research	2017
138	Schiaffino, S; Garcia, P; Amandi, A	eTeacher: Providing personalized assistance to e-learning students	2008
132	Young, T; Rohwer, A; Volminck, J; Clarke, M	What Are the Effects of Teaching Evidence-Based Health Care (EBHC)? Overview of Systematic Reviews	2014
139	Bell, F	Connectivism: Its Place in Theory-Informed Research and Innovation in Technology-Enabled Learning	2011

### 3.3. Country Productivity

We used CiteSpace for country collaboration network analysis, as shown in Figure 2. 109 nodes (countries), 400 connections (collaborations), and a density of 0.068 were obtained. The country nodes differ in size. The United States of America has the most

significant node, followed by CHINA, SPAIN, AUSTRALIA, and ENGLAND. This ranking is consistent with the previous analysis. The booming development of MOOC is precisely from the United States, China due to the number of schools and students that are mainly used, and the research outcomes are abundant.

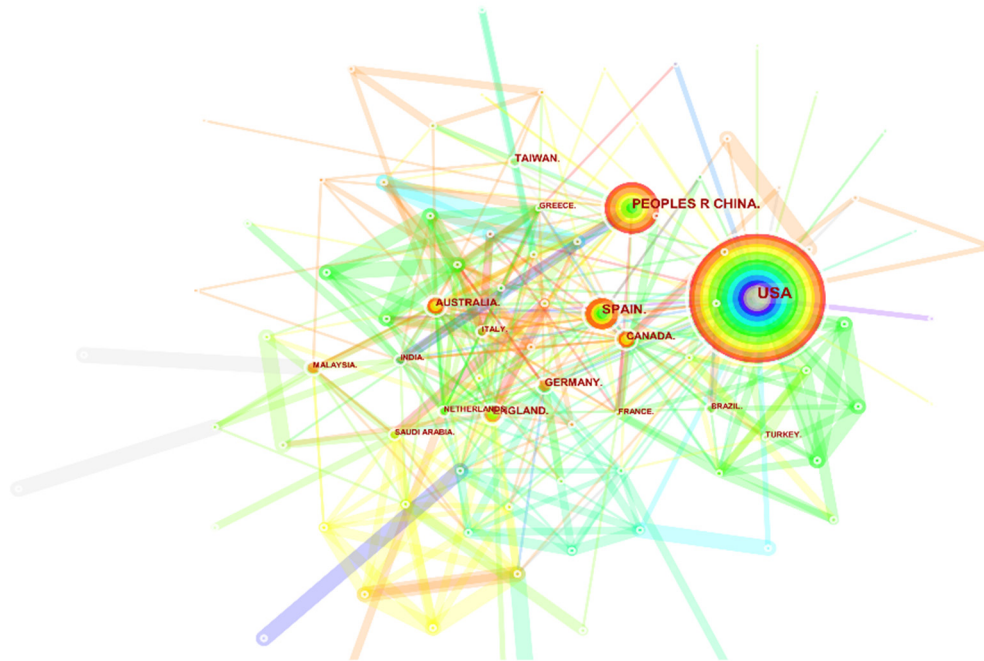


Figure 2: National cooperation network map, 2008-2022

#### 4. RESEARCH HOTSPOTS AND TRENDS

##### 4.1. Research hotspots for MOOC (keyword co-occurrence)

Keywords are the summary of the author's mind. Therefore, keywords with high frequency are often used to identify hot topics. The study was carried out in

CiteSpace software with parameters set, and results were obtained by clicking on "Keyword" in "Node Types." The results obtained 259 nodes (keywords), 355 connections (co-occurrence relations), with a density equal to 0.0106. After essential beautification, precise results of visualization are obtained in Figure 3.

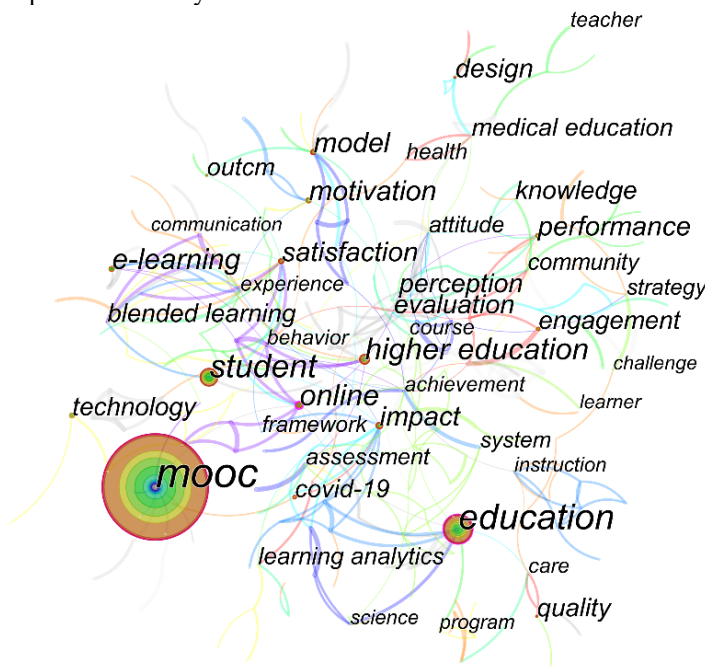


Figure 3: Keyword co-occurrence map

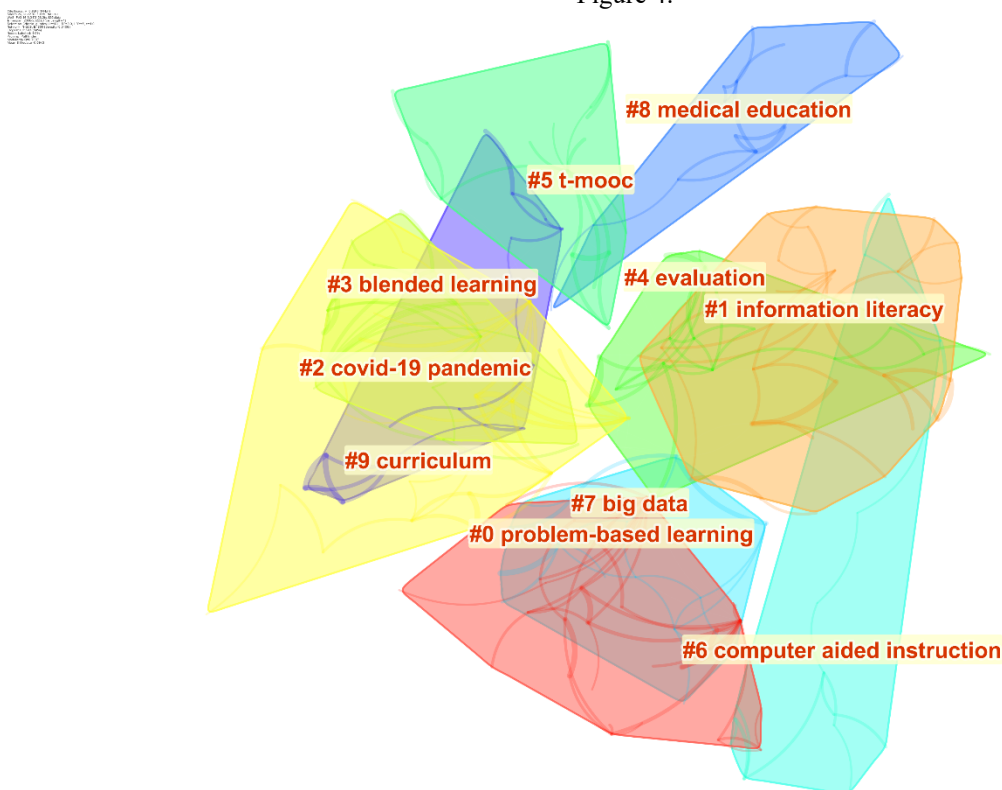
We observe Figure 3 in terms of size and color. The size of the nodes represents the difference in keyword frequency. The larger the node, the higher the keyword frequency. The warm and cold colors of the nodes represent the sequence of the first appearance of the nodes. Among them, the closer the center color of the node is too dark, the earlier the keyword appears for the first time, the lighter the center color of the node is, the later the keyword appears for the first time.

Looking at Figure 3, we can see some clutters in the keywords. This indicates that the research direction of MOOC is not consistent, which again verifies the necessity of choosing "evaluation" and "evaluate" as the entry point of our study. The "MOOC" has the most

significant node and the darkest center color. The following most popular terms are education, student, higher education, online, e-learning, satisfaction, model, performance, design, impact, covid-19, quality, blended learning, motivation, technology.

#### 4.2. High-frequency keyword clusters

The keyword co-occurrence map was too cluttered. The study proceeded with the cluster mapping of keywords for analysis to find the research hotspots. The process is to click "Extract Cluster Keywords" and "Use Keywords" in CiteSpace. The clustering method was selected as LLR, and the clustering map is shown in Figure 4.



**Figure 4:** Co-citation cluster

A total of 10 red labels appear in Figure 4, representing that there are 10 clusters. Three clusters, #0 problem-based learning, #1 information literacy, and #2 covid-19 pandemic, include more keywords and have more lines between internal nodes, indicating a super high degree of co-occurrence between keywords in these fields. The other clusters are #4 evaluation, #5 t-mooc, #6 computer aided instruction, #7 big data, #8 medical education, and #9 curriculum. Some cross-cluster linkages in Figure 4 indicate a high co-citation between these directions.

To further investigate the keywords within each cluster, the Layout-Timeline View function of Citespace was used to observe the temporal pulse variation of the keyword distribution. Figure 5 still has 10 clusters, consistent with Figure 4. The timeline view allows a detailed look at when different topics became hot topics for research. Many keywords appear in Figure 4, but the knowledge graph does not need to be adjusted because there is no severe overlap between each word. This again proves that research topics are broad and change relatively quickly with social context, learner needs, and course providers.

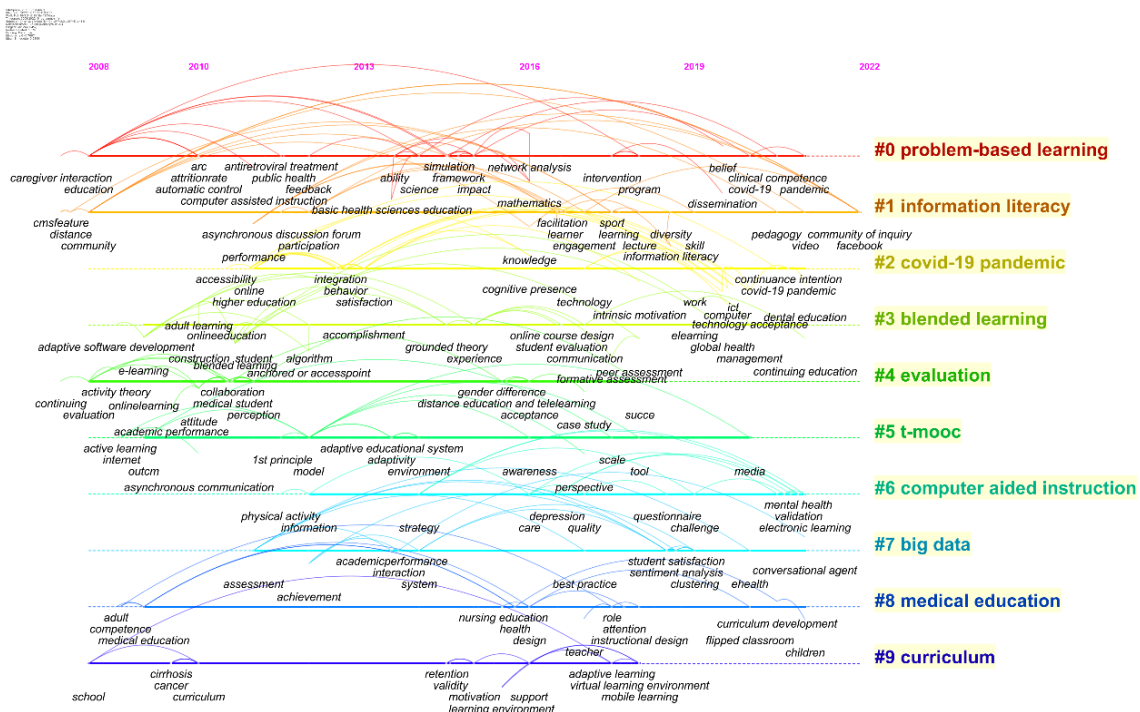


Figure 5: Cluster with a timeline view of keywords

4.3. Research trends for MOOC (keyword emergence)

CiteSpace can find the keywords with a high frequency of occurrence for each period, thus finding the research trends. Figure 6 shows the top 15 keywords. The literature searched is from 2008, when the concept of MOOC emerged, but "e-learning" did not appear until 2011 and continued until 2017. However, the keyword "electronic learning" also appears as an alternative term from 2021 and continues until now. The terms "COVID-

19" and "pandemic" appear from 2020 and continue consistently, which is in line with reality and verifies that CiteSpace was used correctly in this study.

An interesting finding is that from 2021, MOOCs are moving towards specialized areas, especially in combination with more research in the medical field. The principal place of the teacher can start to come to the fore. Completing the instructional design according to the needs and using electronic technology for the whole course will be future research trends.

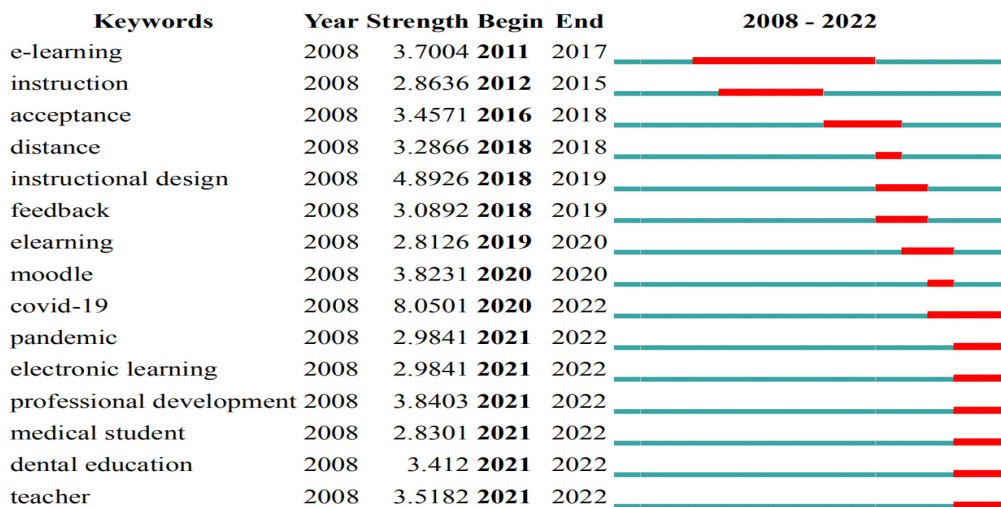


Figure 6: Top 15 Keywords with the Strongest Citation Bursts

## 5. CONCLUSION

This study analyzes MOOC research's hot spots and trends. The literature was sourced from the WOS Core Collection. Given many topics, the theme of "evaluation" was included in the search to make the analysis more accurate. The study used CiteSpace for bibliometrics and generated a clear knowledge map for analysis.

MOOCs have achieved educational equity through technology and have gained more and more participants after COVID-19. However, new requirements have been placed on MOOCs. MOOC providers need to pay attention to the quality of the courses and accurate evaluation and instructional design and implementation based on the professional needs of the courses received. In future research, we need to continue to focus on the changing themes of MOOC research and add more research methods, such as content analysis, to the bibliometric approach to research. Because different disciplines use different instructional designs and teaching methods, we can also study MOOCs in future research based on their disciplinary characteristics.

## ACKNOWLEDGEMENTS

This research was funded by the 2021 Special Project of Philosophy and Social Science Research of Inner Mongolia Education Department, grant number ZSZX21142, 2021 Inner Mongolia Education Science Planning Project, grant number 390, The 8th Teaching and Research Project of Hulunbuir University, grant number JYYB2021045.

## REFERENCES

- [1] Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, 57(3), 359-377.
- [2] Margaryan, A.; Bianco, M.; Littlejohn, A. Instructional quality of Massive Open Online Courses (MOOCs) quality of Massive Open Online Courses (MOOCs). *Comput Educ* 2015, 80, 77-83.
- [3] Ronkowitz, K. (2013). Have You Been MOOCed?. *New Jersey English Journal*, 2(1), 10.
- [4] Zhang, Z.; Chun, D. Research Status and Trend of Employee Innovative Behavior Visualization Analysis Based on CiteSpace and VOSviewer. *Journal of Chongqing Jiaotong University(Social Sciences Edition)* 2021, 21, 93-102.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

