

Bibliometric Knowledge Mapping of e-commerce platform operation on data mining

Min Ye¹, Hongxia LI^{2,1 *}

¹Rattanakosin International College of Creative Entrepreneurship, Rajamangala
University of Technology Rattanakosin, Phutthamonthon, 73170, Thailand

²School of management Science and Engineering, Chongqing Technology and
Business University, 400067, China

Abstract: The e-commerce platform in the digital economy era has evolved into a data platform ecosystem built around data resources and data mining technology systems. The most typical applications of big data are also concentrated in the field of e-commerce. E-commerce companies should first grasp the interactive relationship among the three major factors of data, technology and innovation, e-commerce platform operation is a multidisciplinary research field. It is not easy for researchers to obtain a panoramic view of the knowledge structure in this field. Knowledge graph is a kind of graph that shows the development process and structure relationship of knowledge with the field of knowledge as the object. It is not only a visual knowledge mapping, but also a serialized knowledge pedigree, which provides researchers with a quantitative research method for the development trend of statistics and academic status. The purpose of this research is to help researchers understand the key knowledge, evolutionary trends and research frontiers of current research. This study uses Citespace bibliometric analysis to analyze the data of the Science Net database and finds that: 1) The development of the research field has gone through three stages, and some representative key scholars and key documents have been recognized; 2) the common knowledge mapping of literature The co-occurrence of citations and keywords shows research hotspots; 3) The results of burst detection and central node analysis reveal research frontiers and development trends. Today, the visualization of big data brings different challenges. The abstraction between the world and today's data visualization occurs when the data is captured. Every user sees his own visualization data generated by standardized calculations. At the same time, there are still many controversies in the theoretical model, structure and

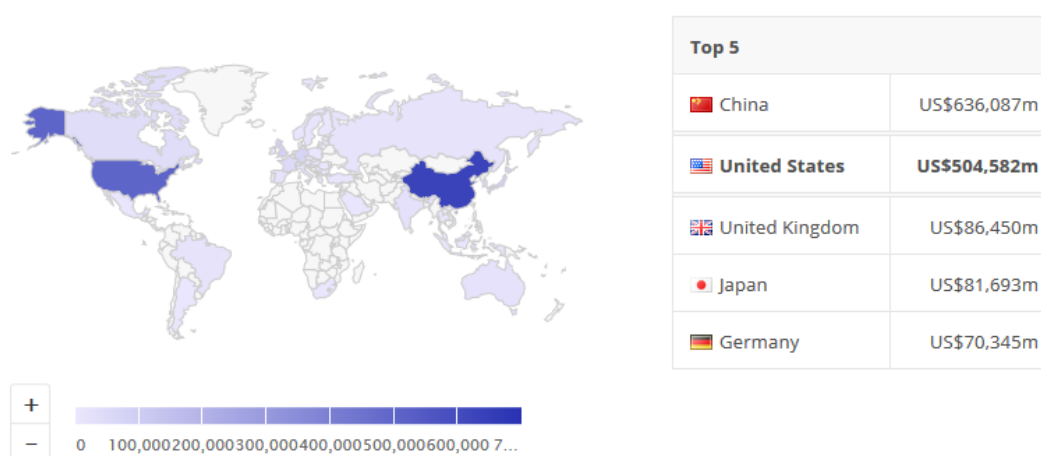
Corresponding author: Hongxia LI, No.19, Xuefu Ave, Nanan District, Chongqing, P.R.China 400067
+8602362769347, lihongxia@ctbu.edu.cn

structural dimensions. This is the direction that future researchers need to further study.

Keywords: E-commerce, big data, bibliometric analysis, knowledge mapping

1.Introduction

E-commerce has entered the "big data era". According to Statista , revenue from the U.S. e-commerce market reached \$504.582 billion in 2018 , up 12.9% year-on-year . China is the world's largest e-commerce market, with revenues of \$636.087 billion in 2018. The United States was close behind (\$504.582 billion). The UK, Japan and Germany are also among the world's top five e-commerce markets, but their revenue levels are far lower than those in China and the US. The compound annual growth rate for 2018-2023 will be around 7.8%, and the market volume will reach US\$735.358 billion by 2023. The number of e-commerce users worldwide in 2018 was 258.5 million, up 1.2% Year-on-Year, with a penetration rate of 78.8% and expected to reach 80.4% by 2023.



E-commerce platform is a kind of virtual "market", with its wireless connectivity and powerful information processing ability, formed a new transaction structure, greatly enhance the value creation and value capture ability of all participants. With the e-commerce platform, buyers and sellers can complete transactions on products or services regardless of space, time, etc., increasing the frequency of transactions and reducing transaction costs. In the e-commerce platform, through the click data flow analysis, we can know that these data to improve the design of the site, improve the affinity of the site, has important significance. Use data warehouse and data mining techniques to process data such as click-through data streams and users' personal information. There is a growing preference for the use of online data, which are properly extracted and processed and can lead to accurate and timely indicators of various economic themes. It's not easy for researchers in the field of e-commerce to get a panoramic view of the structure, evolution and key nodes of knowledge. Before the method of Bibliometric, researchers who wanted quick access to a panoramic view of a discipline or field of study relied primarily on peer-reviewed articles or anthology. One obvious limitation of this approach is that it is subject to peer knowledge and subjective judgment, and it does not fully reveal key literature and emerging research hotspots in many cases and it is easily controversial (Peng, Zhu, Wu, 2020; Latini, 2019). Prior to the advent of Citespace, researchers relied primarily on peer-reviewed

articles or collections of papers to gain a panoramic view of the subject or field of study. Because of the lack of objectivity due to the research peer's knowledge vision and subjective judgment, this method can not reveal the main literature and emerging research hotspots in this field comprehensively, which is easy to cause controversy. The emergence of the Citespace analysis tool provides researchers with another possibility, that is, from simple subjective judgment to subjective judgment combined with objective measurement, this study is guided by this idea.

Citespace is a quantitative analysis method based on scientific and technical literature, which describes, evaluates and predicts the situation and trends in various research fields by mathematical and statistical methods. This paper uses two common literature analysis software: Citespace and VOSviewer, to study the key path of knowledge evolution in the field of e-commerce, to reveal important knowledge inflection points in the field, to analyze the potential dynamic mechanisms of evolution and to explore the frontiers of development by drawing a series of visual maps.

2.Data Sources and Research Methodology

2.1 Data Sources

In this study, we used the Web of Science (WOS) as a data source. WOS integrates the Scientific Citation Index (SCI), the Social Science Citation Index (SSCI), Arts and Humanities Citation Index (A and HCI), and the Conference Institute Citation Index - Social Sciences and Humanities (CPCI-SSH). It is currently the world's deepest and most complete database, providing citation index data. Unlike EBSCO, Springer, Willie Blackwell, and other large databases, the WOS database covers all source journals of SSCI, and its sub-databases enjoy high authority in academia. The retrieval method of this data collection was "Topic= E-Commerce platform operations AND data mining" and to filter out duplicate literature. After searching the four databases of WOS (SSCI, A&HCI, CPCI-SSH, SCI), we obtained 2586 pieces of relevant literature index data. Then we purified the obtained data by three standards of the general article, proceedings paper and review, and obtained 2518 bibliographic data. Data collection time is from 2010 to 2020 (the data collection time for the database is 2010). The retrieval date is November 19, 2020. 1 In this paper, Citespace is used to analyze annual statistical analysis, authors, institutions, countries, literature keywords, synonym keyword combination processing, statistical high frequency keywords, generating word matrix and co-occurrence matrix. Using the visual software CiteSpace to document the key node literature, the knowledge map is drawn, and the research hotspot and research frontier research evolution is analyzed by clustering algorithm and burst word detection function. When this study used Citespace to generate a knowledge map, the parameter threshold was set according to the following steps: 1) the reference data was imported into the Citespace software: 2) Set parameter values in the software to "TimeSlicing=2010-2020", "YearsPerSlice=1", "TermSource" by default, "TermType=BurstTerms," "Pruning=Minimum Spanning Tree," and "NodeType" to be flexibly selected as required by the research. Among them, 2020 in

the data time range refers to the data limited to the data collection time range, that is, the data up to 18th November, 2020.

2.2 Research Methodology

VOSviewer is a JAVA-based knowledge drawing software developed by the Centre for Scientific and Technological Research at Leiden University in the Netherlands. The structure, evolution, cooperation and other relations in the field of literature knowledge are introduced. This paper uses VOSviewer to statistically analyze the age and keywords of the literature related to the operation of the e-commerce platform and combined the synonym keywords to calculate high-frequency keywords. CiteSpace Knowledge Visualization Software is a multidimensional, tense and dynamic knowledge mapping tool developed by Professor Chen Chaomei (2006) of Dressel University in the United States. The software can convert abstract data into graphical expressions (Chen and Liu, 2005), by analyzing and understanding graphics, so that users can visually see the relevant information in the corresponding research field, grasp the research field of structural relations, evolutionary laws and other knowledge characteristics. CiteSpace version 5.7 is used to analyze the literature selected in this article. Firstly, the key node literature is collated by co-introduction analysis, and the research hotspots and frontiers of e-commerce platform operation are analyzed by clustering and burst detection, and secondly, knowledge maps are drawn to analyze the research evolution trend in the field of e-commerce platform operation. Finally, the coupling relationship of literature is used to analyze the internal relationship between disciplines in literature.

3. Data Analysis

3.1 Literature Analysis

3.1.1 Annual Volume Analysis

To some extent, the overall annual distribution of the literature can reflect the research status of the field at a particular stage. In addition, the total number of papers published each year is also an important indicator of scientific research output. Using CiteSpace to extract annual data from WOS data, analyze the overall research trends in the field of e-commerce, and obtain a time trend graph of the number of documents (Figure 1).

According to the trend of the number of publications, the development of e-commerce research is divided into three stages: 1) Initial stage; 2) Slow stage; 3) Development stage. Before 2000, it was the initial research phase, with or without several papers published each year. From 2000 to 2014 is a slow development stage. The number of articles published each year began to grow, sometimes shrinking, from 110 in 2000 to 71 in 2014. 2015 is a stage of rapid development, and the number of articles published each year has shown a clear and rapid growth trend, from 137 in 2015 to 298 in 2020 (Table 1).

Figure 1:

Distribution Map of Annual Total Literature in the E-Commerce Field

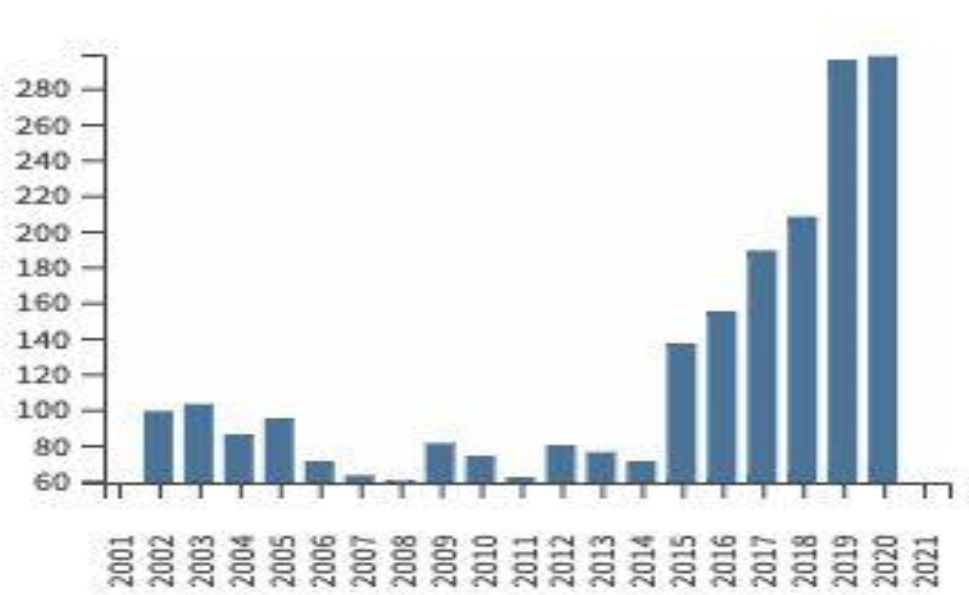


Table 1:

The Annual Number of Publications Related to E-Commerce

Years	The number of publications	Years	The number of publications
2000	110	2011	62
2001	104	2012	80
2002	99	2013	76
2003	103	2014	71
2004	86	2015	137
2005	95	2016	155
2006	71	2017	189
2007	63	2018	208
2008	60	2019	296
2009	81	2020	298
2010	74		

3.1.2 Institutional and Country Analysis

We extracted institutional data from WOS, and the top ranked item by citation counts is (Hong Kong Polytech Univ) in Cluster #1.5, with citation counts of 4. The

second one is (Shandong Univ Finance & Econ) in Cluster #0.5, with citation counts of 3. The third is (Jinan Univ) in Cluster #0.5, with citation counts of 2. The 4th is (Tongji Univ) in Cluster #-0.5, with citation counts of 2. The 5th is (Univ N Carolina) in Cluster #-0.5, with citation counts of 2. The 6th is (Fudan Univ) in Cluster #1.5, with citation counts of 2. The 7th is (RMIT Univ) in Cluster #1.5, with citation counts of 2. The 8th is (Chinese Culture Univ) in Cluster #6.5, with citation counts of 2. The 9th is (Harbin Inst Technol) in Cluster #0.5, with citation counts of 2. The 10th is (Univ Memphis) in Cluster #-0.5, with citation counts of 1. (Figure 2) (Table 2).

Figure 2:

Top 10 Institutions in the Field of E-Commerce

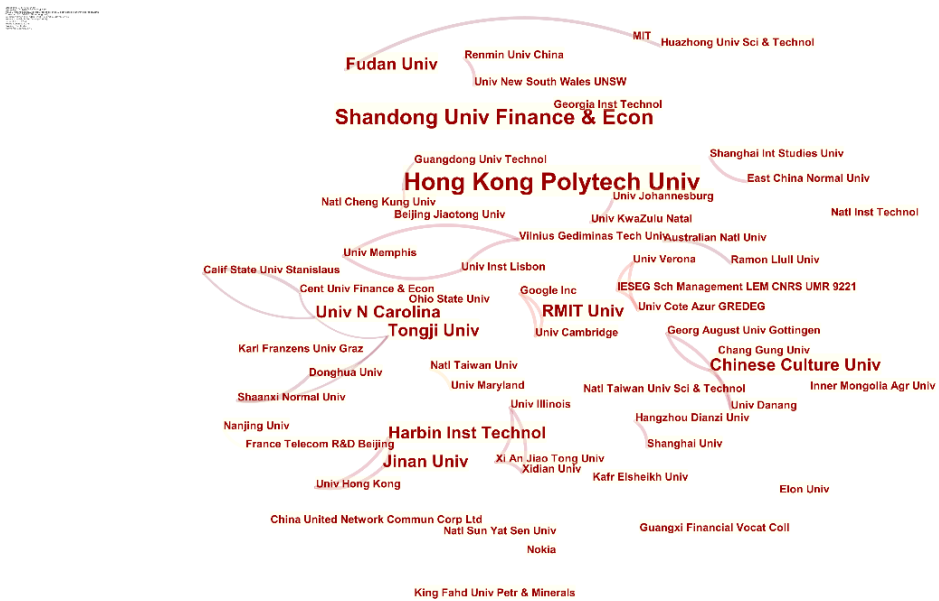


Table 2:

Top 10 Institutions in E-Commerce Rankings

citation counts	References	Cluster #
4	Hong Kong Polytech Univ	1.5
3	Shandong Univ Finance and Econ	0.5
2	Jinan Univ	0.5
2	Tongji Univ	-0.5
2	Univ N Carolina	-0.5
2	Fudan Univ	1.5
2	RMIT Univ	1.5
2	Chinese Culture Univ	6.5
2	Harbin Inst Technol	0.5

Corresponding author: Hongxia LI, No.19, Xuefu Ave, Nanan District, Chongqing, P.R.China 400067
+8602362769347, lihongxia@ctbu.edu.cn

1	Univ Memphis	-0.5
---	--------------	------

3.1.3 Key Scholars Analysis

Key scholars are leading the way in the field of research. The frequency of the author's reference can reveal the author's attention in the field of research and serve as a measure of the author's influence. Based on previous analysis, this has been a rapidly evolving stage since 2015, and analysis of high-volume authors can help researchers find the latest representative authors in the field of e-commerce. The top ranked item by centrality is (WANG Z) in Cluster #0.5, with centrality of 4. The second one is (CHONG AYL) in Cluster #-0.5, with centrality of 3. The third is (SHEN B) in Cluster #-0.5, with centrality of 3. The 4th is (LU QH) in Cluster #-0.5, with centrality of 3. The 5th is (YU Y) in Cluster #-0.5, with centrality of 3. The 6th is (CHOI TM) in Cluster #-0.5, with centrality of 3. The 7th is (LI L) in Cluster #-0.5, with centrality of 3. The 8th is (LIU Z) in Cluster #-0.5, with centrality of 3. The 9th is (ZHANG J) in Cluster #-0.5, with centrality of 3. The 10th is (CHEN X) in Cluster #-0.5, with centrality of 3.

Figure 3:

Analysis of Key Scholars in the Field of E-Commerce

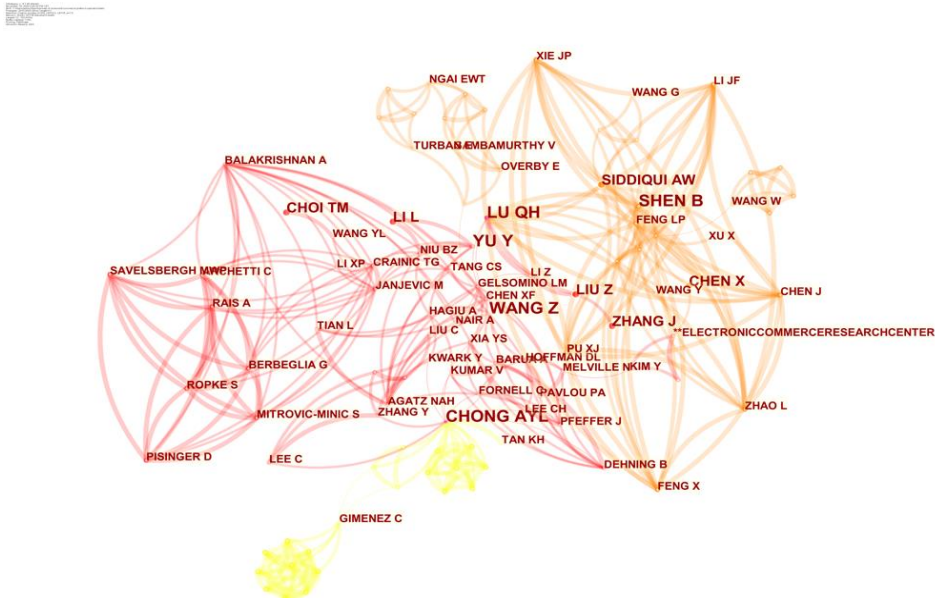


Table 3:

Top 10 Authors Cited in the E-Commerce Field

Centrality	References	Cluster #
4	WANG Z	0.5
3	CHONG AYL	-0.5
3	SHEN B	-0.5
3	LU QH	-0.5
3	YU Y	-0.5
3	CHOI TM	-0.5
3	LI L	-0.5
3	LIU Z	-0.5
3	ZHANG J	-0.5
3	CHEN X	-0.5

3.1.4 High Citation Literature Analysis

The highly cited papers reflect the degree and focus of academic attention to research. In the current research evaluation, citation frequency is an index to measure the importance of the paper. This paper extracts citation data from WOS and obtains the top 10 papers with the highest citation frequency, as shown in Table 4.

Table 4:

Top10 Most Frequent Citations in the Field of E-Commerce

Author	The name of the source publication	The public date	Total number of references
McKnight, DH; Choudhury, V; Kacmar, C	INFORMATION SYSTEMS RESEARCH	SEP 2002	1853
Gefen, D	OMEGA-INTERNATIONAL JOURNAL OF MANAGEMENT SCIENCE	DEC 2000	1196
Srinivasan, SS; Anderson, R; Ponnnavolu, K	JOURNAL OF RETAILING	SPR 2002	907
McKnight, DH; Chervany, NL	INTERNATIONAL JOURNAL OF ELECTRONIC COMMERCE	WIN 2001	771
Dinev, T; Hart, P	INFORMATION SYSTEMS RESEARCH	MAR 2006	684
Gefen, D; Straub, DW	OMEGA-INTERNATIONAL JOURNAL OF MANAGEMENT SCIENCE	DEC 2004	661
Devaraj, S; Fan, M; Kohli, R	INFORMATION SYSTEMS RESEARCH	SEP 2002	570
DeLone, WH; McLean, ER	INTERNATIONAL JOURNAL OF ELECTRONIC COMMERCE	FAL 2004	518
Tsay, AA; Agrawal, N	PRODUCTION and OPERATIONS MANAGEMENT	SPR 2004	461
Ben Schafer, J; Konstan, JA; Riedl, J	DATA MINING and KNOWLEDGE DISCOVERY	JAN-APR 2001	415

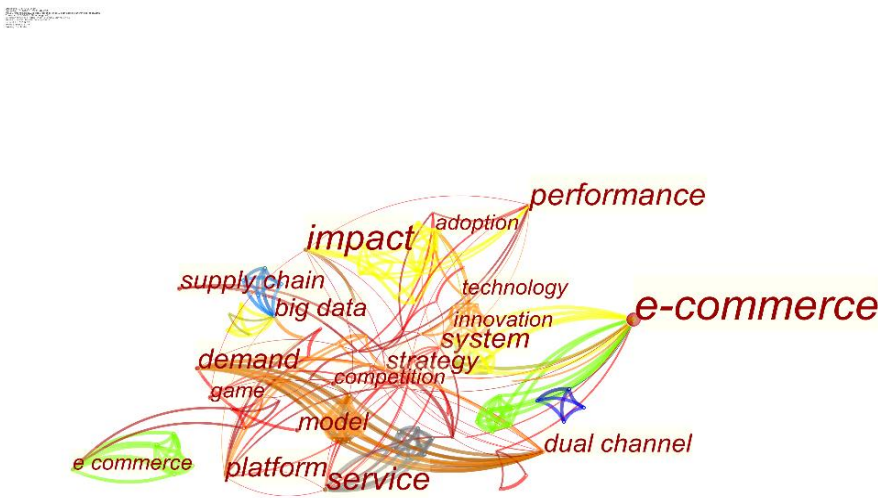
3.2 Research Hotspots Analysis

By analyzing recent hotspots and frontiers, It is helpful for researchers to understand the latest research trends and identify research issues in the near stage. We selected 2010-2020 data as the source data for hot spot and cutting-edge analysis based on the last development stage(2020 data, less than one year, not included in the analysis). In Citespace, "TimeSlicing= 2010-2020"," Node type= Reference ", " Top=25"are set up to obtain a commonly referenced visual knowledge map of the literature(Figure 4). In the figure, there are 235 nodes, 469 links, density 0.0171, modularity 5974 (Q-value,

which is generally considered an important cluster structure for $Q > 0.3$), Silhouette=0.6035 (S-value, which is generally considered a reasonable cluster of $S > 0.5$). The node in the figure is the year ring. The larger the node, the higher the overall frequency of the common reference. In the ring of the year, blue represents the earlier years, and red represents the last few years. The crimson color at the center of the year ring indicates that the literally cited document is a burst document, and that the thickness of the ring is directly related to the frequency with which the document is referenced jointly during the year. Some rings have purple outer rings, indicating that the commonly quoted literature is fairly central. The connection between common reference references means that two common references often appear in the same article. The thicker the connection, the higher the frequency of co-occurrence. Clustering refers to the use of clustering algorithms to group collections of physical or abstract objects into classes composed of similar objects (Chen, C. and Leydesdorf, 2014). Clustering analysis is based on the similarity of analytical objects, and the (LLR) algorithm is used in this paper. Hot spot clustering analysis is based on the natural advantages of co-citation network clustering based on connection relationships rather than node attributes. It clusters the sampling space of any shape and converges to the best global solution. Among them, Q value and S value are two key measures of the overall structure of the network. Keyword phrases are extracted from the title of a particular cluster as a cluster tag to describe and identify the cluster. From Summary Figure 2, we can see that the research hotspots in this field mainly form eight aspects, namely: the item ranked highest by citation count is e-commerce in "cluster" and the citation count is 813. The second is the model in the citation count "cluster" with a citation count of 282. The third cluster, electronic commerce, has a reference count of 273. The fourth is the big data in the cluster, with a reference count of 246. The fifth is performance in the cluster, with a reference count of 221. The sixth is impact in the cluster, with a reference count of 177. The seventh is technology in the cluster, with a reference count of 127. The eighth in the cluster is system, with a reference count of 125. The ninth is adoption in Cluster, with a reference count of 124. In the 10th innovation in Cluster, the reference count is 120. The key article in each cluster block is the node shown in Figure 4.

Figure 4:

Keyword Clustering Graph of E-Commerce



Keyword co-occurrence is another method to detect research hotspots. It is used CiteSpace to generate a keyword co-occurrence matrix and built a keyword co-occurrence network to identify research hotspots in the field. Extract a list of keywords with a total frequency of TOP 25 to display (Figure 5).

Figure 5:

Keyword Clustering Diagram of Common References of E-commerce

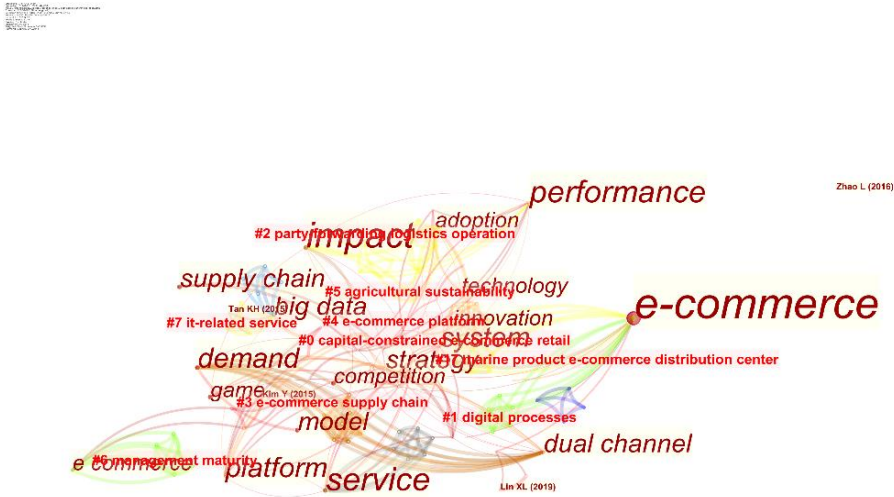


Table 5:

TOP 10 Keywords of Co-occurrence Frequency

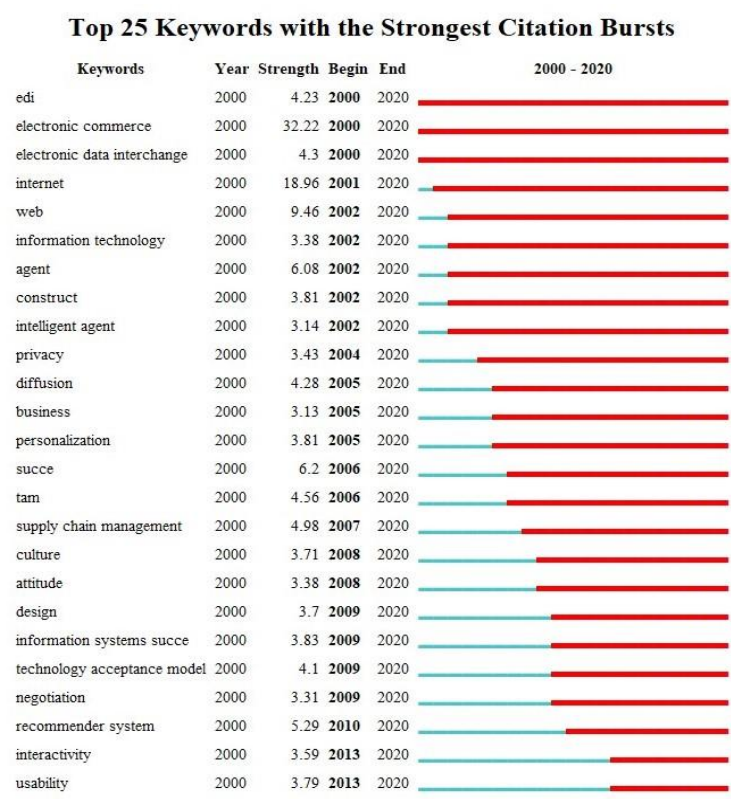
citation counts	References	Cluster #
813	e-commerce	0
282	Model	1
273	electronic commerce	0
246	Big data	3
221	performance	2
177	impact	4
127	Technology	5
125	System	6
124	adoption	7
120	Innovation	7

According to Table 5, we can find that the research objects in the field of e-commerce research are e-commerce, model, electronic commerce, big data, impact, performance, technology, adoption, innovation, system.

To check the literature, we clicked the Burst button to set the minimum duration of burst detection to 1 year. We got the top 25 mutation keywords that revealed different stages and future trends in e-commerce. The keywords in the graph are represented the corresponding keywords, the year represents the first time the record is retrieved, the strength indicates the strength of the mutation, the beginning indicates when the mutation starts, and the end indicates the end time. The start and end times of the year correspond to a small red rectangular block. As shown in Figure 6 below:

Figure 6:

Top 25 Citation Bursts Keywords



3.3 Document Co-citation Analysis

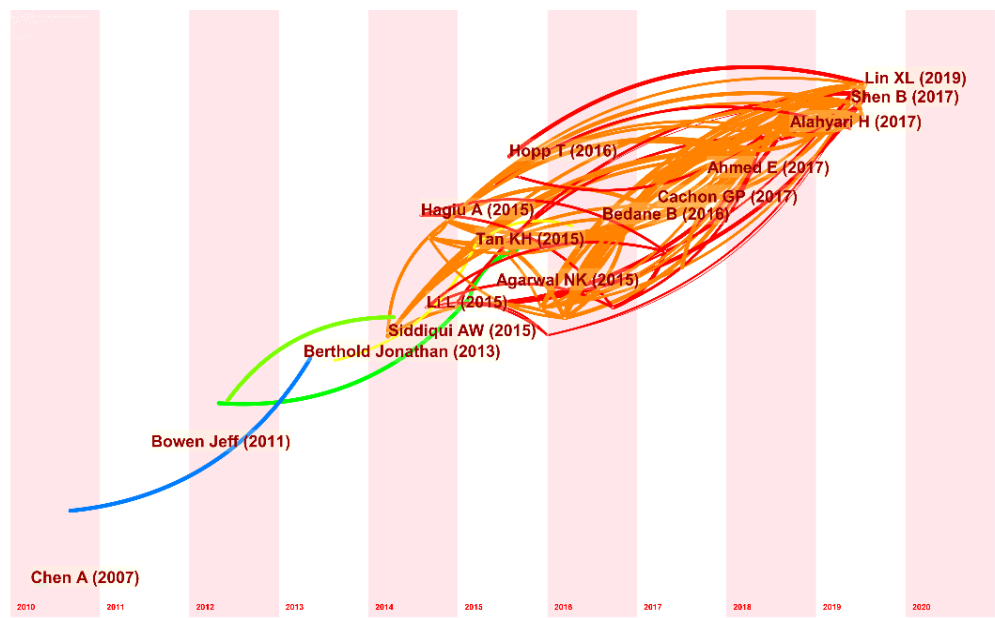
Document Co-citation Analysis (DCA) can be used to identify key literature and research frontiers in the field of research (Chen, C., Ibekwe-SanJuan, F., & Hou, J., 2010). A common citation refers to the phenomenon of two articles quoted in the same article. Generally speaking, the frequency of common citations changes over time, so the literary network based on common citation relationships is a dynamic structure. Common Citation Analysis has dynamic analysis capabilities and is suitable for analyzing the evolution of knowledge in a discipline or field of study (Zeng Zhishen, 2008). Document co-citation analysis (DCA) can be used to identify key literature and research frontiers in a research field. The "cited Reference"" was used as the node in the CiteSpace, set LBY=-1(no limitation of retroactivity) , Timeslice=1, years, and top N=30. The document co-citation map of E-commerce was obtained by using Pathfinder pruning and displaying in a time-zone mode, which is shown in Figure 7. According to Figure 7, it can be seen that the co-citation relationship in the field of e-commerce in 2007-2020 is very close, while the citation contribution rate is high in the period 2015-2019. Among them, two highly cited documents, Chen A and Bowen Jeff, were cited 15 times before 2011; between 2013 and 2015, there were six highly cited documents, mainly Berthold Jonathan, Siddiqui AW, and Li L , Agarwal NK, Tan KH and Hagi A were cited 126 times; in 2016-2019, there were seven highly cited documents, such as

Corresponding author: Hongxia LI, No.19, Xuefu Ave, Nanan District, Chongqing, P.R.China 400067
+8602362769347, lihongxia@ctbu.edu.cn

Bedane B, Hopp T, Cachon GP, Ahmed E, Alahyari H, Shen B and Lin XL, which were cited 178 times. As shown in Figure 7 below:

Figure 7:

Map of Document Co-citation Time Zones

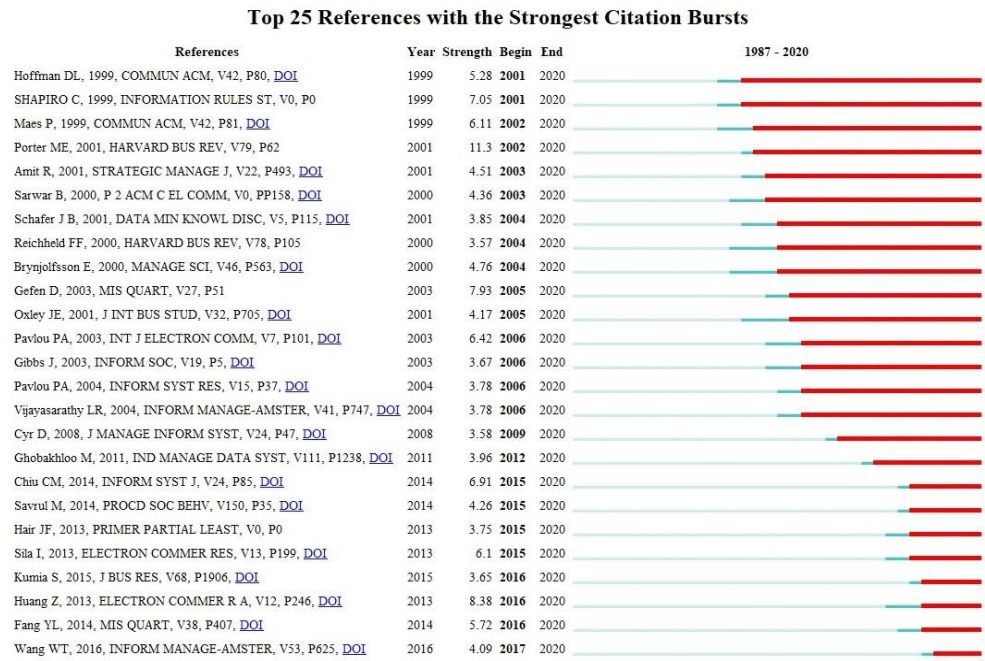


3.4. Research Frontier and Trend Analysis

Research Frontiers refer to "a set of sudden dynamic concepts and potential research issues" (Chen, C. and Leydesdorf,2014). In CiteSpace, Kleinberg's Burst Detection Algorithm was used to identify those cited articles (burst items) that have aroused the academic community's close attention in a specific time, and they are marked with annual red rings inside the nodes. The longer the reference surge of a node lasts, the thicker the red ring of the node. Burst literature contains two dimensions of burst value and burst time. The literature with high emergence value means that it has received particular attention in the corresponding time interval, representing the research frontier of the field to some extent. Burst time interval reflects the dynamic change of a discipline or research field frontier hotspot. To display the latest research frontier information, this paper extracted burst literature from the latest burst time (2020), the authors of key articles representing the frontiers of e-commerce research (Figure 8), as shown in Figure 8 below.

Figure 8:

Top 25 Citation Bursts References



The top ranked item by bursts is (Porter ME) in Cluster #2.5, with bursts of 11.30. The second one is (Gomez-Herrera E) in Cluster #4.5, with bursts of 10.42. The third is (Huang Z) in Cluster #3.5, with bursts of 8.38. The 4th is (Gefen D) in Cluster #3.5, with bursts of 7.93. The 5th is (SHAPIRO C) in Cluster #2.5, with bursts of 7.05. The 6th is (Chiu CM) in Cluster #3.5, with bursts of 6.91. The 7th is (Pavlou PA) in Cluster #3.5, with bursts of 6.42. The 8th is (Maes P) in Cluster #2.5, with bursts of 6.11. The 9th is (Sila I) in Cluster #3.5, with bursts of 6.10. The 10th is (Mangiaracina R) in Cluster #2.5, with bursts of 6.05.

Table 6:

Top 10 Co-cited Literature Authors

bursts	References	Cluster #
11.30	PorterME, HARVARD BUS REV, 79	2.5
10.42	Gomez-HerreraE, INF ECON POLICY, 28	4.5
8.38	HuangZ, ELECTRON COMMER R A, 12	3.5
7.93	GefenD, MIS QUART, 27	3.5
7.05	SHAPIRO C, INFORMATION RULES ST, 0	2.5
6.91	Chiu CM, INFORM SYST J, 24	3.5
6.42	Pavlou PA, INT J ELECTRON COMM, 7	3.5
6.11	Maes P, COMMUN ACM, 42	2.5
6.10	Sila I, ELECTRON COMMER RES, 13	3.5
6.05	Mangiaracina R, INT J PHYS DISTR LOG, 45	2.5

3.5 Disciplinary dual-map Overlay Analysis

Disciplinary dual-map Overlay analysis can use the coupling relationship of literature to analyze the internal relationship between disciplines in literature. Figure 9 is obtained by using the Zscore algorithm to overlay and simplify the literature data of e-commerce, using the journal Double Chart Overlay to see the two areas of common research hot spots, where the thickness of the lines indicates the degree of connectivity between disciplines. Part 1 of the figure consists of the subject relationship of the cited literature, and Part 2 of the area consists of the subject relationship of the cited document, as shown in Figure 9 below.

Figure 9:

Disciplinary Dual-Map Overlay of E-commerce



Figure 9 shows that the research on e-commerce in Part 1 focuses on computer science, economics, management, and business, which provide a developing knowledge base for e-commerce research. Part2 shows that the research literature of e-commerce is quoted as psychology, education, sociology, health science and biometrics. Less distribution in physical education, more often cited in chemical science.

Results and discussions

4.1 Results

In this paper, we used the bibliometrics method to make a statistical analysis of the article's annual publication, research institutions, key authors, keywords, and citations. We used literature citation analysis, keyword clustering algorithm, burst detection algorithm and Z-score algorithm to detect research hotspots, research frontiers and evolutionary trends in the field of e-commerce, providing key knowledge references for relevant researchers.

According to the statistics of the trend of the number of papers, the development of e-commerce research is divided into three stages: 1) initial stage; Before 2000, it was the initial research phase, with or without several papers published each year. From 2000 to 2014 was a slow stage of development. The number of articles published each year began to grow, sometimes shrinking, from 110 in 2000 to 71 in 2014. 2015 was a period of rapid development, with a marked and rapid growth in the number of articles published each year, from 137 in 2015 to 298 in 2020. The analysis cited in the document shows that the common citation relationship in the field of e-commerce was very close in 2007-2020, while the high contribution rate of citations was more concentrated in 2015-2019. Among them, two highly cited documents, Chen A and Bowen Jeff, were cited 15 times before 2011; And Hagi A was cited 126 times; in 2016-2019, seven highly cited documents, such as Bedane B, Hopp T, Cachon GP, Ahmed E, Alahy Hari, Shen B and Lin XL, were cited 178 times. Based on common

Corresponding author: Hongxia LI, No.19, Xuefu Ave, Nanan District, Chongqing, P.R.China 400067
+8602362769347, lihongxia@ctbu.edu.cn

citation statistics and knowledge maps, researchers can quickly acquire important knowledge in the field.

From publishing institutions and national analysis, research institutions in the field of e-commerce are universities, mostly from China. China is the core country in the field of research, accounting for 62.3%, far more than any other country, which is related to China as the world's largest e-commerce market.

WANG Z, CHONG AYL, and SHEN B are highly cited authors in this field, based on the authors' statistical results and citation data. Bedane B, Hopp T, Cachon GP, Ahmed E, Alahyari H, Shen B and Lin XL are highly productive writers in the latest phase. According to citation data and knowledge maps, scholars can access key articles and quickly master the basics of the field.

Research hotspots and cutting-edge test results show that the highest-ranked projects in the field of e-commerce research are e-commerce, model, electronic commerce, big data, impact, performance, technology, adoption, innovation, system. The results show that EDI, electronic commerce, electronic data interchange, information technology, intelligent agent are the latest research frontiers. The key nodes selected by the center show the evolution path of research development, verify the previous analysis results, further reveal the research trends, provide researchers with a clear perspective, and provide researchers with a panoramic view of e-commerce.

A literature measurement method based on a knowledge map. In this paper, a visual analysis method based on knowledge map is used. The Knowledge Map in Literature Bibliometric is a chart that shows the relationship between the development process and the knowledge structure by targeting the domain of knowledge (Chen, C. and Leydesdorf, 2014). It has the dual nature and characteristics of "graph" and "spectrum": it is not only a visual knowledge map, but also a serialized knowledge family tree, showing many complex relationships between knowledge units or groups of knowledge, such as networks, structures, interactions, intersections, evolution or derivation. In the internet age, scientific communication and technological innovation have entered a faster and more efficient pace. In this context, the visualization method can be used to display econometric analysis results, so as to show the research trends and context more clearly. Literature measurement based on knowledge map is a new scientific measurement method. The traditional method of literature is to carry out a simple statistical analysis of literature, and this new method is to dig deep into the literature. It uses deep learning methods to identify entities in the literature and their relationships and displays them as knowledge maps. This new approach opens up more possibilities for knowledge research.

4.2 Discussion

The following is a discussion of research topics and the views of leading scholars in the field of e-commerce. The keyword clustering diagrams in Figures 4 and 5 show that the research topics of e-commerce are mainly concerned with the following. Factors that affect the operation of the e-commerce platform. By analyzing the literature, it can be concluded that many factors affect the operation and transaction of e-

commerce. The mode of E-commerce platform operation, business performance, technological innovation and big data are the main factors.

Data mining Practice and impact in the field of e-commerce academia. One way for e-commerce (Kohavi and Provost, 2001) to bridge the gap from academia to business data mining is to build a vertical bridge and complete the data mining chain from collection to cleanup, mining, action, and validation. The e-commerce and data mining architecture built (Ansari et al., 2001) provides academia with the unique ability to collect more data than is normally available for data mining projects. From the very beginning, important consideration was given to data transformation and analysis needs. This contrasts with one of the challenges of e-commerce business intelligence in the case of post-analysis analysis. In these cases, there is often a gap between the potential value of the analysis and the actual realized value, either because of the limited data collected or because the data must be converted to be efficiently mined (Kohavi, Rothleder, and Simoudis, 2002). Analyzing data is a complex task (Berry and Linov, 1997, 2000) because it requires experience and expertise. Unlike traditional e-commerce, the goal of data mining is not only business-oriented, but also artificial intelligence, big data-oriented. E-commerce has entered the "big data era".

The theoretical model basis for e-commerce platform operation. On the basis of Dynamic Capabilities Theory, Dynamic capability of technological innovation Theory, Technology Acceptance model and RFM model, the researchers gradually perfected the theoretical framework of e-commerce platform operation. Teece and Pisano put forward the concept of dynamic capability (DC) from a resource-based perspective. In the long-term competition, to maintain the company's competitive advantage, it is necessary to continuously develop dynamic capabilities. Through these processes, companies continuously consolidate, reconfigure, update, and re-create resources and capabilities to respond to changing environments to achieve and maintain competitive advantage. The citation frequency of this document is the highest in the statistical backtracking cycle and provides reference for many researchers. Scholar Zhang Lin and others (2018) can be understood as a combination of multiple elements, is a dynamic, need enterprises to accumulate experience and knowledge to form a combination. It will enhance the technology of enterprises, and thus enhance the competitiveness of enterprises. The combination of factors were included technological innovation input ability, technological innovation output ability, technological innovation transformation ability, technological innovation realization mode and technology innovation management ability. Build a knowledge structure diagram of the dynamic ability of technological innovation and verify its rationality. The Technical Acceptance Model (TAM) (Davis, 1989) is one of the core streams of intent to invest in user systems and is used by many scholars. TAM assumes higher system-aware ease of use (PEOU) and perceived usefulness (PU), and therefore has a greater intent to use the system. The technical acceptance model (TAM) is a widely accepted theory in the user's technical acceptance. The RFM (Frequency, Frequency, and Currency) model is a behavior-based model that analyzes the behavior of online customers, such as the customer's browsing history, the customer's purchase history, the customer's transaction history, and so on,

and then forecasts based on database management. Behavior in the database (Hughes, 1996; Yeh et al., 2009). In addition, recency represents the length of the time period since the last purchase, while frequency represents the number of purchases over a specified period of time, and currency indicates that a great deal of research has been done on the analysis and forecasting of customer buying behavior, resulting in a variety of models with high predictive effectiveness, particularly in non-contractual online business environments (Gupta et al., 2006; Van den Boer and Barkinx, 2005). Together, these theories build an impact model for e-commerce platform operations.

Conclusions and limitations

Combining research hotspot and cutting-edge detection with keyword burst detection results and important literature analysis, it can be concluded that innovation, information technology, and big data are important frontier trends in E-commerce. In addition, the social impact of e-commerce participation in different disciplines is also different. Researchers can conduct quantitative comparative studies in the future to provide references for the overall development of e-commerce. With the help of different research methods, research perspectives and research paths, researchers further expand the interdisciplinary research field and enrich the research in this field.

The globalization of society has brought many opportunities, but it has also brought a series of problems. With the rapid development of globalization, these problems have become common problems facing all countries in the world. The field of e-commerce research has become a hot research area for scholars at present. The existing research has made fruitful research results and produced some representative scholars and theoretical articles. However, the results of development analysis show that the scope of research objects needs to be expanded, and there are many disputes in the theoretical model, structure and structural dimension of e-commerce research. There was still a lot of room for development and further research was needed. Future researchers need to enrich and perfect e-commerce theory and research results based on collaboration, integration and expansion, combining the social characteristics of globalization and information technology. E-commerce platform operation and data mining research is a multidisciplinary cross-cutting field, involving behavioral computer science, psychology, education, economics and social science, etc. They enrich the research field by using the research methods, viewpoints and paths of different disciplines, so that the disciplines, methods and perspectives are diversified.

Bibliometrics is a scientific and effective method of library and information science research, which is used for the statistics, description and prediction of academic status quo and development trend. However, the quantity and quality of scientific literature is undoubtedly a measure of technical level. The research in this paper is based on the WOS database, and the research results are relatively objective and fair. However, due to database limitations, the data is not comprehensive enough, and some databases do not support directory downloads. Therefore, there may be some deviations in the results of the study. Future research can consider using data mining technology to expand the scope of source data collection and improve the quantity and quality of

articles.

References

- [1] Ching-Hsue Cheng, You-Shyang Chen. Classifying the segmentation of customer value via RFM model and RS theory. *Expert Systems with Applications* 36 (2009) 4176–4184.
- [2] Chunlai YAN , Hongxia LI. Knowledge Mapping of Virtual Academic Communities: A bibliometric study using visual analysis. *Advances in Science, Technology and Engineering Systems Journal* Vol. 5, No. 2, XX-YY (2020).
- [3] DongHan. The Market Value of Who We Are: The Flow of Personal Data and Its Regulation in China. *Media and Communication*, 2017, Volume 5, Issue 2, Pages 21–30.
- [4] David J. Teece. Dynamic capabilities and entrepreneurial management in large organizations: Toward a theory of the (entrepreneurial) firm. *European Economic Review* 86 (2016) 202–216.
- [5] Gina Colarelli O'Connor. Major Innovation as a Dynamic Capability: A Systems Approach. *J PROD INNOV MANAG* 2008;25:313–330 ©2008 Product Development & Management Association.
- [6] Ivy Bo Peng, Stefano Markidis, Erwin Laure, Daniel Holmes, Mark Bull. A Data Streaming Model in MPI. *National Infrastructure for Computing (SNIC) at PDC*, November 15-20, 2015.
- [7] Isaac Kofi Mensah, Jianing Mi, Feng Cheng. Factors Influencing the Continuance Usage of Online Mobile Payment Apps: A Case Study of WECHAT Users in China. *World Academy of Science, Engineering and Technology International Journal of Humanities and Social Sciences* Vol:12, No:1, 2018.
- [8] Juhani Ukko, Mina Nasiri, Minna Saunila, Tero Rantala. Sustainability strategy as a moderator in the relationship between digital business strategy and financial performance. *Journal of Cleaner Production* 236 (2019) 11762.
- [9] Jelena Šaković Jovanović, Radoje Vujadinović, Elizabeta Mitreva, Cristiano Fragassa and Aleksandar Vujović. The Relationship between E-Commerce and Firm Performance: The Mediating Role of Internet Sales Channels. *Sustainability* 2020, 12, 6993; doi:10.3390/su12176993.
- [10] K.-M. Gremme ,V. Wohlgemuth, Klara-Marie Gremme, Veit Wohlgemuth. Dynamic capabilities: a systematic literature review of theory and practice. *European Journal of Management Issues*. Special Issue: Dynamic Capabilities of Firms Volume 25 (1), 2017, pp. 30-35.
- [11] Martin D. Mileros, Nicolette Lakemond and Robert Forchheimer. Towards a Taxonomy of E-commerce: Characterizing Content Creator-Based Business Models. *Technology Innovation Management Review*, October 2019 (Volume9, Issue 10).
- [12] Mahesh Kumar Singh, Om Prakash Rishi, Sumit Wadhwa. Knowledge Generation through Web Mining Techniques for e-Business Recommendation System.

-
- Journal of Web Development and Web Designing. MAT Journals 2018. Volume 3 Issue 3.
- [13] Mihaela I. MUNTEAN , Diana TÂRNĂVEANU , Alina Raluca ION. E-Commerce Performance. Shopping Cart Key Performance Indicators. Informatica Economică vol. 20, no. 1/2016.
- [14] M. Hajli. An Integrated Model for E-commerce Adoption at the Customer Level with the Impact of Social Commerce. International Journal of Information Science and Management, IJISM, Special Issue (ECDC 2012), 9-10 May 2012.
- [15] Neil A. Morgan . Marketing and business performance. Academy of Marketing Science , 28 July 2011.
- [16] Nieves Arranz,Marta Arroyabe,Jun Li,Juan Carlos Fernandez de Arroyabe. Innovation as a driver of eco-innovation in the firm: An approach from the dynamic capabilities theory. Business Strategy and the Environment, 2020;29:1494–1503.
- [17] Roman Zamecnik ,Rastislav Rajnoha. Strategic business performance management on the base of controlling and managerial information support. Procedia Economics and Finance 26 (2015) 769 – 776.
- [18] RON KOHAVI , LLEW MASON , RAJESH PAREKH, ZIJIAN ZHENG .Lessons and Challenges from Mining Retail E-Commerce Data. Machine Learning, 57, 83–113, 2004 Kluwer Academic.
- [19] Saeed Fayyaz. A Review on Measuring Digital Trade & E-Commerce as New Economic Statistics Products. Statistical Center of Iran , 2019 99 (1), This article is based on contribution at the IAOS (OECD) conference in September 2018 in Paris, France.
- [20] Tian Tian, Tian Baoguang. The Systematical Analysis about the Effect of Knowledge Spillover on Technological Innovation Capability. World Academy of Science, Engineering and Technology International Journal of Humanities and Social Sciences Vol:8, No:3, 2014.
- [21] Víctor Labayen ,Eduardo Magan ,Daniel Morato , Mikel Izal. Online classification of user activities using machine learning on network traffic. Computer Networks 181 (2020) 107557.
- [22] Yuan Shuangyun, Li Hongxia. Knowledge Mapping of Ethnic Identity and Acculturation Based on the Method of Bibliometric. Journal of Ethnic and Cultural Studies Copyright 2020 2020, Vol. 7, No. 3, 17-33.
- [23] Yanyu Wang, Xin Su, Huan Wang, Renyu Zou. Intellectual capital and technological dynamic capability: evidence from Chinese enterprises. Journal of Intellectual Capital©Emerald Publishing Limited 1469-1930 DOI 10.1108/JIC-06-2018-0096.
- [24] Ying Wanga, Fu Jiab, Tobias Schoenherrc, Yu Gongd, Lujie Chene. Cross-border e-commerce firms as supply chain integrators: The management of three flows. Industrial Marketing Management 89 (2020) 72–88.
