7th International Conference on Information Technology and Quantitative Management (ITQM 2019)

The last five years of Big Data Research in Economics, Econometrics and Finance: Identification and conceptual analysis

José Ricardo López-Robles[a],*, Marisela Rodríguez-Salvador[b], Nadia Karina Gamboa-Rosales[c], Selene Ramirez-Rosales[d], Manuel Jesús Cobo[e]

[a] University of the Basque Country, Alameda Urquijo S/N, 48013, Bilbao, Spain
[b] Tecnológico de Monterrey, Av. Eugenio Garza Sada No. 2501 Sur, Col. Tecnológico Monterrey, 64849, Nuevo León, Mexico
[c] CONACYT – Autonomous University of Zacateca, Av. Ramón López Velarde No. 801. Col. Centro, 98000, Zacatecas, Mexico
[d] Autonomous University of Queretaro, Cerro de las Campanas s/n, 76010, Cto Universitario, Querétaro, Mexico
[e] University of Cadiz, Av. de la Universidad 10, 11519, Cádiz, Spain

Abstract

Today, the Big Data term has a multidimensional approach where five main characteristics stand out: volume, velocity, veracity, value and variety. It has changed from being an emerging theme to a growing research area. In this respect, this study analyses the literature on Big Data in the Economics, Econometrics and Finance field. To do that, 1.034 publications from 2015 to 2019 were evaluated using SciMAT as a bibliometric and network analysis software. SciMAT offers a complete approach of the field and evaluates the most cited and productive authors, countries and subject areas related to Big Data. Lastly, a science map is performed to understand the intellectual structure and the main research lines (themes).

1. Introduction

Nowadays, large number of organizations and governments are under pressure to face challenge such extract the maximum value to the data and information. To meet the growing demands of the data age, the main stakeholders are incorporating a new capabilities into their strategies and core process (i.e. Business Intelligence, Competitive Intelligence, Strategic Intelligence, among others), with special consideration to the Big Data [1-4].

However, the Big Data concept has many definitions. Nevertheless, most of them coincide that the Big Data has five main characteristics: velocity, volume, veracity, variety and value. Additionally, the technologies

* Corresponding author. Tel.: +34-650-213-662; fax: +34-946-014-038
E-mail address: ricardolopezrobles@outlook.com.

enablers related to these characteristics are another dimension of the Big Data to be taken into consideration. In view of these approaches, Big Data can be well-defined as a term describing the analysis and storage of complex and large data sets using a series of advanced techniques and tools [5].

To understand the evolution and structure of the Big Data term, it is helpful to has a holistic background. To do it, a performance and bibliometric network analysis are a suitable framework to drive an integrative, comparative and objective analysis of the main research themes related to Big Data in the Economics, Econometrics and Finance field and assess its evolution. Moreover, it makes it possible to include prospective support the identification of research opportunities and gaps and future decisions [6, 7].

In view of the foregoing, the objectives of this research are understand the Big Data research in the Economics, Econometrics and Finance field and to identify the intellectual structure of research area using a bibliometric software SciMAT [8, 9]. In this way, the main indicators related to bibliometric performance (number of documents, received citations, geographic distribution, among others) have been analyzed and measured.

2. Methodology and dataset

To achieve the objectives of this study, a holistic bibliometric approach based on science mapping analysis and performance indicators is employed with the purpose of uncover the status of the Big Data research in the Economics, Econometrics and Finance field.

2.1. Methodology

A complete bibliometric analysis based on performance indicators and science mapping analysis is performed to evaluate the Big Data field in the Economics, Econometrics and Finance field. First, the performance analysis is based on bibliometric indicators which measure the production of different actors, and the international impact achieved. Also, the most cited articles of the field are identified using the h-index and h-classic approach [10].

Second, a conceptual science mapping analysis based on co-words network is performed using SciMAT, which is based on an approach focused on four stages [9]: (1) research themes detection, (2) visualizing themes (research lines) and its relationship as thematic network, (3) discovery of thematic areas and (4) performance analysis.

To that end, the research themes identified have been set out in a strategic diagram. It is a two-dimensions map separated in four different areas in accordance with their relevance, where the research lines (themes) are shown as a sphere and their volume being equivalent to the total of publications linked with the theme: (i) The upper-right quadrant: Q1-Motor themes (themes within this quadrant are relevant to structure and develop the research field). (ii) Upper-left quadrant: Q2- Highly developed and isolated themes (themes within this area are significant but do not have enough importance to be considered more than a very specialized or peripheral activity for the Big Data research field. (iii) Lower-left quadrant: Q3-Emerging or declining themes (the research lines within this quadrant are weakly, but this weakness can be understood as emerging or disappearing themes. (iv) Lower-right quadrant: Q4-Basic and transversal themes (themes included in this quadrant are not enough developed, but could be relevant for the knowledge area.

2.2. Raw data and dataset

In order to carry out the performance and science mapping analysis, the publications associated to Big Data field in the Economics, Econometrics and Finance field during the last five years (2015-2019) have been collected and processed.

The data belonging to the Big Data research in the Economics, Econometrics and Finance field was retrieved from Scopus employing the next advance query: TITLE-ABS-KEY ("big-data" OR "big data" OR "bigdata")

AND (LIMIT-TO (SUBJAREA, "ECON")). AND (LIMIT-TO (PUBYEAR, 2019) OR LIMIT-TO (PUBYEAR, 2018) OR LIMIT-TO (PUBYEAR, 2017) OR LIMIT-TO (PUBYEAR, 2016) OR LIMIT-TO (PUBYEAR, 2015)).

The advance query used in this research retrieved a total of 1,034 documents from 2015 to 2019, of which 137 corresponded to Open Access (Articles publishes in "Gold" OA, including full OA journals, hybrids, Open Archive and Promotional Access). The total citations of these publications are also used in this study and these were calculated up to July 28, 2019.

The documents and its raw data were retrieved from Scopus as plan text and processed into SciMAT to set the knowledge base for the performance analysis and science mapping. In order to improve the data quality a de-duplicating and meaningless keywords process were carried out (e.g., "BIG-DATA", "BIG-DATA-(BD)" and "BIGDATA" were merged as "BIG-DATA" | "ALGORITHM", "SYSTEM" or "APPLICATION" were stopped).

3. Performance analysis

The bibliometric performance analysis assesses the documents and its citations with the objective of evaluating and testing the field growth. Additionally, it analyzes the authors, publications, sources, most relevant documents and subject areas to evaluate the impact of the these [11].

The Figure 1 shows the distribution of citations and documents related to Big Data field in the Economics, Econometrics and Finance field.

	2015	2016	2017	2018	2019
Documents	126	201	336	233	138
Documents acum	126	327	663	896	1.034
Citations acum	2.445	3.912	4.670	5.047	5.103

Fig. 1. Distribution of documents by year from 2015 to 2019

Taking into account the results of the figure, the documents have been increasing year by year from 2015 to 2017, and this evolution reveals the growing interest in its main components and application. Nevertheless, there were fewer publications in the previous year than in 2017, because the accepted publications in 2018 had not been indexed before the collection date indicated earlier. In addition, the last year has to be considered still active. Finally, the most cited authors and the most productive authors are presented in Table 1.

Table 1. Most cited and productive authors from 2015 to 2019

Cites (documents)	Most cited authors	Documents (cites)	Most productive authors
601 (4)	Akter, S.	6 (338)	Gunasekaran, A.
381 (2)	Fosso Wamba, S.	6 (28)	Li, X.
389 (1)	Van der Aalst, W.	6 (2)	Liu, Y.
379 (1)	Chopin, G.	5 (4)	Chen, J.
379 (1)	Edwards, A.	5 (3)	Li, K.
379 (1)	Gnanzou, D.	5 (16)	Zhang, Y.
372 (1)	Hastie, T.	4 (601)	Akter, S.
372 (1)	Tibshirani, R.	4 (126)	Dubey, R.
372 (1)	Wainwright, M.	4 (4)	Liu, C.
338 (6)	Gunasekaran, A.	4 (2)	Wang, Y.
222 (3)	Wamba, S.F.	4 (2)	Zhang, L.
222 (3)	Xiang, Z.		

It is important to highlight that Akter, S. and Gunasekaran, A. are included in both rankings during the evaluated period. Moreover, it should be noted that in some positions, a tie was recorded between different authors, thus, all of them have been arranged in alphabetical order.

In this context, the most productive regions and countries associated to the Big Data research in the Economics, Econometrics and Finance field during the last five years (2015-2019) are United States (297), United Kingdom (108), China (106), Germany (83), India (72), Australia (68), France (47) and Spain (38).

Furthermore, the most productive organizations are University College London (10), Tsinghua University (8), University of New South Wales Australia (8), The University of Sydney (8), Ohio State University (7), Hong Kong Polytechnic University (7), Renmin University of China (7), George Mason University, Fairfax Campus (7), University of California, Berkeley (7), University of Oxford (7), City University of Hong Kong (7) and University of South Australia (7).

Finally, the most productive sources are Studies in Systems Decision and Control (22), Statistical Journal of the Iaos (20), International Journal of Production Economics (14), Asta Wirtschafts Und Sozialstatistisches Archiv (11), Electronic Markets (11), Concurrences (8), Electronic Commerce Research (8), Financial Innovation (8), Journal of Econometrics (8), Economie et Statistique (7), Journal of Business and Economic Statistics (7) and Journal of Marketing Analytics (7).

4. Conceptual analysis

In this section, an overview of the relations between core themes of the Big Data in Economics, Econometrics and Finance field is provided. This section is structured in two parts, the first one is the strategic diagram from 2015 to 2019, and the second ones are the sub research themes within in the main research themes and its relationship.

The main research themes identified have been classified into four classes according their relevance. Moreover, the main research themes included in the strategic diagram are represented as spheres. These spheres are proportional to the number of documents related with each research theme and it considers the number of citations achieved by each one (in parenthesis).

From 2015 to 2019, twelve research lines (themes) associated to Big Data research in the Economics, Econometrics and Finance field could be identified and arranged (see Figure 2). Therefore, six themes are considered key due to their contribution to the field growth (Q1-Motor themes and Q4-Basic and transversal themes): DATA-MINING, DATA-HANDLING, INTERNET-OF-THINGS-(IOT), BIG-DATA-ANALYTICS-(BDA), SUPPLY-CHAIN-MANAGEMENTS-(SCM) and DECISION-MAKING-PROCESS.

Table 3 shows the performance of the main bibliometric indicators (h-index, citations achieved and number of publications) from 2015 to 2019 and the strategic quadrant position of each research themes. According to the

results, in terms of production the most significant themes of the Big Data research within Economics, Econometrics and Finance field are MACHINE-LEARNING-(ML), DATA-SETS and BIG-DATA-ANALYTICS-(BDA). On the other hand, the most important themes in terms of citations are BIG-DATA-ANALYTICS-(BDA), SUPPLY-CHAIN-MANAGEMENTS-(SCM) and DECISION-MAKING-PROCESS.
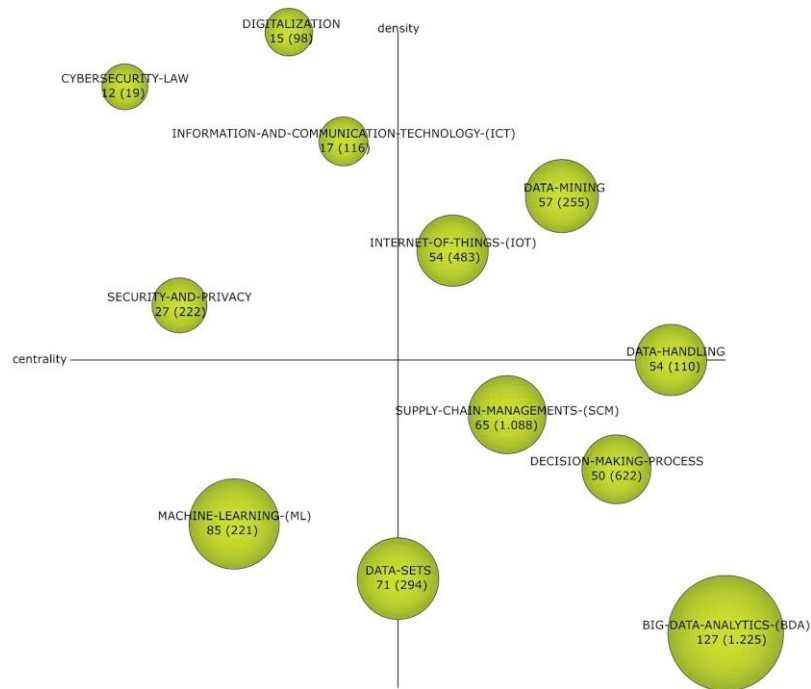

Fig. 2. Strategic diagram from 2015 to 2019

Table 2. Performance of the research themes from 2015 to 2019

Theme	Quadrant	Documents	Citations	h-index
BIG-DATA-ANALYTICS-(BDA)	4	127	1,225	13
MACHINE-LEARNING-(ML)	3	85	221	8
DATA-SETS	3	71	294	9
SUPPLY-CHAIN-MANAGEMENTS-(SCM)	4	65	1,088	12
DATA-MINING	1	57	255	6
DATA-HANDLING	1	54	110	4
INTERNET-OF-THINGS-(IOT)	1	54	483	9
DECISION-MAKING-PROCESS	4	50	622	8
SECURITY-AND-PRIVACY	2	27	222	7
INFORMATION-AND-COMMUNICATION-TECHNOLOGY-(ICT)	2	17	116	4
DIGITALIZATION	2	15	98	3
CYBERSECURITY-LAW	2	12	19	2

The important features of the research lines (themes) identified from 2015 to 2019 and their main research areas are shown in the Figure 3 and described below.
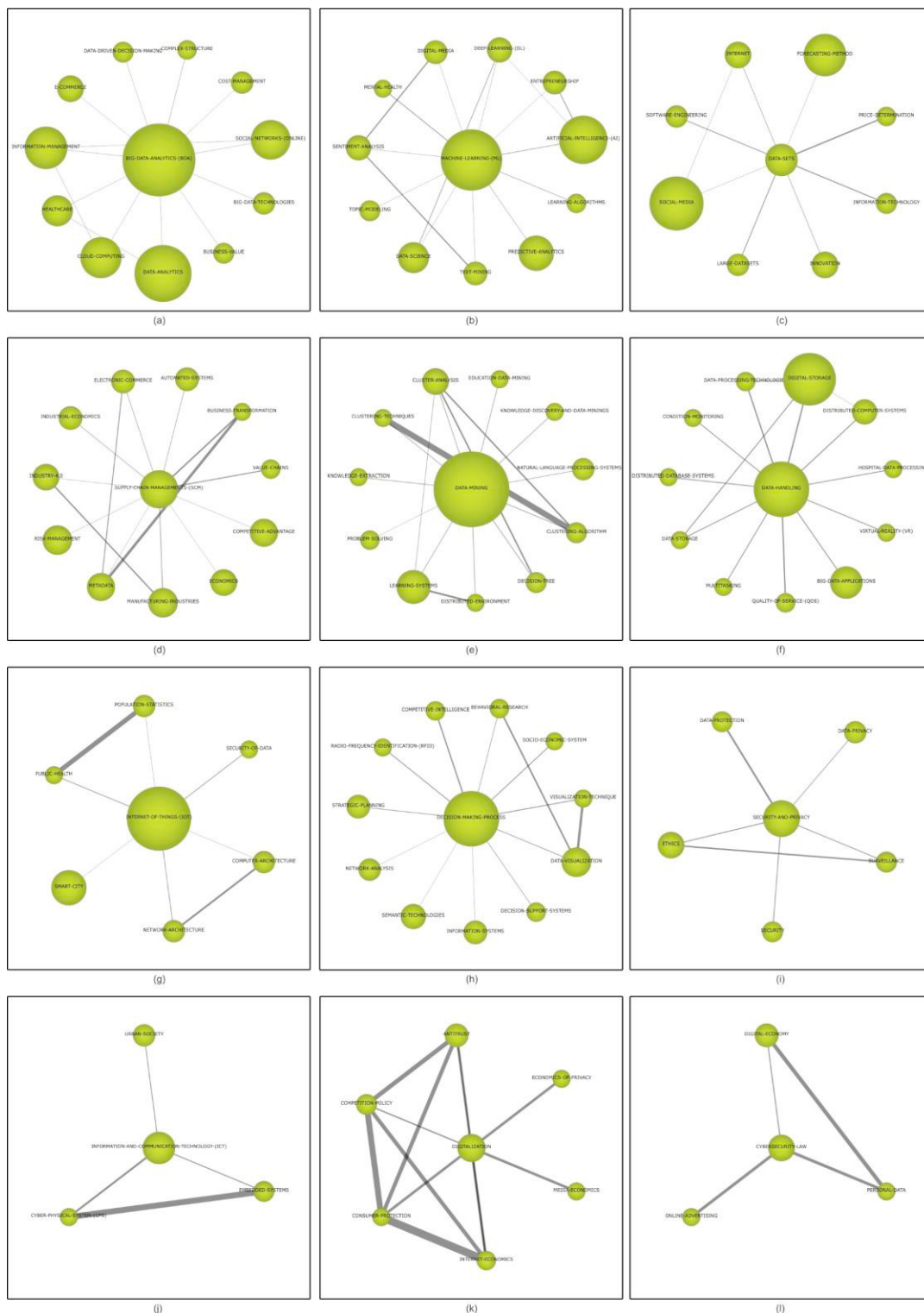
Fig. 3. Intellectual structure for each research theme from 2015 to 2019

BIG-DATA-ANALYTICS-(BDA) (Figure 3.(a)) is the first research theme ranked in terms of number of citations and documents achieved. This theme provides a means to evaluate data sets and define conclusions about them to support organizations make informed strategic decisions. The most important subthemes within this research themes are (documents): DATA-ANALYTICS (29), INFORMATION-MANAGEMENT (19), CLOUD-COMPUTING (18) and SOCIAL-NETWORKS-(ONLINE) (17).

MACHINE-LEARNING-(ML) (Figure 3.(b)) is the second theme ranked in terms of number of documents. It is primarily focused on the development and application of models and algorithms that information systems use in order to perform a detailed assignment effectively. The most important subthemes within this research themes are (documents): ARTIFICIAL-INTELLIGENCE-(AI) (23), PREDICTIVE-ANALYTICS (14), DATA-SCIENCE (11), SENTIMENT-ANALYSIS (7) and DEEP-LEARNING-(DL) (7).

DATA-SETS (Figure 3.(c)) is referred to the data in a collection . The most important subthemes within this research themes are (documents): SOCIAL-MEDIA (27), FORECASTING-METHOD (19), INNOVATION (8) and INTERNET (8).

SUPPLY-CHAIN-MANAGEMENTS-(SCM) (Figure 3.(d)) is focused on the implementation of the big data in the value chain. The most important subthemes within this research themes are (documents): MANUFACTURING-INDUSTRIES (10), RISK-MANAGEMENT (10), COMPETITIVE-ADVANTAGE (9) and METADATA (9).

DATA-MINING (Figure 3.(e)) is devoted to the process of determining and discovering patterns in large data sets involving approaches at the connection of database systems, statistics and machine learning. The most important subthemes within this research themes are (documents): LEARNING-SYSTEMS (13), CLUSTER-ANALYSIS (8) and NATURAL-LANGUAGE-PROCESSING-SYSTEMS (5).

DATA-HANDLING (Figure 3.(f)) deals with the process of ensuring that research data is stored, archived or disposed. The most important subthemes within this research themes are (documents): DIGITAL-STORAGE (26), BIG-DATA-APPLICATIONS (12) and DISTRIBUTED-COMPUTER-SYSTEMS (7).

INTERNET-OF-THINGS-(IOT) (Figure 3.(g)) is related to the use of data by IOT devices. The most important subthemes within this research themes are (documents): SMART-CITY (14), COMPUTER-ARCHITECTURE (5), NETWORK-ARCHITECTURE (5) and POPULATION-STATISTICS (5).

DECISION-MAKING-PROCESS (Figure 3.(h)) is an efficient way to make thoughtful, informed decisions by gathering information, assessing alternative resolutions and identifying a decision. The most important subthemes within this research themes are (documents): DATA-VISUALIZATION (10), SEMANTIC-TECHNOLOGIES (7), INFORMATION-SYSTEMS (6) and STRATEGIC-PLANNING (6).

SECURITY-AND-PRIVACY (Figure 3.(i)) deals with the privacy and data protection in Big Data environments. The most important subthemes within this research themes are (documents): ETHICS (8), DATA-PRIVACY (5), SURVEILLANCE (4), SECURITY (4) and DATA-PROTECTION (4).

INFORMATION-AND-COMMUNICATION-TECHNOLOGY-(ICT) (Figure 3.(j)) is mainly focused on the integration of the new ICT. The most important subthemes within this research themes are (documents): URBAN-SOCIETY (5), EMBEDDED-SYSTEMS (4) and CYBER-PHYSICAL-SYSTEM-(CPS) (2).

DIGITALIZATION (Figure 3.(k)) is related to the process of converting information into a digital format and its utilization. The most important subthemes within this research themes are (documents): ANTITRUST (6), COMPETITION-POLICY (4), INTERNET-ECONOMICS (3), MEDIA-ECONOMICS (2), CONSUMER-PROTECTION (2) and ECONOMICS-OF-PRIVACY (2).

CYBERSECURITY-LAW (Figure 3.(l)) is mainly connected to policies, regulation and directives that safeguard computer systems and information technology. The most important subthemes within this research themes are (documents): DIGITAL-ECONOMY (6), PERSONAL-DATA (2) and ONLINE-ADVERTISING (2).

5. Conclusions

This research presents the most extensive analysis of the evolution and intellectual structure of Big Data field in Economics, Econometrics and Finance field available in literature. One of the objectives of this research is to show the main trends in the knowledge field and to provide a starting point that allows academics and professionals to find new opportunities, or strengthen their research lines.

In performance terms, the dimension and weight of literature related to Big Data in Economics, Econometrics and Finance field exposed a noticeable increase from 2015 to 2019. Taking into consideration the large volume of documents and citations recorded, it is expected that the development and application of it will be seen as part of other knowledge areas.

The Big Data strategic map reveals that it has to different approaches. The first approach is related to Computer Sciences (data mining, data handling, data set, machine learning and big data analytics); the second one is nearly to the organizations, innovation and competitiveness (decision making process, supply chain management, information and communication technology and digitalization), and the last one is focused on the security and privacy (security and privacy, and cybersecurity law). Finally, the current study underlines the fact that Big Data is a growing field of research with different approaches. Moreover, some future research lines are necessary to provide a more in-depth examination of the Big Data in other fields and their relationship.

Acknowledgements

References

[1] Oussous A, Benjelloun FZ, Lahcen AA, Belfkih S, Big Data technologies: A survey, Journal of King Saud University-Computer and Information Sciences 2018; 30:431-448.
[2] López-Robles JR, Otegi-Olaso JR, Porto Gómez I, Cobo MJ, 30 years of intelligence models in management and business: A bibliometric review, Int J Inf Manage 2019; 48:22-38.
[3] López-Robles JR, Otegi-Olaso JR, Arcos R, Gamboa-Rosales NK, Gamboa-Rosales H, Mapping the structure and evolution of JISIB: A bibliometric analysis of articles published in the Journal of Intelligence Studies in Business between 2011 and 2017, J. Intell. Stud. Bus. 2018; 8.
[4] López-Robles JR, Otegi-Olaso JR, Gamboa-Rosales NK, Gamboa-Rosales H, Cobo MJ, 60 Years of Business Intelligence: A Bibliometric Review from 1958 to 2017, New Trends in Intelligent Software Methodologies, Tools and Techniques: Proceedings of the 17th International Conference SoMeT_18, IOS Press; 2018; p. 395.
[5] McAfee A, Brynjolfsson E, Davenport TH, Patil DJ, Barton DC, Big data: the management revolution, Harv. Bus. Rev. 2012; 90:60-68.
[6] López-Robles JR, Otegi-Olaso JR, Porto Gómez I, Gamboa-Rosales NK, Gamboa-Rosales H, Robles-Berumen H, Bibliometric Network Analysis to Identify the Intellectual Structure and Evolution of the Big Data Research Field, International Conference on Intelligent Data Engineering and Automated Learning, Springer; 2018; p. 113-120.
[7] López-Robles JR, Guallar J, Otegi-Olaso JR, Gamboa-Rosales NK, El profesional de la información (EPI): bibliometric and thematic analysis (2006-2017), El profesional de la información 2019; 28:e280417.
[8] Cobo MJ, SciMat: herramienta software para el análisis de la evolución del conocimiento científico. Propuesta de una metodología de evaluación, Granada: Universidad de Granada; 2012.
[9] Cobo MJ, López‐Herrera AG, Herrera‐Viedma E, Herrera F, SciMAT: A new science mapping analysis software tool, J. Am. Soc. Inf. Sci. Technol. 2012; 63:1609-1630.
[10] Martínez MA, Herrera M, López-Gijón J, Herrera-Viedma E, H-Classics: Characterizing the concept of citation classics through H-index, Scientometrics 2014; 98:1971-1983.
[11] Gutiérrez-Salcedo M, Martínez MA, Moral-Munoz JA, Herrera-Viedma E, Cobo MJ, Some bibliometric procedures for analyzing and evaluating research fields, Appl Intell 2018; 48:1275-1287.