



Bibliometric Mapping of Big Data (BD) in Higher Education (HE): Towards a comprehensive framework

Baomin Qi

Business School, Liverpool Hope University

Emilia Vann Yaroson

University of Huddersfield

Atif Sarwar

Liverpool Hope University (Business School)

Bibliometric Mapping of Big Data (BD) in Higher Education (HE): Towards a comprehensive framework

Abstract

Although big data (BD) has emerged rapidly over the last decade, its importance within higher education (HE) has remained scarce in academic literature. This research aims to develop a comprehensive framework for using big data in HE. We achieved our research objective by conducting a bibliometric analysis of the available literature on BD in HE published in the English language between 2013 -2021. A total of 4312 articles were considered for analysis. Our results showed that most studies focused on the technical specifications of BD, such as data mining and Hadoop. There was a slight reference to the operationality and management functions. However, it is pertinent to note that data privacy, advanced analytics, and machine learning were highlighted as emerging topics. It, therefore, suggests the importance of advanced data analytics and data privacy in establishing a comprehensive framework for managing BD in HE.

Introduction

Big Data Analytics (BDA) has emerged rapidly in the last decade and has become prominent in managing modern businesses (Chaurasia et al, 2017). Applications of BDA adoption include traffic analysis, shopping recommendations and students module results prediction demonstrating the power of big data, and substantial investment in research and technology development in BDA has been augmented recently. Big data analytics has been adopted by public education across the globe and become a vital force in Higher Education (HE) teaching reform (Daniel, 2014). In general, technological development in education has stalled over the past decades, BDA technologies present a new wave of technological advancement and hold the key to the future of HE.

Big educational data plays a critical role in HE, the analysis and utilisation of such data enhances students' academic performance and the improved student retention. The potential benefits of applying BDA in HE have interested many researchers over the past two decades; the number of academic publications has increased exponentially and is still growing rapidly. Despite the increased number of publications in this area, there is a lack of systematic literature revealing the research landscape and future trends. This research aims to perform a bibliometric mapping of the BDA researches in HE and provide a systematic and transparent review of the current literature.

Big Data Analytics in HE

BD and BDA have recently become buzzwords in academia and industry. Universities and colleges across the globe have been trailing ways of applying BDA to improve both teaching and learning (Tatt, 2019). Many benefits are realised from mining the big data captured by HE institutions (HEIs). Such benefits range from tracking and monitoring students' learning progress to predicting students' academic performance and achievements. Students' data such as their learning journeys, readings, coursework completion, and test and examination results were captured and analysed to predict their academic performance. The insights also allows teachers to promptly identify each student's learning situation and potential issues enabling tailored guidances and support to those concerned students. Applications and tools have been developed in this domain and utilised by HEIs to support teaching and learning (Hassan, 2015). The wealth of educational data is analysed and trends in student scores, attendance, dropout

rates, and retention been identified at early stages of learning (Marín-Marín et al., 2019). Results from BDA could provide personalised experience and services for students too. By analysing the valuable data in the learning process, unique learning experience and personalised learning support services could be provided. The application of BDA in the HE sector has been rapidly developed in recent years and various applications have been implemented in HEIs (Javidi, et al., 2017; Dinter et al, 2017)

Despite a large number of researches published in BDA in HE our initial review of literature revealed that major focus in this domain is largely on the exploration of benefits and challenges of BDA. The aim of our research is to conduct a comprehensive review of the current literature and to identify the current research landscape and establish the future research direction. Researchers in various disciplines have performed bibliometric analysis to map out future research pathways, including service marketing (Donthu et al. 2021); Covid-19 and consumer behaviour (Cruz-Cárdenas et al., 2021); supply chains and carbon taxes (Zhou et al., 2021); BD in supply chain management (Mishra et al., 2018). However, there are limited studies that have performed bibliometric analysis to examine BDA trends in HE (Marin-Marín et al., 2019; Rodríguez Jiménez et al., 2019). This study extends the existing literature by exploring BDA research trends in HE and developing research themes to identify areas for further research.

Research Methodology

Our research investigates BDA research in higher education (HE), the literature was systematically reviewed following Tranfield et al. (2003) and the updated PRISMA guideline by Page et al. (2021) recommendations. We considered using the systematic review of literature due to the transparent and replicable process provided, which enhances the overall quality of the review (Tranfield et al., 2003). We also align our bibliometric analysis following the guidelines provided by Donthu et al. (2021). The bibliometric analysis is a cross-disciplinary technique used in analysing large volumes of data to identify for example research topics and trends within a discipline.

Article selection

Our data was collected from articles in Scopus database. We deemed it appropriate because of its reputation as the most significant database, with over 77 million bibliographic information on peer-reviewed articles. As world-leading multidisciplinary research publisher Scopus are extensively used for bibliometric analysis (Donthu et al., 2020; Goh and See et al., 2021). We established our review process following the guidelines by Tranfield et al. (2003) and PRISMA (Page et al., 2021) (see Figure 1). To this end, we imposed restrictions on our data and limited them to full-length articles and conference papers published in English language between January 2013 and January 2022. Our sample time frame suggests the emergence of published articles within this period. These articles were empirical and conceptual in nature.

We also restricted our data sample to articles published in business studies/management, decision sciences, environmental sciences, transportation sciences, business finance and social sciences, operations, research management, economics subject areas, with the aim to achieve a wider coverage and comprehensive list of articles. Our search criteria were based on advanced search queries of ** Big Data, Higher Education** in the title, abstract and keywords. We used standard Boolean operators 'OR' and 'AND' in our search queries. Our final sample set from our literature search included 4,621 documents and these were used in our bibliometric analysis.

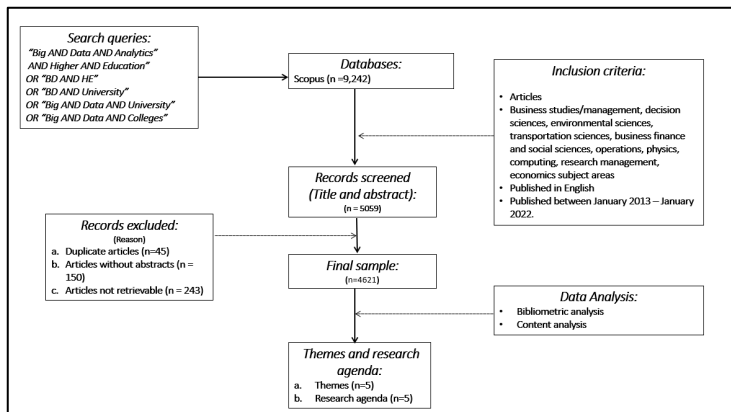


Figure 1: Literature Search criteria

Analytical methods

We analysed our data using a two-step process involving the bibliometric analysis and meta-analysis techniques. These methods were considered suitable due to their transparent and objective lenses centred on conceptual structures. We also employed the meta-analysis technique, made up of co-word analysis and Intertopic Distance Map (IDM), to understand our study's progress of research evolution. The co-word analysis provided the core themes developed, and the IDM validated the developed themes through visual cues of closely linked keywords (Zhang et al., 2021). It implies that the developed themes in our study of BDA in HE emanated from the density of strongly linked keywords. Biblioshiny, an open-source application on R, analysed our data (Aria and Cuccurull 2017). Biblioshiny supports extensive bibliometric data from various sources like Scopus and has successfully mapped scientific literature.

Bibliometric analysis and results

Bibliometric Information

The bibliometric information shows that conference papers account for most of our sample at 3061, and full-length articles account for 1560. We needed to include conference papers in our sample due to the novelty of our research domain.

Description	Results
MAIN INFORMATION ABOUT DATA	
Timespan	2013:2022
Sources (Journals, Books, etc)	1853
Documents	4621
Average years from publication	2.53
Average citations per documents	4.888
Average citations per year per doc	1.042
References	1
DOCUMENT TYPES	
article	1560
conference paper	3061
DOCUMENT CONTENTS	
Keywords Plus (ID)	15643
Author's Keywords (DE)	10789
AUTHORS	
Authors	9972
Author Appearances	13630
Authors of single-authored documents	938
Authors of multi-authored documents	9034
AUTHORS COLLABORATION	
Single-authored documents	1198
Documents per Author	0.463
Authors per Document	2.16
Co-Authors per Documents	2.95
Collaboration Index	2.64

Figure 2 Bibliometric Information

Publication trend and trend topics

For the period under review (2013-2022), Figure 3 shows the top 20 journals that have published articles relating to BDA in HE. It shows that 22% of the articles were published in the *Journal of Physics Conference Series*, *Advances in Intelligent Systems and ACM conference proceedings*. It implies that the research around BD in HE have focused more on the computing aspects and implications. However, there are limited full-length papers and limited research published in operations management journals depicting the budding nature of research within this field.

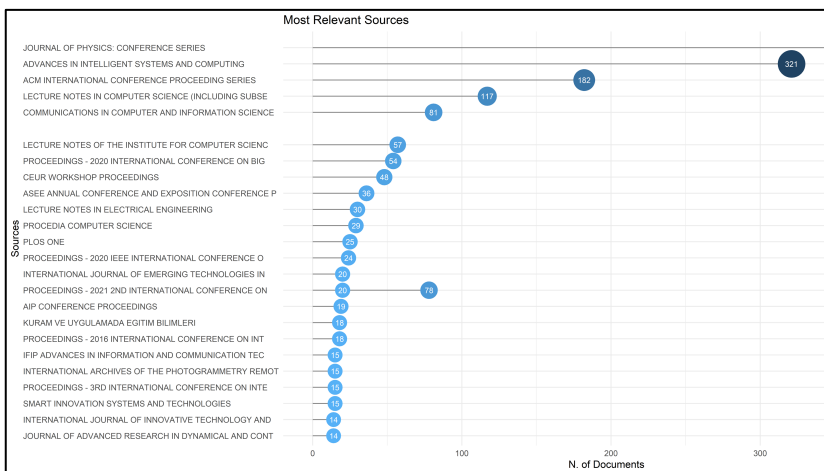


Figure 3 Top Journals Publishing Related Articles

Trend Topics

The research topics were identified by using the top 50 keywords with a 12 words minimum frequency and five words per year. We find that between 2013 and 2015, topics related to BD and HE included "big datum", "research", "world-wide-web", "industry", "mobile devices", and "algorithms"- all indicating business processes. These topics depicted research on BD in HE focused on the explosion of big data. The use of big data in HE was sparsely recognised. The research trends between 2015 and 2017 can be categorised as education database systems as the topic trends evolved to terms such as "education", "Hadoop", "map-reduce", "distributed computer systems", "learning analytics", "digital storage", "information systems", "data reduction", and "database systems". It implies that research in BD and HE between this period focused on how to handle the explosion of big data in HE with limited focus on how HE's used these big data. After 2017 however, we began to see topics including "decision-making", "teaching models", "artificial intelligence", "information management", and "advanced analytics" emerge. The topic trend here suggests that research focuses on how BD can be used in HE, including developing better teaching models and decision-making.

Co-occurrence network of keywords

We present the co-occurrence network of keywords in Figure 4, where there are two clusters of keywords. The first cluster is "big data", which centres around data security, quality control, decision making, intelligent systems, digital storage and data visualisation. This cluster also branches out to include data mining, social media, information classification and deep learning. The second cluster focuses on students with keywords such as e-learning, teaching, curricula, college students and education. Thus one cluster (students) researches how data is generated, and the other is concerned with privacy, security and digital control. The co-occurrence map corroborates the topic trend diagram in Figure 4 and reiterates the dimensions of research within BD in HE. The co-occurrence map identifies several research paths. For instance, we find that data analytics, machine learning and advanced analytics are small clusters between the two significant clusters. In the scheme of things understanding BD in HE requires data analytics and advanced analytics to develop efficient frameworks.

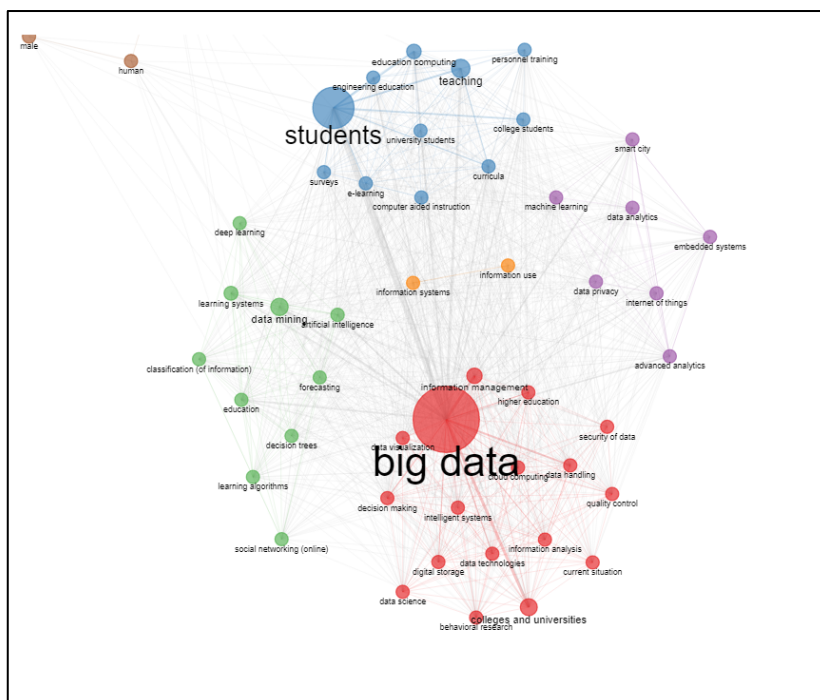


Figure 4: Co-occurrence Network

Conceptual structure map

The conceptual structure map is presented in Figure 5. The graph was plotted with 100 terms and a minimum of 5 documents per year using the multiple correspondence analysis. We identified three clusters of documents communicating related concepts. From left to right, the first cluster mapped in red represents well-researched areas, the green cluster represents emerging research, and the blue cluster represents under-researched topics within our research nexus. Against this backdrop, crucial words that require further research in BD and HE include data analytics, data privacy, machine learning, embedded systems and advanced analytics, indicating the future research paths.

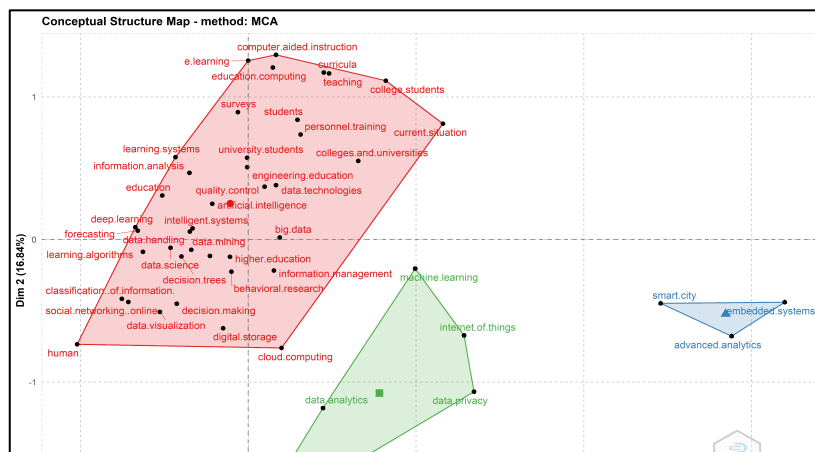


Figure 5: Conceptual structure map of BD in HE

Conclusion and future research

BDA has attracted the research attentions in the HE sector, our research conducted a bibliometric analysis and examined available literature. It identified the publication trend in this area and the leading publications; it mapped the co-occurrence network of keywords revealing the two clusters of key words in this research area: "big data" concerning such as data security, quality of data and data visualisation. Our research also constructed a Conceptual structure map using the multiple correspondence analysis, revealing that future research directions in BDA in HE are data analytics, data privacy, machine learning, embedded systems and advanced analytics.

The main limitation of our research is the only use of Scopus in our literature search, it will be more comprehensive if WoS, google scholars and Dimensions were also selected. The next stage of our research will focus on the detailed scrutinising of the key publications identified and develop a comprehensive framework to aid the BDA implementations in HE.

References

- Aria, M. and Cuccurullo, C. (2017) Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), pp.959-975.
- Chaurasia, SS and Frieda Rosin, A. (2017) From Big Data to Big Impact: analytics for teaching and learning in higher education. *Industrial and Commercial Training*, Vol. 49 No. 7/8, pp. 321-328. <https://doi-org.ezproxy.hope.ac.uk/10.1108/ICT-10-2016-0069>
- Cruz-Cárdenas, J., Zabelina, E., Guadalupe-Lanas, J., Palacio-Fierro, A. and Ramos-Galarza, C. (2021) COVID-19, consumer behavior, technology, and society: A literature review and bibliometric analysis. *Technological forecasting and social change*, 173, p.121179.
- Daniel, B. (2014) *British journal of educational technology*, 2015-09, Vol.46 (5), p.904-920

- Reyes, J.A. (2015). The skinny on big data in education: Learning analytics simplified. *TechTrends*, 59(2), pp.75-80.
- Dinter, B., Kollwitz, C. and Fritzsche, A. (2017) Teaching Data Driven Innovation-Facing a Challenge for Higher Education. In *AMCIS*.
- Donthu, N., Kumar, S. and Pattnaik, D. (2020) Forty-five years of journal of business research: a bibliometric analysis. *Journal of Business Research*, 109, pp.1-14.
- Donthu, N., Kumar, S., Mukherjee, D., Pandey, N. and Lim, W.M. (2021) How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, 133, pp.285-296.
- Donthu, N., Kumar, S., Ranaweera, C., Pattnaik, D. and Gustafsson, A. (2021) Mapping of journal of services marketing themes: a retrospective overview using bibliometric analysis. *Journal of Services Marketing*.
- Ertz, M. and Leblanc-Proulx, S. (2018) Sustainability in the collaborative economy: A bibliometric analysis reveals emerging interest. *Journal of Cleaner Production*, 196, pp.1073-1085.
- Goh, K.H. and See, K.F. (2021) Twenty years of water utility benchmarking: A bibliometric analysis of emerging interest in water research and collaboration. *Journal of Cleaner Production*, 284, p.124711.
- Hassan, S. (2015) A model recommends best machine learning algorithm to classify learners based on their interactivity with moodle. In *2015 Second International Conference on Computing Technology and Information Management (ICCTIM)* (pp. 49-54). IEEE.
- Javidi, G., Rajabion, L. and Sheybani, E. (2017) December. Educational data mining and learning analytics: Overview of benefits and challenges. In *2017 International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 1102-1107). IEEE.
- Kumar, S., Pandey, N., Lim, W.M., Chatterjee, A.N. and Pandey, N. (2021) What do we know about transfer pricing? Insights from bibliometric analysis. *Journal of Business research*, 134, pp.275-287.
- Marín-Marín, J.A., López-Belmonte, J., Fernández-Campoy, J.M. and Romero-Rodríguez, J.M. (2019) Big data in education. A bibliometric review. *Social Sciences*, 8(8), p.223.
- Mishra, D., Gunasekaran, A., Papadopoulos, T. and Childe, S.J. (2018) Big Data and supply chain management: a review and bibliometric analysis. *Annals of Operations Research*, 270(1), pp.313-336.
- Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E. and Chou, R. (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *International Journal of Surgery*, 88, p.105906.
- Rodríguez Jiménez, C., Sanz Prieto, M. and Alonso García, S. (2019) Technology and higher education: A bibliometric analysis. *Education Sciences*, 9(3), p.169.
- Tatt, S. (2019). *Do you know how universities and colleges use Big Data?* Big Data Made Simple. Online available at: <https://bigdata-madesimple.com/do-you-know-how-universities-and-colleges-use-big-data/?cv=1> [Accessed 9 May 2021].
- Tranfield, D., Denyer, D. and Smart, P. (2003) Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, 14(3), pp.207-222.
- Verma, S. and Gustafsson, A. (2020) Investigating the emerging COVID-19 research trends in the field of business and management: A bibliometric analysis approach. *Journal of Business Research*, 118, pp.253-261.
- Zhang, Y. (2018) Model innovation and teaching effect evaluation of accounting teaching in higher vocational colleges in the era of big data. *Educational Sciences: Theory & Practice*, 18(6).

Zhang, Y., Hou, Z., Yang, F., Yang, M.M. and Wang, Z. (2021) Discovering the evolution of resource-based theory: Science mapping based on bibliometric analysis. *Journal of Business Research*, 137, pp.500-516.