RELATIONSHIP BETWEEN BIG DATA AND DECISION SUPPORT SYSTEMS

Relação entre Big Data e Sistemas de Apoio a Decisão

Gustavo Grander[1], Luciano Ferreira da Silva[1], Alan Tadeu de Moraes[1], Paulo Sergio Gonçalves de Oliveira[2],
[1] Universidade Nove de Julho – Uninove – São Paulo - Brazil
[2] Universidade Anhembi Morumbi, São Paulo, Brazil
E-mail: grandergustavo@gmail.com, lf_silvabr@yahoo.com.br,
alant.moraes@gmail.com, psgoliveira@hotmail.com

ABSTRACT

This article aimed to identify relationships between Big Data and Decision Support Systems. For this, we conducted a search in the Scopus database and as a result, we identified a report according to the increased frequency of publications, frequency of publications in journals and, using the VOSviewer software, we performed an analysis of words co-citation. We identified 5 groups of keywords that suggest different areas of study (e.g., logistics, health and social media), as well as a more recent focus on studies aimed at sustainable development, machine learning, analytical techniques and decision-making processes decision. An important contribution that should also be highlighted was the strong relationship between the keywords Big Data, artificial intelligence and decision making, suggesting studies involving the three terms in a large number of works.
Keywords: Big Data; Decision Support System; Machine learning; VOSviewer; Artificial Intelligence; Bibliometry.

RELAÇÃO ENTRE BIG DATA E SISTEMAS DE APOIO A DECISÃO

Relationship between big data and decision support systems

Gustavo Grander¹, Luciano Ferreira da Silva¹, Alan Tadeu de Moraes¹, Paulo Sergio Gonçalves de Oliveira²,
¹ Universidade Nove de Julho – Uninove – São Paulo - Brazil
² Universidade Anhembi Morumbi, São Paulo, Brazil
E-mail: grandergustavo@gmail.com, lf_silvabr@yahoo.com.br,
alant.moraes@gmail.com, psgoliveira@hotmail.com

RESUMO

Este artigo teve como objetivo identificar as relações entre Big Data e sistemas de suporte a decisões. Para isso, realizamos uma pesquisa no banco de dados Scopus e, como resultado, identificamos um relatório de acordo com o aumento da frequência de publicações, frequência de publicações em periódicos e, utilizando o software VOSviewer, realizamos uma análise da co-citação de palavras. Identificamos 5 grupos de palavras-chave que sugerem diferentes áreas de estudo (por exemplo, logística, saúde e mídias sociais), bem como um foco mais recente em estudos voltados para o desenvolvimento sustentável, aprendizado de máquina, técnicas analíticas e decisão de processos de tomada de decisão. Uma contribuição importante que também deve ser destacada foi a forte relação entre as palavras-chave Big Data, inteligência artificial e tomada de decisão, sugerindo estudos envolvendo os três termos em um grande número de trabalhos.
Palavras-chave: Big Data; Decision Suporte System; Aprendizado de máquina; VOSviewer; Inteligência Artificial; Bibliometria

INTRODUCTION

Big Data can be defined as a generic term for any collection of large, complex datasets that are difficult to store, process, analyze, and understand using traditional database processing tools (Huang & Chaovalitwongse, 2015). Big Data has emerged as a paradigmatic shift on how organizations make decisions (Mortenson, Doherty, & Robinson, 2015). There is a common acceptance that Big data can be adopted at strategic, tactical, and operational levels, thus improving existing decision-making practices (Wamba, Ngai, Riggings, & Akter, 2017).

The term "Big Data" has become hugely popular in the business world in recent years (Gupta, Chen, Hazen, Kaurd, & Gonzalez, 2018). Its use is either to support decision making or to make automated decisions (Davenport & Dyche, 2013). Through Decision Support Systems (DSS), it is possible to process large volumes of data using output models and outputs with interfaces that increasingly permeate the professions with a high level of knowledge (Constantiou & Kallinikos, 2014).

From this context, we try to understand what relations Big Data have with DSS. Our goal was to understand how the use of Big Data contributes to DSS in different aspects of use, given its great applicability. To do this a literature mapping was performed, this being a variation of systematic literature review, used in this study to understand the interaction between keywords and to answer the question of research that guided this study: what Big Data relationships have with DSS?

We performed a search in the Scopus database and analyzed the data of 93 articles through the VOSviewer software. We identified 5 clusters through network analysis, recent studies with 'machine learning' and 'sustainable development' terms related to Big Data and DSS, and a strong relationship of the term 'artificial intelligence' to the search terms used in this research.

Our study is organized as follows: We present a brief theoretical basis in item 2, then we describe the methodological procedure and data collection in item 3, then we present the results in item 4, and finally the conclusion in item 5 along with suggestions of future research opportunities in depth to confirm our findings as well as to increase the understanding about the topic that has been presented as very relevant for society.

1. THEORETICAL BACKGROUND

As highlighted by Laney (2001), Big Data has three primary characteristics: volume, velocity and variety, and later complemented by SAS (https://www.sas.com/en_us/insights/big-data/what-is-big-data.html) with the characteristics of variability and complexity. These dimensions required more powerful computers, an ubiquitous network, and algorithms capable of connecting datasets in order to make possible analyzes that would not be possible until then, and this convergence made possible the commercial application of data Science (Provost & Fawcett, 2013). Consequently, this has driven a new generation of technologies to extract value from Big Data (Gantz & Reinsel, 2011).

According to Manyika et al. (2011), Big Data can play a significant economic role for the benefit not only of private trade, but also of national economies and their citizens. The authors also highlight the value generated by the use of Big Data in five domains: health, public sector, retail, manufacturing and global personal location data. There is enormous potential in big data analytics in the healthcare industry, especially with regard to a better understanding of strategic implications (Wang, Kung, & Byrd, 2018). Modern systems produce collections of data sets so large and complex that it is impossible to store and process them manually (Burattin, Cimitile, Maggi, & Sperduti, 2015). Analytical techniques through the DSS can provide for example, annotations, medical prescriptions, images and laboratory information (Raghupathi & Raghupathi, 2014).

The use of big data also helps decision making in public policies by combining geographical information with health care data, such as consumption of tobacco, alcohol and economic factors, public managers can map

areas most likely to experience social problems (Fredriksson, 2018). Big Data processing has also been studied as a business process monitoring and analysis system (Vera-Baquero, Colomo-Palacios, Molloy & Elbattah, 2015), logistics, service and planning (Brinch, Stentoft, Jensen & Rajkumar, 2018). On the other hand, there is a discussion about the reduction of privacy that occurs in highly centralized control environments, where governments can interfere with citizens' freedom (Power, 2016).

To capture the value of big data, techniques and technologies are needed that use a series of disciplines such as mathematics and statistics (Chen & Zhang, 2014) through computational tools (Fayyad, Piatetsky-Shapiro and Smyth (1996). And Big Data analysis must include the phases of data generation, acquisition, storage and analysis (Chen, Mao & Liu (2014).

Machine learning, artificial intelligence, and cognitive computing are dominating conversations about how advanced analytics can provide companies, for example, with a competitive advantage for business. Machine Learning is a form of Artificial Intelligence that allows a system to learn from data rather than explicit programming, and enables data scientists and business analysts to make predictions based on data-based analytical models (Hurwitz & Kirsch, 2018).

2. DATA AND METHODOLOGY

For this study, we used the Scopus database because it is one of the main databases in the area of Social Sciences. We performed the research on 05/30/2019 with the Boolean term: "Big Data" AND ("Decision Theory" or "decision support system*" or "decision-support system*"). The use of the '*' symbol was so that plurals, gerunds or nouns were also identified in the search. As an initial restriction, we limited the search so that the words were contained in title, abstract, or keywords.

As a result, we obtained 1,031 results that were reduced to 243 results after the filtering of areas (limiting to Business, Management and Accounting, Decision Sciences; Social Sciences) and after a new filter was limited to articles only, we obtained 93 results. These papers underwent an individual exploratory analysis with title reading, summary and when necessary, reading directed to the article to identify the possible contributory character to answer our research question.

To help us with the research question, we use the grouping technique using the VOSviewer software. Grouping techniques play a prominent role in bibliometric research by providing clusters of authors, publications and related keywords (Van Eck & Waltman, 2017). This analysis conditioned us to understand the most relevant connections between the articles identified in the research through the relationship between the keywords of the works.

This method is described by Zupic and Čater (2015) as useful in a phase prior to a literature review, for example, as it guides the researcher to the fields of research by reducing subjective bias. Thus, we perform co-occurrence analyzes of keywords through three ways of visualization via VOSviewer: network, overlap and density. The methodological procedures followed in this study can be visualized in Figure 1.
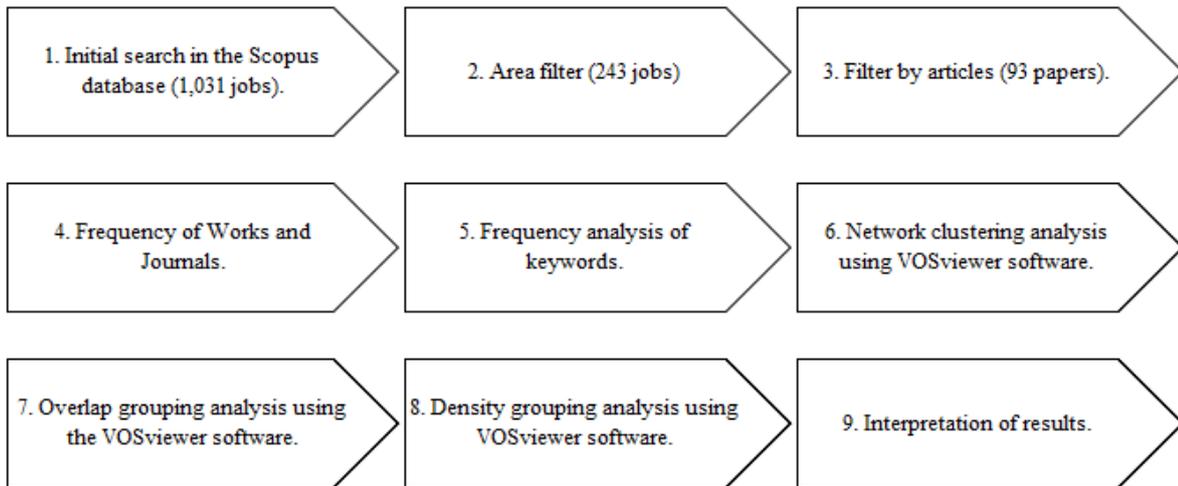
Figure 1. Methodological procedure of this work.

3. RESULTS

Initially, we analyzed the frequency of publications of the 93 selected papers, as can be seen in Figure 2. Note that in the first two years (2012 and 2013) only one publication per year occurred and from 2014 there was an exponential growth of publications, including with the year 2019 already with 8 publications, a year that is still partially counted due to the period of research. Due to our research question, we decided to keep the year 2019 even if we did not have the total of publications for the year.
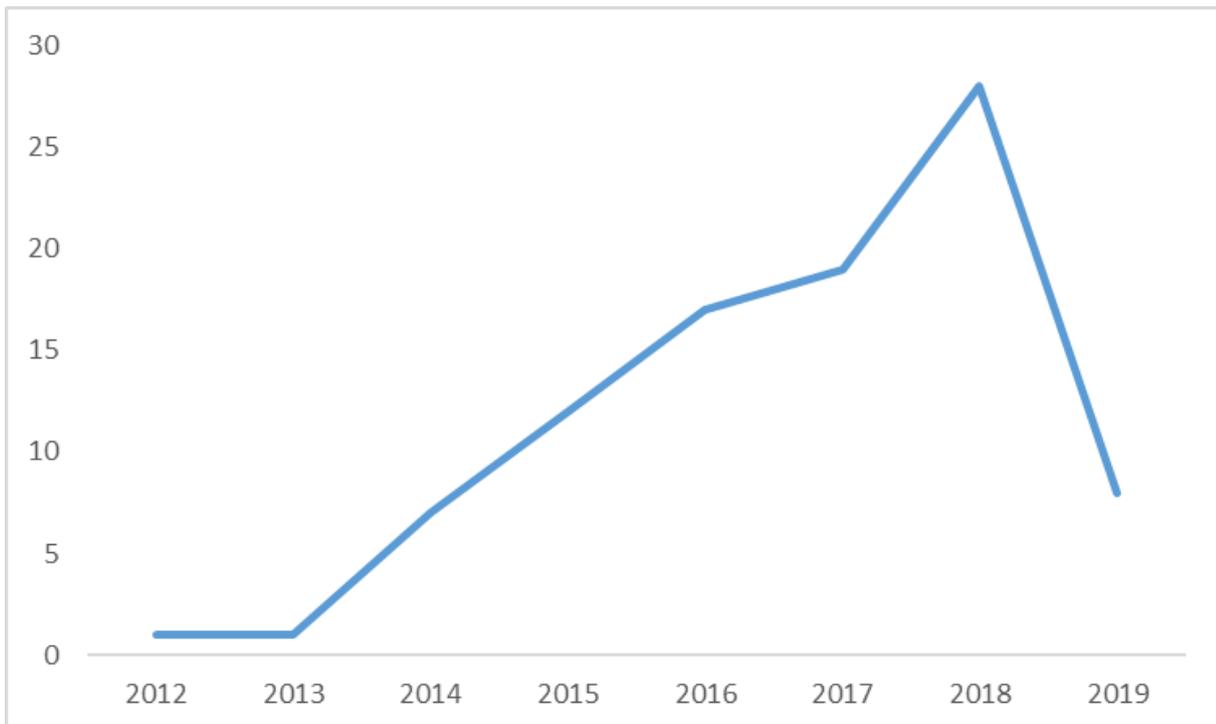

Figure 2. Publications by year

We also count the frequency of works per journal over the period, as presented in Table 1 and we can verify that they were "Journal Of Decision Systems" and "Decision Support Systems" with 9 and 7 papers, respectively. Of the total number of Journals with publications registered in the research, 13 represented 48.39% of the sample (these with more than one registered work), the other Journals had only one work registered, thus representing 51.61%, and were not mentioned in the table.

Table 1. Frequency of works per journal over the period.

Journal	Frequency (n)	Frequency (%)	Cumulative frequency (%)
Journal Of Decision Systems	9	9,68%	9,68%
Decision Support Systems	7	7,53%	17,20%
Lecture Notes In Business Inf. Processing	4	4,30%	21,51%
Sustainability Switzerland	4	4,30%	25,81%
European Journal Of Operational Research	3	3,23%	29,03%
Information Sciences	3	3,23%	32,26%
Knowledge Based Systems	3	3,23%	35,48%
Computers And Operations Research	2	2,15%	37,63%
Computers Envir. And Urban Systems	2	2,15%	39,78%
Dili Xuebao Acta Geographica Sinica	2	2,15%	41,94%
Information And Management	2	2,15%	44,09%
Technol. Forecasting And Social Change	2	2,15%	46,24%
Transportation Research P. Emerg. Technol.	2	2,15%	48,39%

We note that the journals could be classified into four different editorial scopes: decision making, sustainability and the environment, operational research and intelligent and information systems. This shows us different areas of research that are being discussed in the studied subjects and consequently different applications of concepts related to the areas.

Table 2 shows the frequency that the keywords were cited in the analyzed works. As an exclusion criterion, we chose to filter out keywords that had been quoted at least four times, and as a result 32 keywords from a total of 931 were recorded, which were cited 389 times out of a total of 1,400, representing 27.79% of total citations.

It should be considered that the keywords identified with IDs 2 and 5 are separated only by the fact that the first one is in the plural and the second in the singular, but could be considered as the same keyword and in this case it would be the most frequent word (n = 75), the same occurred with the keywords in ID 8 and 10, but we opted to maintain the originality of the data extracted from the software. The 32 most frequently used keywords in the papers therefore represent 27.79% of the total frequency of keywords.

Some terms that are quite recurring have caught our attention, e.g. data mining, which is a set of techniques to extract valuable information from the data and which may involve methods and statistics from Machine Learning (Chen & Zhang, 2014) and cloud computing, which has a strong relationship with Big Data because Big Data offers users the ability to use computing to process queries distributed across multiple data sets and return the resulting sets in a timely manner (Hashem et al., 2015).

Table 2. Most cited keywords in the analyzed database.

ID	Keywords	Frequency (n)	Frequency (%)	Cumulative frequency (%)
1	big data	74	5,29	5,29
2	decision support systems	58	4,14	9,43
3	decision making	36	2,57	12,00
4	artificial intelligence	35	2,50	14,50
5	decision support system	17	1,21	15,71
6	information management	15	1,07	16,79
7	data mining	14	1,00	17,79
8	decision supports	12	0,86	18,64
9	decision theory	11	0,79	19,43
10	decision support	8	0,57	20,00
11	digital storage	8	0,57	20,57
12	data analytics	7	0,50	21,07
13	information analysis	7	0,50	21,57
14	cloud computing	6	0,43	22,00
15	analytics	5	0,36	22,36
16	big data analytics	5	0,36	22,71
17	data handling	5	0,36	23,07
18	information systems	5	0,36	23,43
19	knowledge management	5	0,36	23,79
20	logistics	5	0,36	24,14
21	predictive analytics	5	0,36	24,50
22	social media	5	0,36	24,86
23	social networking (online)	5	0,36	25,21
24	business intelligence	4	0,29	25,50
25	costs	4	0,29	25,79
26	data visualization	4	0,29	26,07
27	decision making process	4	0,29	26,36
28	decision support system (dss)	4	0,29	26,64
29	health care	4	0,29	26,93
30	machine learning	4	0,29	27,21
31	strategic decision making	4	0,29	27,50
32	sustainable development	4	0,29	27,79

From the analysis of frequency of keywords, we analyze the grouping of these by means of the network visualization as shown in Figure 3. From this verification, we can notice the formation of 5 groupings discriminated, these identified with the colors red, blue, green, purple and yellow.
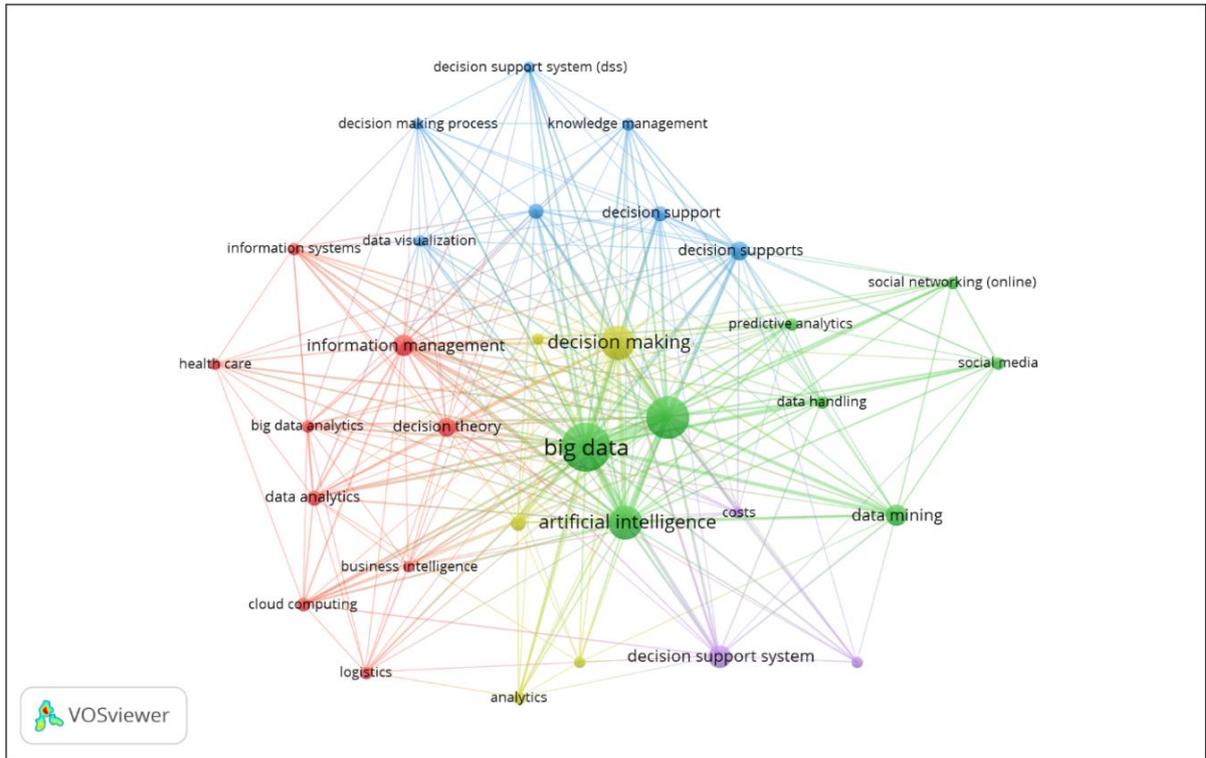
Figure 3. Networking of keywords

After the network visualization, we perform the keyword overlapping visualization, which organizes them according to the frequency identified in a time horizon so that the yellow words are the most recent and as they darken characterize the frequency in an older period (Figure 4).
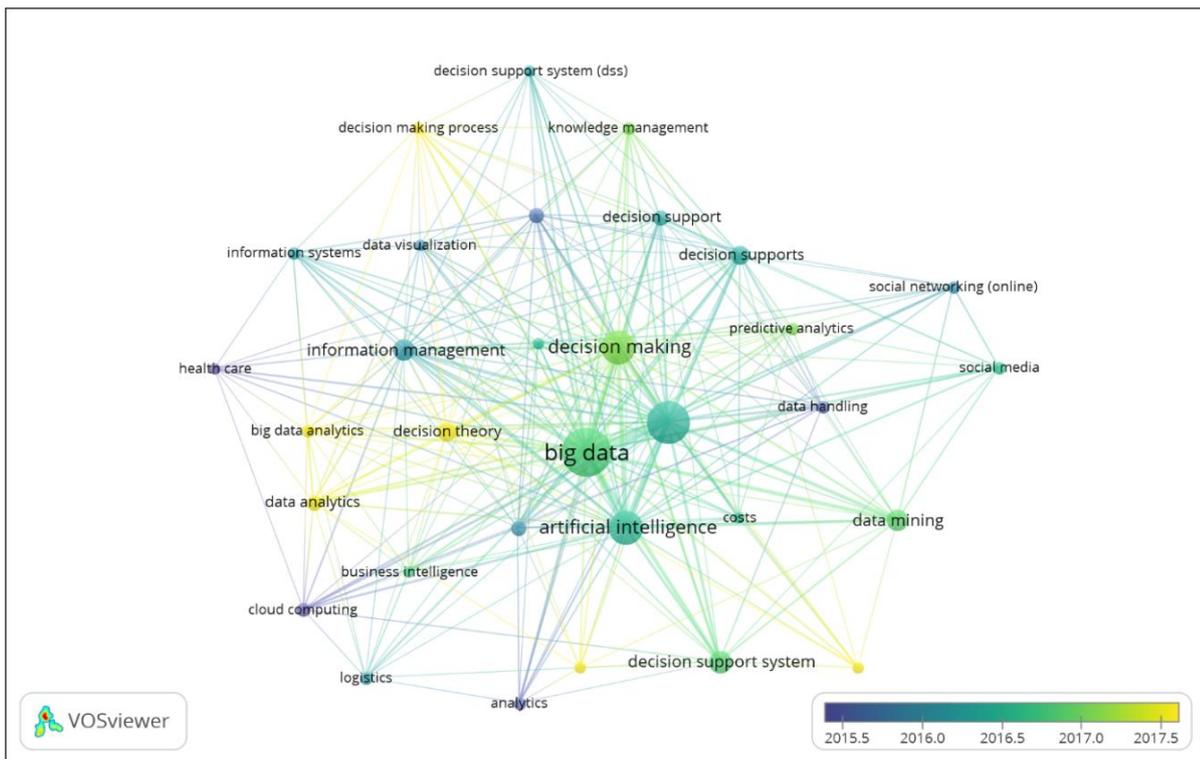

Figure 4. Keyword overlay view

Finally, we perform density analysis, which intensifies color (with warm colors) according to frequency and relation to other words (Figure 5).
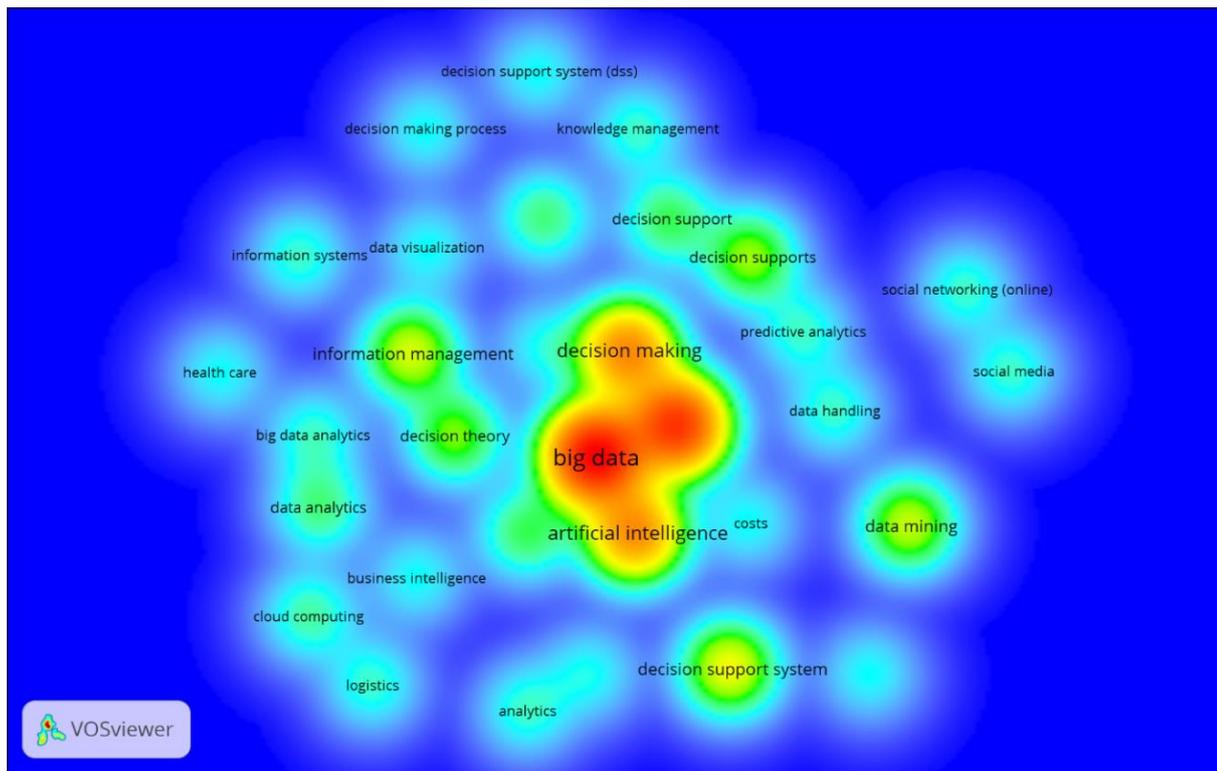

Figure 5. Keyword density view.

From the descriptive analysis we could verify that for the Boolean term used, 5 groups were identified that characterize large areas: the cluster identified in red (Figure 3) suggests studies directed at the areas of health, logistics and IT as well, it is possible to infer that cloud computing and business intelligence technologies and analytical techniques are related to this group that also has the expression related to decision theory as a recurrent.

The cluster identified with the blue color (Figure 3) suggests studies aimed at decision-making systems/processes, knowledge management and data visualization techniques. In this cluster, no specific area was predominant to the point of standing out. The grouping identified with the green color indicates studies aimed at data mining techniques, data handling, artificial intelligence and predictive analysis. The areas that stood out were social media and online social networks.

The small cluster identified with the purple color indicates studies aimed at decision support systems oriented towards sustainable development (word extended for space and omitted by the software). Finally, the cluster identified with the yellow color, also small, indicates studies aimed at strategic decision making via machine learning (word extended for space and omitted by the software).

We could also highlight, from the verification of Figure 4, that recent studies are directed towards sustainable development and machine learning (words below in the image and that were omitted by the software due to the little space for the legend), words that close in the image indicate that studies may even be addressing both subjects simultaneously. On the right side of Figure 4 we can still see that analytical techniques may be discussed simultaneously with decision theory, and above but in a slightly more isolated way in a chronological aspect, decision-making processes are also shown as a recent term in studies.

Finally, through the visualization of keyword density (Figure 5), we noticed a strong density and centralization of the words Big Data, artificial intelligence and decision making, indicating a high capillarity in the studies in general.

CONCLUSION

This article aimed to identify the relationships between Big Data and DSS; and to achieve this objective a search in the Scopus database was carried out and interpreted using the VOSviewer software. Initially, we analyzed the frequency of publications of the research carried out, as well as journals with the respective frequency of published works, enabling us to understand the scope of the main journals and thereby relate to the published works. As a result, we have seen exponential growth since 2014 and highlight "Journal of Decision Systems" and "Decision Support Systems" as the journals with the most published work.

Next, we identified 5 groups of keywords that suggest different areas of study (e.g. logistics, health and social media), as well as a more recent focus on studies aimed at sustainable development, machine learning, analytical techniques and decision-making processes decision. An important contribution that should also be highlighted was the strong relationship between the keywords Big Data, artificial intelligence and decision making, suggesting studies involving the three terms in a large number of works.

As a limitation to the study we can highlight that the combination of the term Boolean directs the frequency of keywords extracted from the database, so a different combination can result in different relationships between the keywords, so we cannot gauge that the results found in this research truly reflect the relationships between surveys. To mitigate this uncertainty, as a suggestion of future studies, we propose an extension to this preliminary bibliometric analysis by means of citation and co-citation analyzes of the works.

Citations, according to Zupic and Čater (2015) are used as a measure of influence, so that if an article is very quoted it is considered important based on the assumption that the authors cite documents that they consider important for their work. Co-citations, defined as the frequency with which two units are cited together (Small, 1993) indicate that the more items are quoted together, the more likely the content is related to them (Zupic & Čater, 2015). Thus, it would be possible to support the findings of this work, to relate relevant authors to the identified areas and also to condition a systematic review of the literature for a deepening of the identified factors, as suggested by (Zupic & Čater, 2015).

REFERENCES

BURATTIN, A., CIMITILE, M., MAGGI, F. M., & SPERDUTI, A.. Online Discovery of Declarative Process Models from Event Streams. *IEEE Transactions on Services Computing,* 2015, *8*(6), 833–846. doi:10.1109/tsc.2015.2459703

CHEN, M., MAO, S., & LIU, Y. Big Data: A Survey. *Mobile Netw Appl,* 2014, *19*, 171–209.

CHEN, C. L. P., & ZHANG, C. Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences, 2014*, *275*, 314–347. doi:10.1016/j.ins.2014.01.015

CONSTANTIOU, I. D., & KALLINIKOS, J. New games, new rules: big data and the changing context of strategy. *Journal of Information Technology,* 2014, *30*(1), 44–57. doi:10.1057/jit.2014.17

DAVENPORT, T., & DYCHE, J. Big Data in Big Companies. *International Institute for Analytics*, 2013, 1–31.

FAYYAD, U., PIATETSKY-SHAPIRO, G., & SMYTH, P. From Data Mining to Knowledge Discovery in Databases. AI Magazine, 1996, 17(3).

FREDRIKSSON, C. Big data creating new knowledge as support in decision-making: practical examples of big data use and consequences of using big data as decision support. *Journal of Decision Systems,* 2018, *27*(1), 1–18. DOI:10.1080/12460125.2018.1459068

GANTZ, J., & REINSEL, D. *Extracting Value from Chaos*. IDC IView, 2011.

GUPTA, S., CHEN, H., HAZEN, B. T., KAUR, S., GONZALEZ, E. D. R. S. Circular economy and big data analytics: A stakeholder perspective. *Technological Forecasting and Social Change,* 2019, vol. 144, issue C, 466-474. DOI:10.1016/j.techfore.2018.06.030

HASHEM, I. A. T., YAQOOB, I., ANUAR, N. B., MOKHTAR, S., GANI, A., KHAN, S.U. The rise of "big data" on cloud computing: Review and open research issues. *Information Systems,* 2015, *47*, 98–115. DOI:10.1016/j.is.2014.07.006

HUANG, S., & CHAOVALITWONGSE, W. A. Computational Optimization and Statistical Methods for Big Data Analytics: Applications in Neuroimaging. *The Operations Research Revolution*, 2015, 71–88. DOI:10.1287/educ.2015.0135

HURWITZ, J., & KIRSCH, D. *Machine Learning For Dummies*. IBM Limited Edition: John Wiley & Sons, Inc. 2018

LANEY, D. 3D Data Management: Controlling Data Volume, Velocity, and Variety. *Application Delivery Strategies*, 2001, 949:1-4.

MANYIKA, J., CHUI, M., BROWN, B., BUGHIN, J., DOBBS, R., ROXBURGH, C., & BYERS, A.H. *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute: Washington, Dc, USA. 2011.

MORTENSON, M. J., DOHERTY, N. F., & ROBINSON, S. Operational research from Taylorism to Terabytes: A research agenda for the analytics age. *European Journal of Operational Research, 2015, 241*(3), 583–595. DOI:10.1016/j.ejor.2014.08.029

POWER, D. J. "Big Brother" can watch us. *Journal of Decision Systems, 2016*, *25*, 578 – 588.

PROVOST, F., & FAWCETT, T. Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data,* 2013 (1), 51–59. DOI:10.1089/big.2013.1508

RAGHUPATHI, W., & RAGHUPATHI, V. Big Data analytics in healthcare: promise and potential. *Health Information Science and Systems,* 2014 *2*(1). DOI: 10.1186/2047-2501-2-3

SMALL, H. Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science,* 1973, *24*(4), 265–269. doi:10.1002/asi.4630240406

VAN ECK, N. J., & WALTMAN, L. Citation-based clustering of publications using CitNetExplorer and VOSviewer. *Scientometrics,* 2017*, 111*(2), 1053–1070.   DOI: 10.1007/s11192-017-2300-7

VERA-BAQUERO, A., COLOMO-PALACIOS, R., MOLLOY, O. &, ELBATTAH, M. Business process improvement by means of Big Data based Decision Support Systems: A Case Study on Call Centers. *International Journal of Information Systems and Project Management, 2015, 3*(1), 5-26. DOI: 10.12821/ijispm030101.

WAMBA, S. F., NGAI, E.W.T., RIGGINGS, F.J. AKTER, S. Transforming operations and production management using big data and business analytics: future research directions. *International Journal of Operations and Production Management, 2017, 37*(1):2–9.

WANG, Y., KUNG, L., & BYRD, T. A. Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change,* 2018*, 126*, 3–13. DOI:10.1016/j.techfore.2015.12.019

ZUPIC, I., & ČATER, T. Bibliometric Methods in Management and Organization. *Organizational Research Methods,* 2014, *18*(3), 429–472. DOI: 10.1177/1094428114562629