TRACKING BUSINESS INTELLIGENCE RESEARCH TRENDS: A BIBLIOMETRIC APPROACH

Ardita DORTI & Ömür AKDEMIR
Ankara Yıldırım Beyazıt University, Ankara, Turkey
arditadorti1@gmail.com omurakdemir@gmail.com

ABSTRACT
Business intelligence is considered as a fundamental necessity for any type of organization. The field of business intelligence has shown continuous enhancements and has kept pace with rapidly changing business environments. The purpose of this study is to explore research trends in scientific articles pertaining to business intelligence. The data set was extracted from the Web of Science and bibliometric method was used to analyze keywords. After the identification of research trends based on frequency, co-occurrence analysis was used to construct and visualize a bibliometric network. The results comprise twenty-nine research trends and four visualizations of the network based on frequency of occurrences, average publication year and densities. Practitioners and scholars may benefit from this study to gain an understanding of the past research in business intelligence field.
Keywords: Bibliometric Network, Business Intelligence, Research Trend

Introduction

Business intelligence has a pivotal role in each organization, as it can accomplish various tasks including collecting, storing, analyzing and spreading data. It is valued as a top technology investment in Europe and considered as an important factor to assist organizations in gaining competitive advantage (Derksen & Luftman, 2014; Miller et al., 2006).

Business intelligence has a history of approximately 60 years since it was first mentioned by Hans Peter Luhn (Luhn, 1958). It is premised that the evolution of business intelligence has received support from concepts and developments of management information systems (Rodrigues, 2002). The studies of Scott Morton (1967) and Gerrity (1971) investigated the connection of business intelligence with decision support systems and chronologically by time lastly, it is claimed that enterprise information systems have influenced developments in business intelligence field (Arnott & Pervan, 2005).

According to Pirttimäki (2007), it is difficult to reach a universal definition for business intelligence. However, some studies define it in the line of transforming data to information and then knowledge (Vizgaitytė & Skyrius, 2012; Stylianou et al., 2013). The definition of Vercellis (2009) was used as the base for this study and it is stated: “set of mathematical models and analysis methodologies that exploits the available data to generate information and knowledge useful for complex decision making processes”.

As claimed by Parker, Benson and Trainor (1988) classical analyzing techniques of cost-benefit are not the most proper ones for evaluation of investments in information technology. A shift in paradigm for evaluation techniques of information technologies was represented by Saaty (1998) and which gives more importance to issues different than financial such as intangibles.

Previous similar studies have conducted bibliometric methodology in various approaches such as the most productive nations or cities, distribution of research areas and publications, the most influential publications and trends in online databases (Chen et al., 2012; Liang & Liu, 2018; López-Robles et al., 2019; Wyskwarski, 2019; Zou et al., 2019). This study differs from others because it adopts an approach based on basic technological elements of business intelligence. Haustein & Larivière (2015) conclude that tracking of trends is indispensable for understanding the research landscape of any field.

The Study

Quantitative approach and bibliometrics method were adopted for the purpose of this study to identify research trends in scientific articles related to the business intelligence. As stated by Liang and Liu (2018) bibliometrics comprises statistical and quantitative analysis of texts and authors based on frequency, centrality and connection. Despite descriptive statistics, network analyses such as co-word, co-occurrence, co-citation, co-authorship or citation may be used to explore publication trends, impacts of a subject, impact of an academic output, authors network or citation patterns (Liang & Liu, 2018). In this research the focus were the texts composed of keywords, which were analyzed by descriptive statistics based on high frequency and bibliometric network visualization based on co-occurrence links.

The data set was retrieved from the Web of Science Core Collection database. The sample and target population were extracted from the population by applying the purposeful sampling. Creswell (2012) listed reasons like convenience, presence and availability of certain characteristics of participants as cause why researches choose non probability sampling for selecting the sample (Creswell, 2012). The population included 7664 academic outputs that have “business intelligence” terms in all possible searchable fields, all available science citation indexes, all languages and time span from 1945 to 2019. The 1727 academic outputs part of target population extracted from population own these characteristics: scientific articles, Science Citation Index Expanded (SCI-EXPANDED), Social Sciences Citation Index (SSCI), Arts & Humanities Citation Index (A&HCI), Emerging Sources Citation Index (ESCI) and English language. The final stage of purposeful sampling consisted in extracting 198 scientific articles from target population. The scientific articles own the characteristics of target population and at least one of basic technological elements terms combined with “business intelligence” terms in all possible searchable fields. In their study Rikhardsson and Yigitbasioglu (2018) have outlined that the major vendors ((*Gartner, 2017. Magic Quadrant for Business Intelligence and Analytics Platforms*, n.d.; Elliot & Woodward, 2015; Howson & Arnold, 2013; Troyansky et al., 2015; Volitich & Ruppert, 2012) and academic literature (Chae & Olson, 2013; Chaudhuri et al., 2011; C Howson, 2013; Cindi Howson, 2007; Sheikh, 2013) identify data analyses, data management, infrastructure and information delivery as basic technological elements of business intelligence and analytics. The set of rules of Web of Science for issues like grammar, synonyms, lemmatization and stemming are taken into consideration for all search queries (*Web of Science Core Collection Help*, n.d.). The Figure 1 represents the search queries of population, target population and sample.

Figure 1: The Search Queries for Population, Target Population and Sample

ALL=("business intelligence") *Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI Timespan=1945-2019*	Population
(ALL=("business intelligence")) *AND* LANGUAGE: (English) *AND* DOCUMENT TYPES: (Article) *Indexes=SCI-EXPANDED, SSCI, A&HCI Timespan=1945-2019*	Target Population
((ALL=(("business intelligence") AND ("data management"OR"data analyses"OR"data analysis"OR"information delivery"OR"information deliveries"OR"infrastructure"OR"infrastructures")))) *AND* LANGUAGE: (English) *AND* DOCUMENT TYPES: (Article) *Indexes=SCI-EXPANDED, SSCI, A&HCI Timespan=1945-2019*	Sample

In order to analyze the data set of sample VOSviewer 1.6.13 software tool and co-occurrence analysis were used to construct and visualize bibliometric network of keywords. Compared to other computer programs that serve for bibliometric mapping, VOSviewer graphical representation is considered as more advanced (Van Eck & Waltman, 2010).

Initially, the bibliometric network was constructed that means places of items and links with each other are determined. Items may obtain various attributes in VOSviewer and some of which are weight, score and cluster number attributes (Van Eck & Waltman, 2019). In accordance with data set and research aim of this study and among other custom weight attributes, occurrences attribute was selected. The data set of sample amounts to 1115 keywords in total from author keywords or keywords plus of scientific articles. The keywords with at least 5 occurrences (by default value) and the highest occurrences are selected by VOS text mining methodology that performs part-of-speech tagging and calculation of distributions (Van Eck & Waltman, 2016). With regard to standard weight attributes, link strength represents the number of scientific articles where two keywords have co-occurred and total link strength represents total number of scientific articles where a keyword has co-occurred with other keywords of the network. The strength of links between items in a network gained from co-occurrence analysis depends on the counting method applied (Perianes-Rodriguez et al., 2016) and in this study full counting was applied to construct the bibliometric network. Average publication year score attribute was selected and its values for each keyword indicated colors of respective items in overlay visualization. Cluster number attribute represents the number of the cluster where each keyword belongs, based on the calculations of VOS unified mapping and clustering technique (Van Eck et al., 2010).

The steps that follow network construction are listed in this order normalization, mapping, clustering and visualizing. In pursuance of normalizing the differences between items of the network, VOSviewer applies association strength. The calculations of association strength deploy the numbers of links of each item, links between two items and all links in the network (Waltman et al., 2010). The unified approach of mapping and clustering executes the logic that the higher similarity between items the smaller distance and based on overall calculations and resolution parameter locates items into clusters(Van Eck & Waltman, 2010, 2014).

Networks can be illustrated with three different types of visualizations in VOSviewer software tool. Network visualization presents the items based on custom and standard weight attributes which respectively indicate the size of labels and circles and the thickness of lines. The color of an item in network visualization depends on the cluster where it belongs, each cluster has a different color. Related to label and circle size, thickness of lines and distance between items, network visualization and overlay visualization are undifferentiated. Each overlay visualization is accompanied by a color bar which range the spectrum of colors from lower to higher score value. Density visualization is divided into two subtypes item and cluster, which follow the same calculations but they differentiate in the concept of neighborhood. Item density visualization considers as neighborhood all the items around the referred item while cluster density visualization considers as neighborhood all the items around and that belong to the same cluster with the referred item. Each item density is calculated based on kernel width parameter, items number and weight in the neighborhood and weight of item itself. The color of an item in density visualization changes from prominent to pale as item density decreases (Van Eck & Waltman, 2019).

Findings

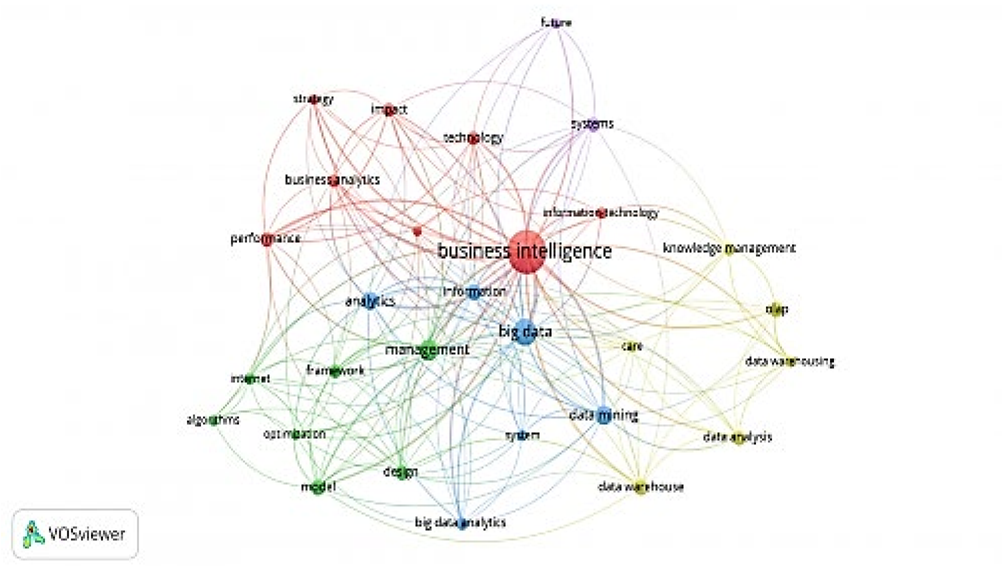
The high frequency of a keyword is a crucial point in this study because it indicates the prominence given by research to a particular topic (keyword). The keywords with high frequency of occurrences can be called research trends. The results, obtained from usage of VOSviewer text mining technique and descriptive statistics in order to select the keywords with the highest frequencies of occurrences, are presented in Table 1. The keywords in Table 1 are listed in descending order.

Table 1: Frequency of Occurrences of Keywords

Item	Occurrences
business intelligence	74
big data	28
management	18
data mining	15
analytics	13
olap	11
model	11
information	10
data warehouse	10
systems	9
performance	9
design	9
data analysis	9
technology	8
impact	8
framework	8
knowledge management	7
business analytics	7
big data analytics	7
system	6
internet	6
information-technology	6
data warehousing	6
care	6
strategy	5
optimization	5
future	5
cloud computing	5
algorithms	5

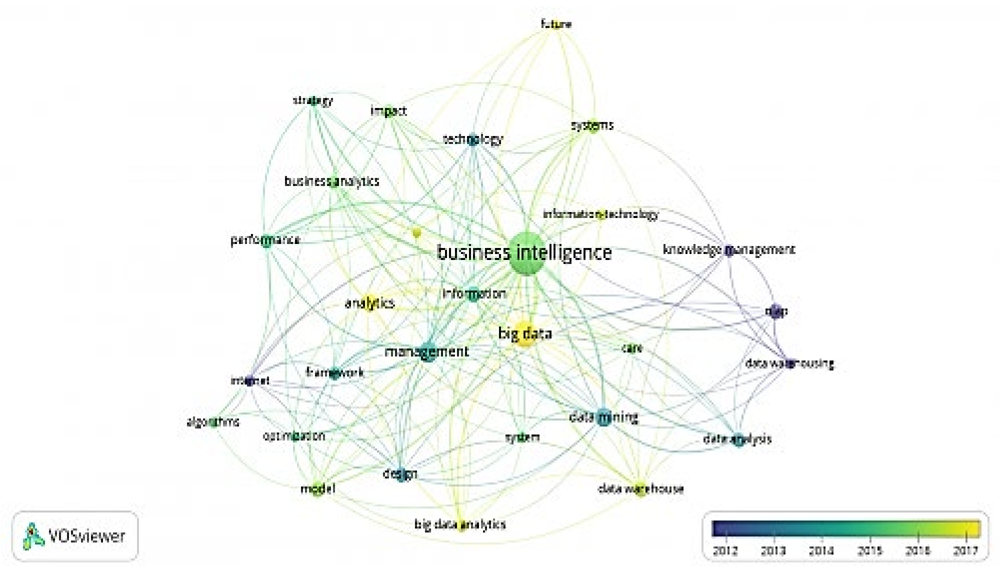
Network visualization of this study, which is illustrated in Figure 2, has 29 items (keywords) connected with each other by 178 links with an overall link strength of 374 and divided into 5 clusters. The keywords, which fulfill the condition that the frequency of occurrences (of a keyword) is equal to or higher than the minimum default value of occurrences, are part of the bibliometric network as items. Network visualization illustrate the items(keywords) in the way that the smaller the distance between two items means higher relevance to each other in terms of co-occurrence links; the thicker the lines between two items means the higher number of scientific articles where two keywords have co-occurred; the bigger the size of label and circle means higher frequency of occurrences of a particular keyword. The range of colors in network visualization reveal the number of clusters and the higher the resolution parameter the higher the number of clusters.

Figure 2: Network Visualization



In the overlay visualization illustrated in Figure 3, the size of labels and circles of items, the thickness of the lines that connect them and distance in between are exactly the same as network visualization in Figure 2. Score attribute range between 2009 and 2017 related to the average publication year of each item (keyword). The color of circles and lines change from blue to green and then yellow as the average publication year increases.

Figure 3: Overlay Visualization



Density visualizations were illustrated based on each item density which was calculated with kernel width parameter of 1.2. In Figure 4 item density visualization is presented and where the density of an item relies on all the items around it. In Figure 5 cluster density visualization is presented and where the density of an item relies on the items around that are also part of the same cluster. In both visualizations the color of items changes from pale to prominent as item density increases.

Figure 4: Item Density Visualization

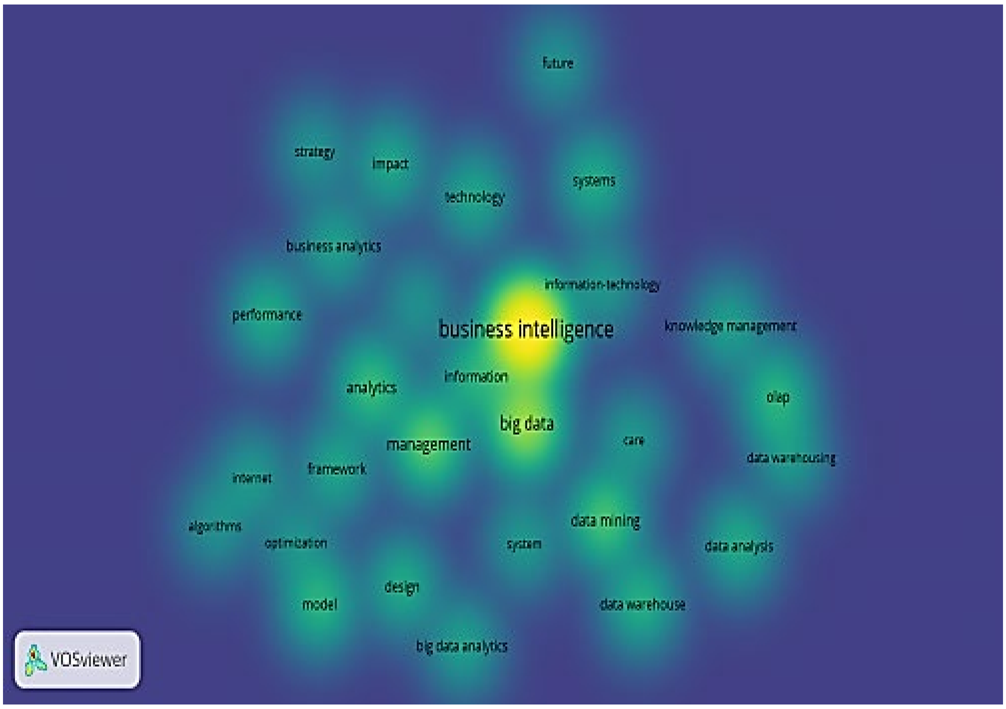
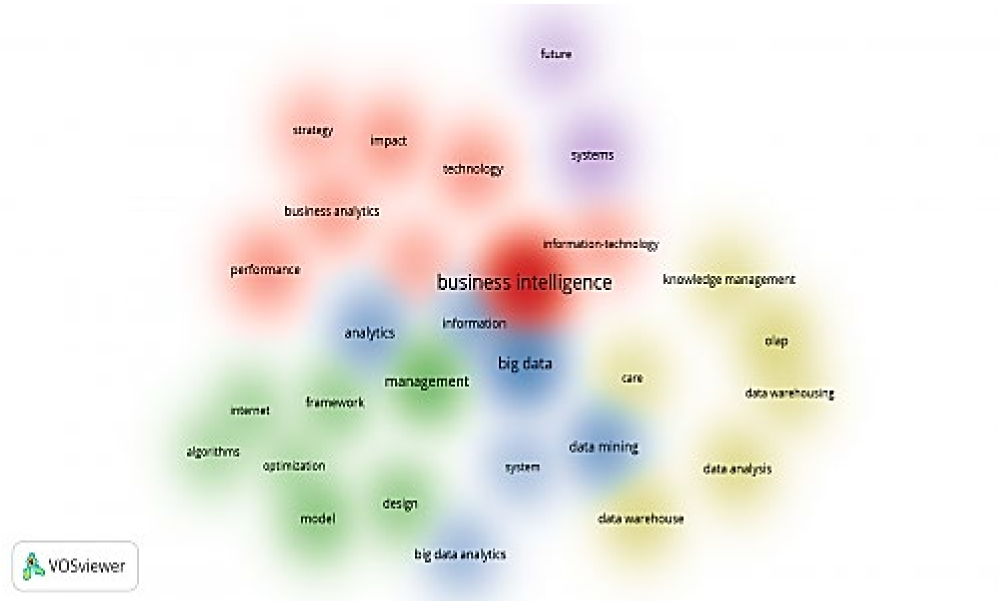


Figure 5: Cluster Density Visualization



Conclusion and Suggestions

Business intelligence is evaluated as an emerging field (Derksen & Luftman, 2014; Miller et al., 2006). Thus the tracking of its research trends may be essential because it provides insights into the list of sub-fields that has gained more attention. This study employed bibliometric approach and co-occurrence analysis associated with scientific articles to explore research trends and co-occurrence links between them. The results consist of 29 research trends which are connected by 178 co-occurrence links with an overall link strength of 374. The research trends are separated into 5 clusters and illustrated in network, overlay, item density and cluster density visualizations, the colors of which are indicated respectively by cluster parameter, average publication year item density and kernel width parameter. The aim of this research was fulfilled because the results of analysis are all terms related to the business intelligence field. The first research trend ranked in terms of frequency of occurrences is business intelligence, followed by big data and management. The other research trends don't have big difference among each other in terms of frequency of occurrences and they are sorted from highest to lowest based on their frequency of occurrences are as follows: data mining, analytics, olap, model, information, data warehouse, systems, performance, design, data analysis, technology, impact, framework, knowledge management, business analytics, big data analytics, system, internet, information-technology, data warehousing, care, strategy, optimization, future. Both scholars and practitioners may benefit from this study in order to understand trends up to now and can plan their future research questions based on the trends in the business intelligence field.

References

Arnott, D., & Pervan, G. (2005). A critical analysis of decision support systems research. *Journal of Information Technology*, *20*(2), 67–87. https://doi.org/10.1057/palgrave.jit.2000035

Chae, B. K., & Olson, D. L. (2013). Business analytics for supply chain: A dynamic-capabilities framework. *International Journal of Information Technology and Decision Making*, *12*(1), 9–26. https://doi.org/10.1142/S0219622013500016

Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An overview of business intelligence technology. *Communications of the ACM*, *54*(8), 88–98. https://doi.org/10.1145/1978542.1978562

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data To Big Impact. *MIS Quarterly*, *36*(4), 1165–1188. www.freakonomics.com/2008/02/25/hal-varian-answers-your-questions/

Creswell, J. W. (2012). *Educational Research* (pp. 141–146). Pearson.

Derksen, B., & Luftman, J. (2014). European key IT and management issues & trends for 2014. In *1*.

Elliot, A., & Woodward, W. (2015). *SAS Essentials: Mastering SAS for Data Analytics*. Wiley.

Gartner, 2017. Magic Quadrant for Business Intelligence and Analytics Platforms. (n.d.). Retrieved April 2, 2020, from https://www.gartner.com/en/documents/3611117

Haustein, S., & Larivière, V. (2015). The Use of Bibliometrics for Assessing Research: Possibilities, Limitations and Adverse Effects. In *Incentives and Performance* (pp. 121–139). Springer, Cham.

Howson, C. (2013). *Successful Business Intelligence: Unlock the Value of BI & Big Data*. McGraw-Hill.

Howson, C, & Arnold, E. (2013). *SAP Business Objects BI&A 4.0 The Complete Reference* (3rd ed.). McGraw-Hill.

Howson, Cindi. (2007). *Successful Business Intelligence: Secrets to Making BI a Killer App [Hardcover]* (p. 244). McGraw-Hill. https://doi.org/10.1036/0071498516

Liang, T. P., & Liu, Y. H. (2018). Research Landscape of Business Intelligence and Big Data analytics: A bibliometrics study. *Expert Systems with Applications*, *111*(128), 2–10. https://doi.org/10.1016/j.eswa.2018.05.018

López-Robles, J. R., Otegi-Olaso, J. R., Porto Gómez, I., & Cobo, M. J. (2019). 30 years of intelligence models in management and business: A bibliometric review. *International Journal of Information Management*, *48*, 22–38. https://doi.org/10.1016/j.ijinfomgt.2019.01.013

Luhn, H. P. (1958). A Business Intelligence System. *BM Journal of Research and Development*, *2*(4), 314–319. https://doi.org/doi:10.1147/rd.24.0314

Miller, G. J., Brautigam, D., & Gerlach, S. V. (2006). *Business Intelligence Competency Centers: A Team Approach to Maximizing Competitive Advantage.* (p. 3). SAS Institute Inc.

Parker, M. M., Benson, R. J., & Trainor, H. E. (1988). *Information economics: Linking business performance to information technology*. Prentice-Hall.

Perianes-Rodriguez, A., Waltman, L., & Van Eck, N. J. (2016). Constructing bibliometric networks: A comparison between full and fractional counting. *Journal of Informetrics*, *10*(4), 1178–1195. https://doi.org/10.1016/j.joi.2016.10.006

Pirttimäki, V. (2007). *Business Intelligence as a Managerial Tool in Large Finnish Companies.* Tampere University of Technology.

Rodrigues, L. C. (2002). Business intelligence: The management information system next step. *Management Information Systems*, 269–278.

Saaty, T. L. (1998). Reflections and Projections on Creativity in Operations Research and Management Science: A Pressing Need for a Shift in Paradigm. *Operations Research*, *46*, 9–17.

Sheikh, N. (2013). *Implementing Analytics: A Blueprint for Design, Development, and Adoption*. Morgan Kaufmann.

Stylianou, V., Savva, A., & Spyrou, S. (2013). Business Intelligence Made Simple. In *Innovations and Advances in Computer, Information, Systems Sciences, and Engineering* (pp. 1001–1012).
Troyansky, O., Gibson, T., Leichtweiss, C., & Bjork, L. (2015). *QlikView Your Business: An Expert Guide to Business Discovery With QlikView and Qlik*. Wiley.
Van Eck, N. J., & Waltman, L. (2010). Software survey : VOSviewer , a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3
Van Eck, N. J., & Waltman, L. (2014). Visualizing bibliometric networks. In *Measuring scholarly impact: Methods and practice* (pp. 285–320). Springer.
Van Eck, N. J., & Waltman, L. (2016). Text Mining and Visualization using VOSviewer. *ISSI Newsletter*, *7*(3), 50–54. https://doi.org/10.1201/b19007
Van Eck, N. J., & Waltman, L. (2019). *VOSviewer Manual*.
Van Eck, N. J., Waltman, L., Dekker, R., & Van Den Berg, J. (2010). A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS. *Journal of the American Society for Information Science and Technology*, *61*(12), 2405–2416. https://doi.org/10.1002/asi.21421
Vercellis, C. (2009). Business Intelligence. In *Business Intelligence: Data Mining and Optimization for Decision Making* (p. 6). Wiley. https://doi.org/10.1002/9780470753866
Vizgaitytė, G., & Skyrius, R. (2012). Business Intelligence in the Process of Decision Making : Changes and Trends. *Ekonomika/Economics*, *91*(3), 147–157.
Volitich, D., & Ruppert, G. (2012). *IBM Cognos Business Intelligence 10: The Official Guide*. McGraw-Hill.
Waltman, L., Van Eck, N. J., & Noyons, E. C. M. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics*, *4*(4), 629–635. https://doi.org/10.1016/j.joi.2010.07.002
Web of Science Core Collection Help. (n.d.). Retrieved April 2, 2020, from https://images.webofknowledge.com/WOKRS534DR2/help/WOS/hs_allfields.html
Wyskwarski, M. (2019). *Business intelligence – Publication analysis using the R language*. *137*.
Zou, Z., Cheng, J., Huang, K., & Gui, Y. (2019). Research on the Developments of Business Intelligence and its Enlightenment based on Bibliometric Statistics and Knowledge Map Analysis. *Journal of Physics: Conference Series*, *1176*(4). https://doi.org/10.1088/1742-6596/1176/4/042089

