Annals of Operations Research

Big Data and Supply Chain Management: A Review and Bibliometric Analysis

--Manuscript Draft--

Manuscript Number:	ANOR-D-16-00312R1
Full Title:	Big Data and Supply Chain Management: A Review and Bibliometric Analysis
Article Type:	S.I. : Big Data Analytics in Operations & Supply Chain Management
Keywords:	Big Data; Supply chain management; Bibliometric analysis; Network analysis
Corresponding Author:	Thanos Papadopoulos, PhD University of Kent Chatham, Kent, UNITED KINGDOM
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University of Kent
Corresponding Author's Secondary Institution:	
First Author:	Deepa Mishra, PhD
First Author Secondary Information:	
Order of Authors:	Deepa Mishra, PhD
	Angappa (Guna) Gunasekaran, PhD
	Thanos Papadopoulos, PhD
	Stephen J Childe, PhD
Order of Authors Secondary Information:	
Funding Information:	

Abstract:	As Big Data has undergone a transition from being an emerging topic to a growing research area, it has become necessary to classify the different types of research and examine the general trends of this research area. This should allow the potential research areas that for future investigation to be identified. This paper reviews the literature on 'Big Data and supply chain management (SCM)', dating back to 2006 and provides a thorough insight into the field by using the techniques of bibliometric and network analyses. We evaluate 286 articles published in the past 10 years and identify the top contributing authors, countries and key research topics. Furthermore, we obtain and compare the most influential works based on citations and PageRank. Finally, we identify and propose six research clusters in which scholars could be encouraged to expand Big Data research in SCM. We contribute to the literature on Big Data by discussing the challenges of current research, but more importantly, by identifying and proposing these six research clusters and future research directions. Finally, we offer to managers different schools of thought to enable them to harness the benefits from using Big Data and analytics for SCM in their everyday work.

Big Data and Supply Chain Management: A Review and Bibliometric Analysis

Deepa Mishra
Department of Industrial and Management Engineering
IIT Kanpur
Kanpur, 208016
India
Cell. No. +917753920946
E-mail: dmishra@iitk.ac.in

Angappa Gunasekaran
Charlton College of Business
University of Massachusetts Dartmouth
North Dartmouth, MA 02747-2300
USA
Tel: (508) 999-9187
Fax: (508) 999-8646
E-mail: agunasekaran@umassd.edu

Thanos Papadopoulos*
Kent Business School, University of Kent.
Sail and Colour Loft
The Historic Dockyard
Chatham, Kent ME4 4TE

United Kingdom
Tel: +44 1634 88 8494
E-mail: A.Papadopoulos@kent.ac.uk

Stephen J. Childe
Plymouth Business School
Plymouth University
Plymouth, PL4 8AA
United Kingdom
E-mail: stephen.childe@plymouth.ac.uk

*Corresponding Author

Big Data and Supply Chain Management: A Review and Bibliometric Analysis

Abstract

As Big Data has undergone a transition from being an emerging topic to a growing research area, it has become necessary to classify the different types of research and examine the general trends of this research area. This should allow the potential research areas that for future investigation to be identified. This paper reviews the literature on 'Big Data and supply chain management (SCM)', dating back to 2006 and provides a thorough insight into the field by using the techniques of bibliometric and network analyses. We evaluate 286 articles published in the past 10 years and identify the top contributing authors, countries and key research topics. Furthermore, we obtain and compare the most influential works based on citations and PageRank. Finally, we identify and propose six research clusters in which scholars could be encouraged to expand Big Data research in SCM. We contribute to the literature on Big Data by discussing the challenges of current research, but more importantly, by identifying and proposing these six research clusters and future research directions. Finally, we offer to managers different schools of thought to enable them to harness the benefits from using Big Data and analytics for SCM in their everyday work.

Keywords: Big Data; Supply chain management; Bibliometric analysis; Network analysis

1. Introduction

What is Big Data? And why is it significant for academics and professionals to study this concept? There are several definitions of Big Data which might not be universally accepted (Mayer-Schonberger and Cukier 2013; Song et al. 2016). As the name itself suggests, ‘size’ was conceived as its main characteristic. But later on, Gartner Inc. observed that size may not be the only criterion to adjudge ‘data’ as ‘Big Data’. Big Data has been identified by both Gobble (2013) and Strawn (2012) as being very important for innovation the “fourth paradigm of science” (p.34) respectively. According to McKinsey & Co., Big Data is “the next frontier for innovation, competition and productivity”. McAfee and Brynjolfsson (2012) viewed Big Data as an approach that transforms decision making processes by enhancing the visibility of firms’ operations and improving the performance measurement mechanisms. In this regard, Brown et al. (2011) claimed that the logic behind these facts lies in the capability of ‘Big Data’ to change competition by “transforming processes, altering corporate ecosystems, and facilitating innovation” (p.3). Not only does Big Data influence competition and growth for individual companies, but it also enhances productivity, innovation, and competitiveness for different sectors and economies.

Hence, the study of Big Data is significant because Big Data has the ability to transform entire business processes. A firm's competitive advantage could depend on its ability to extract Big Data and analyse it to gain business insights (Wong 2012) and outperform its competitors (Oh et al. 2012). In this regard, McKinsey and Company claimed that "collecting, storing, and mining Big Data for insights can create significant value for the world economy, enhancing the productivity and competitiveness of companies and the public sector and creating a substantial economic surplus for consumers " (Manyika et al. 2011: p. 1). It has also been pointed out (Bozarth et al 1998; Tsai et al 2013) that firms can identify the preferences and needs of customers by taking advantage of Big Data derived from loyalty cards and social media. With regards to social media, a leading eyeglasses manufacturer, SPEC, collects and analyses Big Data from social media (i.e., tweets, Google, Facebook, etc.) to generate new product ideas (Tan et al., 2015). Further, Thibeault and Wadsworth (2014) noted that on Facebook, around 10 billion messages including photos and videos are sent per day, the "share" button is clicked 4.5 billion times and 350 million new pictures are uploaded each and every day. By utilizing the hidden value of Big Data, retailers can increase their operating margins by 60 percent (Werdigier 2009). While huge assets and time are invested in creating Big Data platforms and technologies, it offers extensive long-term benefits related to the achievement of competitive advantage (Terziovski 2010).

Big Data has potential value that is yet to be explored. In 2011, research by Oxford Economics found that only 25% of industry executives were of the belief that in the next five years the manufacturing sector would be highly impacted by digital transformation. Nonetheless, it has been observed that "every manufacturer has an unbelievable amount of data that is never put to use. They are literally drowning in it, and when they begin to gather it, analyse it and tie it to business outcomes, they are amazed by what comes out" (Records and Fisher 2014: p. 2). Manyika et al. (2011) estimated that US health care may benefit by 300 billion dollars a year by using Big Data creatively and effectively to drive efficiency and quality. Exploitation of personal location data across the world can generate a commercial value of 600 billion dollars annually (Davenport and Harris 2007; LaValle et al. 2011). The significance of Big Data can be realized from the fact that it was regarded as the national priority task in supporting healthcare and national security by the White House in 2010 (Mervis 2012).

Applications of Big Data have been seen in diverse fields including medicine, retail, finance, manufacturing, logistics, and telecommunications (Feng et al. 2013). Researchers (Chen et al. 2012; Fosso-Wamba et al. 2015; Dubey et al. 2015; Wang et al. 2016; Song et al. 2016) have endeavoured to explore different dimensions of Big Data and capture the potential benefits to supply chain management (SCM). It is important for supply chain managers to understand the role of Big Data in enhancing the efficiency and profitability of a firm. The senior solutions principals for HCL Retail and CPG Consulting Practice claim the information provided by Big Data can maximize productivity, collaboration, speed and visibility and improve relationships with supply chain stakeholders (SC Digital, 2014). Schoenherr and Speier-Pero (2015) have identified several benefits of Big Data and predictive analytics on supply chain performance.

In recent years, scholars (Sagiroglu and Sinanc 2013; Fosso-Wamba et al. 2015; Gandomi and Haider 2015; Khorheh et al. 2015; Wang et al. 2016; Mishra et al. 2016a, 2016b) have reviewed the literature on Big Data. While these studies have been able to provide insight into the field through structured reviews and classification into future research themes, apart from Mishra et al. (2016a; 2016b), they have not used additional analyses such as bibliometric and network analyses that could help in identifying the established and emerging areas of research. The papers by Mishra and colleagues have used bibliographic and network analyses, but focused on either Internet of Things applications or concepts, trends and challenges of Big Data, rather than upon SCM applications, which is the focus of this paper. Therefore, the analyses in this work are important – no matter if

Big Data is still in its infancy – to provide the reader (i.e. academician and/or practitioner) with an overview of the current state of the field with regards to authors, countries, and topics and areas; and to suggest emerging clusters and encourage researchers towards collaborating and further expansion of the knowledge of the field.

Therefore, to address the SCM area, this paper (i) reviews the literature on 'Big Data and supply chain management', dating back to 2006; (ii) provides a thorough insight into the field by using the technique of bibliometric and network analysis and by evaluating 286 articles published in past 10 years, and identifies top contributing authors, countries and key research topics related to the field; (iv) obtains and compares the most influential works based on citations and PageRank; and (v) identifies and proposes six established and emerging research clusters which would encourage scholars to expand research on Big Data and SCM.

In this study, bibliometric tools were used to thoroughly review the publications on Big Data and SCM. Initially, we obtained 7868 articles which were further filtered to obtain 286 articles containing the most influential works and researchers. The findings of this study offer additional insights on the current state of the field and highlight potential future research directions. In the next section, we review the literature on Big Data and SCM followed by the research methodology. Then, we present a thorough analysis using rigorous bibliometric tools. The paper ends with conclusion, limitations and future research directions.

2. Review of the literature on Big Data and supply chain management

In this section we report on the literature by discussing Big Data and its characteristics followed by its role and applications in SCM.

2.1 Big Data

Although the term 'Big Data' is ubiquitous these days, its origin dates back to mid-1990s. Diebold (2012) noted that the term "Big Data . . . probably originated in lunch-table conversations at Silicon Graphics Inc. (SGI) in the mid-1990s, in which John Mashey figured prominently" (p. 5). The popularity of Big Data can be attributed to the fact that this topic was Google-searched 252,000 times in November 2011 (Flory 2012) and then reached the impressive number of 801,000,000 hits in October 2015 (Mishra et al. 2016b). McKinsey Global Institute (2011) defined Big Data as the "datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyse" (p. 1). This definition is not confined to data size, since data sets will increase in the future. It highlights the necessity of technology to cope up with the rapid growth in available data. Other characteristics have been put forward to define the Big Data concept (Mishra et al. 2016b) and these will be reviewed below.

2.2 Big Data characteristics

Volume reflects the magnitude of data, which has increased drastically in the past few years. The size of Big Data may vary from multiple terabytes to petabytes. Fosso-Wamba et al. (2015) provided a definition of volume as "the large amount of data that either consume huge storage or entail of large number of records data" (p. 3). As the amount of data crossing the internet per second has increased tremendously, firms have an opportunity to work with many petabytes of data in a single dataset. In SCM, high volume data may relate to, for example, data from RFID and other types of sensors used for identification and transportation of products/components, cell phone GPS signals, and purchase transaction records. For example, in Walmart it is estimated that more than 2.5 petabytes of data every hour is collected from customer transactions (McAfee and Brynjolfsson 2012).

Variety refers to the "structural heterogeneity in a dataset" (Gandomi and Haider 2015: p. 138). In the work of Russom (2011a, b), variety in Big Data is defined as when the "data generated from greater variety of sources and formats contain multidimensional data fields" (Fosso-Wamba et al., 2015: p. 3). Firms are using various types of data; structured, semi-structured, and unstructured. Structured data refers to the tabular data available in spreadsheets and amounts to only 5% of all the existing data (Cukier 2010) whereas, unstructured data is more plentiful in the form of text, images, audio, and video. A continuum between these two types of data is referred as semi-structured data which does not follow any particular standards. A classic example of semi-structured data is Extensible Mark-up Language (XML) which is used for exchanging data on the internet. As an example in SCM, Tata motors analyses around 4 million text messages every month ranging from product complaints and service appointment reminders to new product announcements and customer satisfaction surveys (Fosso-Wamba et al. 2015).

Velocity refers to the "rate at which data are generated and the speed at which it should be analysed and acted upon" (Gandomi and Haider 2015: p. 138). Owing to the rapid growth in digitalization, data is getting generated at an exceptional rate which drives the need for real time analytics and evidence based planning. Since conventional data management systems are inefficient to handle large data sets, Big Data technologies act as a safeguard by helping firms in creating real-time intelligence from high volumes of perishable data (Gandomi and Haider 2015). An SCM example is Amazon that manages every day a constant flux of products, suppliers, customers, and promotions while being dependable at the same time (Fosso-Wamba et al. 2015).

Besides the "3Vs", three other characteristics, that is, veracity, variability and value have been introduced. *Veracity,* known as the fourth V, reflects the "unreliability inherent in some sources of data" (Gandomi and Haider 2015: p. 139). White (2012) suggests that veracity deals with data quality and its importance, as well as the level of trust accorded to a source of data.

Variability (and *Complexity*) are the two dimensions of Big Data which were introduced by Statistical Analysis Software (SAS). Usually, the velocity of Big Data is inconsistent and has variation in data flow rates, termed as 'variability' of Big Data (Gandomi and Haider 2015). Related to this, Complexity arises when the Big Data comes from innumerable sources. Thus, there exists a need to connect, match, cleanse and transform data received from these sources (Gandomi and Haider 2015). For instance, in the previous example of Amazon, the company needs to understand (in order to deal with *veracity*) and cleanse the data (in order to deal with *variability*) in order to make sense out of it. IBM has also reported that data quality is important for Big Data since the inherent unpredictability and complexity of data cannot be removed by even the best data cleansing methods. *Value* reflects the economic benefits from Big Data (Forrester 2012; Oracle 2012). It is important for firms to acknowledge the substantial amount of data and from this data, what is meaningful to be extracted for further analysis. In an SCM example, Tesco has increased their operating margins while analysing Big Data that related to temperature and weather patterns, thereby conducting better forecasts of temperatures and associated changes in consumer demand (Patil, 2014).

2.3 Big Data and supply chain management

Big Data in forms such as data from social media and networking applications, is widely used in business and marketing. However, research evaluating its role, usage, and potential in SCM seems to be lagging behind (Casemore 2012; O'Leary 2011). Several studies (Chae and Olson 2013; Hazen et al. 2014; Trkman et al. 2010) have been conducted on the use of data and analytical capabilities for SCM, focusing mainly on the application and impact of traditional data sources and analytical techniques in supply chain planning and execution. There have also been calls for researchers to consider the use of Big Data in the field of SCM (Huang et al. 2014; Waller and Fawcett 2013).

The significance of data analytics for SCM was highlighted by Waller and Fawcett (2013) who defined 'SCM data science' as the "application of quantitative and qualitative methods from a variety of disciplines in combination with SCM theory to solve relevant SCM problems and predict outcomes, taking into account data quality and availability issues" (p. 79). In their study on Big Data analytics (BDA), Bi and Cochran (2014) discussed the impact of Big Data on manufacturing information systems and identified BDA as critical to data acquisition, storage, and analytics in modern manufacturing. In addition, the problem of data quality in SCM was studied by Hazen et al. (2014) who emphasized that it is crucial to monitor and control the quality of data in supply chain processes. They also noted that "supply chain professionals are inundated with data, motivating new ways of thinking about how data are produced, organized, and analysed. This has provided an impetus for organizations to adopt and perfect data analytic functions (e.g. data science, predictive analytics, and Big Data) in order to enhance supply chain processes and, ultimately, performance" (p. 72). Recently, Chae (2015) noted that Big Data and social media have not been thoroughly examined in the field of SCM. Hence, Chae proposed an analytical framework through which supply chain tweets can be analysed, the current usage of Twitter in the context of supply chain can be examined and the potential role of Twitter in supply chain research can be explored. It has also been argued that the competition is no longer between firms, but between entire supply chains (see for example Craighead et al. 2009; Ketchen and Hult 2007; Slone 2004; Whipple and Frankel 2000). As an outcome of this increasing attention on SCM, managers are now forced to reassess their competitive strategies (Zacharia et al. 2011). Since both technology and data are available, it is important for companies to decide how to use them to win (Hopkins et al. 2010). Supply chain managers are getting increasingly dependent upon data for gaining visibility on expenditure, identifying trends in costs and performance, and for supporting process control, inventory monitoring, production optimization, and process improvement efforts. As a matter of fact, there are several companies that are flooded with data and try to capitalize on data analysis in an attempt to gain competitive advantage (Davenport 2006). Having an ability to exploit data, firms such as Google, Amazon outperform their competitors by developing potential business models. Barton and Court (2012) highlighted that through Big Data, firms can change the way they do business and deliver performance gains similar to the ones achieved in 1990s when companies redesigned their core processes. They also pointed out that the adoption of data-driven strategies will soon become a significant point of competitive differentiation. McAfee and Brynjolfsson (2012) observed that productivity rates and profitability of companies can be enhanced by 5% to 6%, if they incorporate Big Data and analytics into their operations.

3. Research methodology and data statistics

Rowley and Slack (2004) proposed a five step methodology to carry out a literature review, which includes scanning documents, making notes, structuring the literature review, writing the literature review, and building the bibliography. In this study, we adopted a similar five step literature review process to identify the influential works, ascertain the recent areas of research and offer insights into current research interests and directions for future research in the field.

3.1 Defining keywords

While selecting keywords for this study and to ensure that the topic of the study was fully captured, we used Big Data and supply chain as the two major keywords for data collection. Two combinations were made: (1) Big Data and (2) Big Data AND Supply Chain.

3.2 Initial results

The data was collected from Scopus database only since it is the largest abstract and citation database of over 20,000 peer-reviewed journals belonging to publishing houses, namely, Elsevier, Emerald, Informs, Taylor and Francis, Springer and Inderscience, and covering fields of science, technology, medicine, social sciences, and arts and humanities (Fahimnia et al. 2015). On comparing Scopus and Web-of-Science (WoS) databases, Yong-Hak (2013) observed that Scopus database is the more comprehensive since WoS includes only ISI indexed journals, only 12,000 titles.

We searched for the aforementioned keywords in "title, abstract, keywords" of articles belonging to Scopus database. The initial search resulted in 7868 articles. When using "Big Data" as a keyword, the research yielded 6534 articles, whereas when using "Big Data and supply chain" 1334 articles. These results contain information about the title of the paper, author names and affiliations, abstract, keywords and references which were then saved in RIS format.

3.3 Refining the initial results

To refine our search results, we removed the duplicates as few papers may be present in more than one combination of keywords. On eliminating them, 5486 papers were left. Since Rodriguez and Navarro(2004) categorised articles and reviews as certified knowledge, we restricted ourselves to only scientific publications (articles and reviews) which appeared in peer-reviewed journals (p. 982)

Unpublished articles, working papers and magazine articles were excluded during the data purification process. This search resulted in 2564 relevant documents, published during a 10-year period i.e. 2006-2016. For using “Big Data”, the search yielded 1659 articles, whereas for “Big Data and supply chain” 905 articles. These refinements in the RIS file were made by using Endnote bibliography software, and the final RIS data file was stored for future analysis.

3.4 Initial data statistics

In order to understand the role of the different journals, we identified the top 20 journals appearing in the data, and it was found that these journals have published 286 articles in this field of research. Table 1 shows the number of articles published in each of these journals during the time period 2006-2016. It also depicts the total number of articles published in each year (Please see Table 1A in the Appendix for a list of all abbreviations).

Table 1: Journal wise publication break down table

Source	Publication year										
	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	Total
JCP					1	1			8	12	2	24
BDR									3	21		24
TRC-ET							2		4	14		20
IS				1					8	9	2	20
Scientometrics	1		1		3	2		1	4	5	1	18
JICS				2	1	1	2		6	5		17
IJPR			1	·		3		1	3	9		17
ICS	1					1	2		3	7	1	15
CLSR				2				1	7	4	1	15
CFS			1		3	2	2	3	1	3		15
IMDS		2			1		2		2	7		14
IEEE-SP				1		2		2	7	1		13
IJPE		2		1					1	9		13
HBR	1	2	2	1		1	5	1				13
DSS	1						1	3	4	3	1	13
JBR		1			2				1	2	3	9
McKQ						3		3	1	1		8
IJIM								1	1	1	3	6
JBL								2	3	1		6
MS					1		1	1	2	1		6
Total	4	7	5	8	12	16	17	19	69	115	14	286

Figure 1 demonstrates the changing pattern of publications in the selected journals in each year from 2006 until the beginning of 2016. It can be clearly seen from the figure that the number of publications on Big Data increased slowly from 2006 to 2013, but since then it has been increasing dramatically. This indicates that the field of Big Data in SCM is gaining increasing attention.

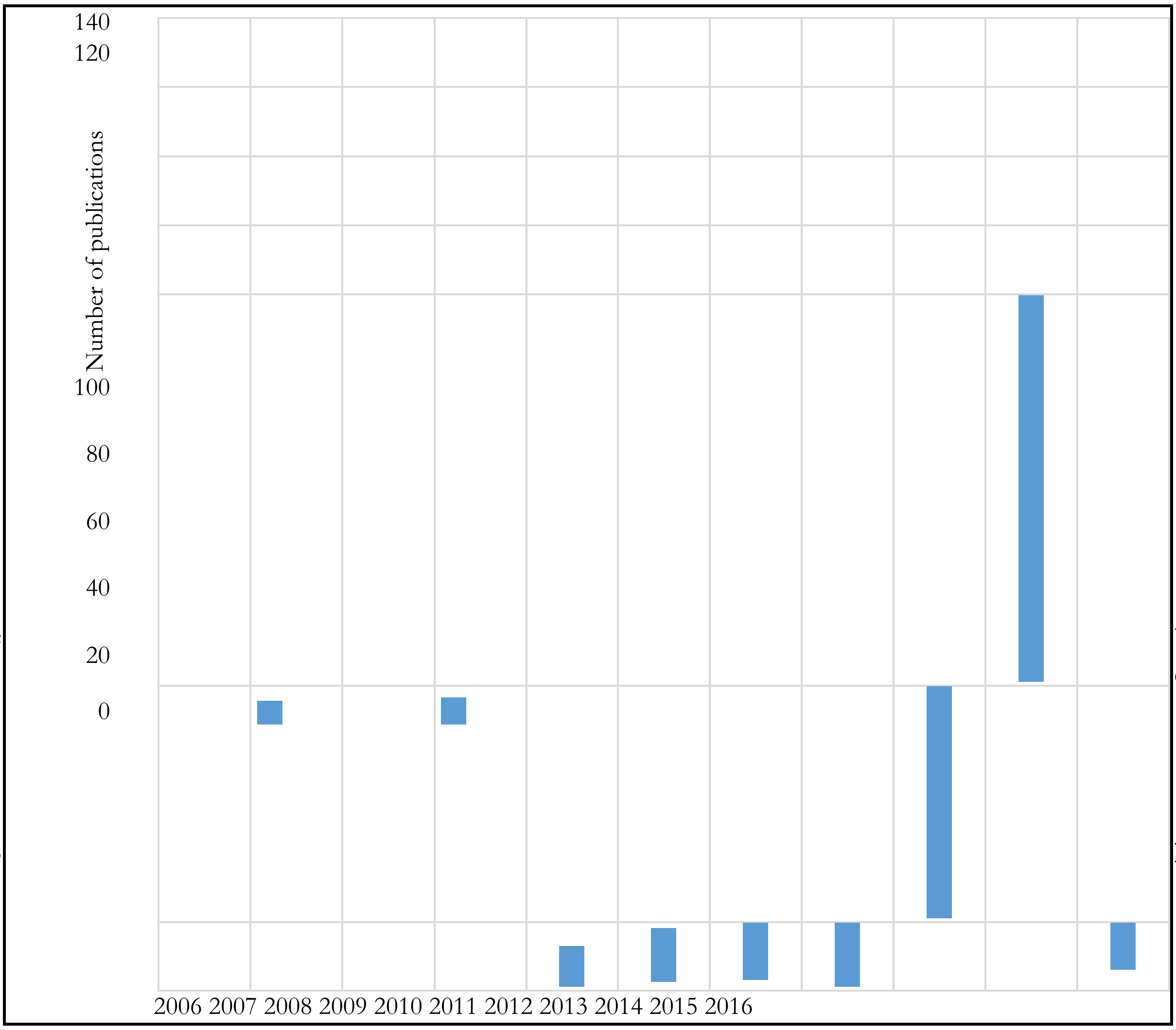

Figure 1: Distribution of articles published

3.5 Data analysis

We performed data analysis in two steps. In the first step, bibliometric analysis was performed using BibExcel software and in the second, network analysis was conducted using Gephi. BibExcel provides data statistics containing author, affiliation and keyword statistics. We decided to use this software because of its flexibility and ability to handle big amounts of data, as well as because of its compatibility with applications such as Excel, Pajek and Gephi (Persson et al. 2009). The data prepared in BibExcel software was then transferred to Gephi for further analysis. We chose Gephi over other software such as Pajek (Batagelj and Mrvar 2011) and VOSviewer (van Eck and Waltman 2013) as it has the ability to handle large data sets efficiently and can produce a range of innovative visualization, analysis and investigation options.

4. Bibliometric analysis

Bibliometric analysis can be conducted by using different software packages, such as Publish or Perish, HistCite, and BibExcel. Since other software packages have their own capabilities and limitations, we chose BibExcel in this study because it is highly flexible in handling data from different databases like Scopus and WoS. Another advantage of using BibExcel is its ability to offer an extensive data analysis which can be further used by network analysis tools, namely, Gephi, VOSviewer and Pajek. However, HistCite can only work with data imported from WoS, while Publish or Perish works with Google Scholar and Microsoft Academic Search. It is worth mentioning that apart from BibExcel these tools do not generate data for network analysis.

The data entered in BibExcel is in RIS format and contains all the necessary bibliographic information related to the papers. In our analysis, we focussed on information regarding authors, title, journal, publication year, keywords, affiliations, and references. During these analyses, the RIS file is converted into different formats and, as a result, various file types are produced. To get thorough knowledge about the processes and applications of BibExcel, readers may refer to Paloviita (2009) and Persson et al. (2009). In the forthcoming sub-sections, we present statistics on author, affiliation and keyword obtained from BibExcel analysis.

4.1 Author influence

To analyse the influence of authors using BibExcel, the author field was extracted from the RIS data file and the frequency of occurrence of these authors was noted. Table 2 shows the top ten contributing authors along with their number of publications. It can be clearly observed that Wang with 6 publications dominates the list, and is followed by Li and Wang each with 5 publications.

Table 2: Top 10 contributing authors

Author	Number of publlished articles
Wang, H.	6
Li, H.	5
Wang, J.	5
Zhang, J.	4
Li, X.	4
Li, Z.	4
Waller, M.A.	4
Zhang, Y.	4
Fawcett, S.E.	4
Wang, Y.	4
Court, D.	4

4.2 Affiliation statistics

In a similar manner, we used BibExcel to extract the affiliation of authors from the RIS data file. Then, corresponding to each affiliation, the country in which the institution is situated was taken out for further analysis. From Table 3 it can be seen that institutions in United States, China and United Kingdom are the major contributors. In fact, researchers across the world are attracted towards the area of Big Data.

Table 3: Top 20 contributing countries

Country	Number of papers	Country	Number of papers
United States	88	Taiwan	5
China	47	Canada	5
United Kingdom	17	Singapore	4
Germany	9	Sweden	4
India	7	Switzerland	4
Australia	7	France	4
South Korea	7	Spain	4
Greece	6	UK	4
Italy	6	Finland	4
Hong Kong	6	Poland	3

4.3 Keyword statistics

In this section we present the results of our keyword analysis. Such a discussion assists in revealing the intellectual core and identity construction of the discipline by looking into keywords used by research papers (and top-cited authors) and their aggregation (Scott and Lane 2000; Sidorova et al. 2008).

We adopted a similar approach to identify the most commonly used words in the paper titles and the list of keywords. The top 20 keywords used in the paper titles and most popular keywords from the list of keywords are shown in Tables 4 and 5 respectively. By comparing these two tables it can be seen that there is a uniformity in the use of keywords in the title and the list of keywords. For instance, in both tables the top keywords include a combination of Big Data, supply chain management, data mining and privacy. It is to be noted here that the most popular keywords which occur in Table 4 are the actual search keywords used for this study.

Table 4: Top 20 keywords search results

Word	Frequency	Word	Frequency
Big data	180	Privacy	14
Decision making	25	Algorithms	13
Data mining	25	Energy utilization	12
Commerce	17	Data handling	12
Information management	17	social media	12
Social networking (online)	17	Data analytics	12
Cloud computing	16	Forecasting	12
Data privacy	15	Security of data	12
Supply chain management	15	Security	12
Artificial intelligence	14	Manufacture	11

Table 5: Top 20 commonly used words in paper titles

Word	Frequency	Word	Frequency
Data	133	Research	14
Big	107	Social	12
Analytics	28	Privacy	12
Based	21	Information	12
Analysis	21	Case	12
Chain	18	Science	11
Study	17	Network	10
Management	16	Twitter	9
Approach	15	Mining	9
Supply	15	Predictive	8

5. Network Analysis

For conducting network analysis, the most widely used software packages are Pajek, VOSviewer, HistCite Graph Maker, and Gephi. In this paper we used Gephi as it provides flexible visual aids, powerful filtering techniques, an inherent toolkit for network analysis and capability to handle different data formats (Mishra et al. 2016a). Due to the flexibility provided by its multi-task architecture, Gephi can deal with complicated datasets and generate purposeful visualisations (Gephi, 2013). As input to Gephi we could not use the bibliographic data we obtained from Scopus, which was saved in RIS format. To deal with this problem, we used BibExcel to reformat the data to a graph dataset or .NET file. This file was saved for future network analysis.

5.1 Citation analysis

Citation analysis evaluates the citation frequency and subsequently is used to rank (i) journals in terms of their significance in a particular area of research (Garfield, 1972), and (ii) scholars and indicate their scientific research impact (Sharplin and Marby 1985; Culnan 1986). Therefore, citation analysis can provide insights regarding the popularity of articles over time (Pilkington and Meredith 2009). Despite the criticisms, it is still used for analysing literature and identifying influential authors, journals, or articles within a research area (Mac Roberts and Mac Roberts 1989, 2010; Vokurka 1996).

Figure 2 demonstrates the top ten most influential works published between 2006 and 2016. The most influential article during this period, having received 804 citations, is the work published by Chiu et al. (2006). The paper integrates the Social Cognitive Theory and the Social Capital Theory to construct a model for investigating the motivations behind people's knowledge sharing in virtual communities. Another important contribution was made by Boyd and Crawfold (2012) who offered six provocations to spark conversations about the issues of Big Data: a cultural, technological, and scholarly phenomenon that rests on the interplay of technology, analysis, and mythology that provokes extensive utopian and dystopian rhetoric. This work received 351 citations which reflects the significance of the article in this field. Furthermore, the article by McAfee and Brynjolfsson (2012), cited 153 times, highlighted the significance of Big Data by stating that it allows the managers to measure and thus, acquire thorough knowledge of the business which can be used to improve decision making and performance. Authors also claimed that Big Data enables firms to take decisions based on evidence rather depending upon instinct and gut feeling.

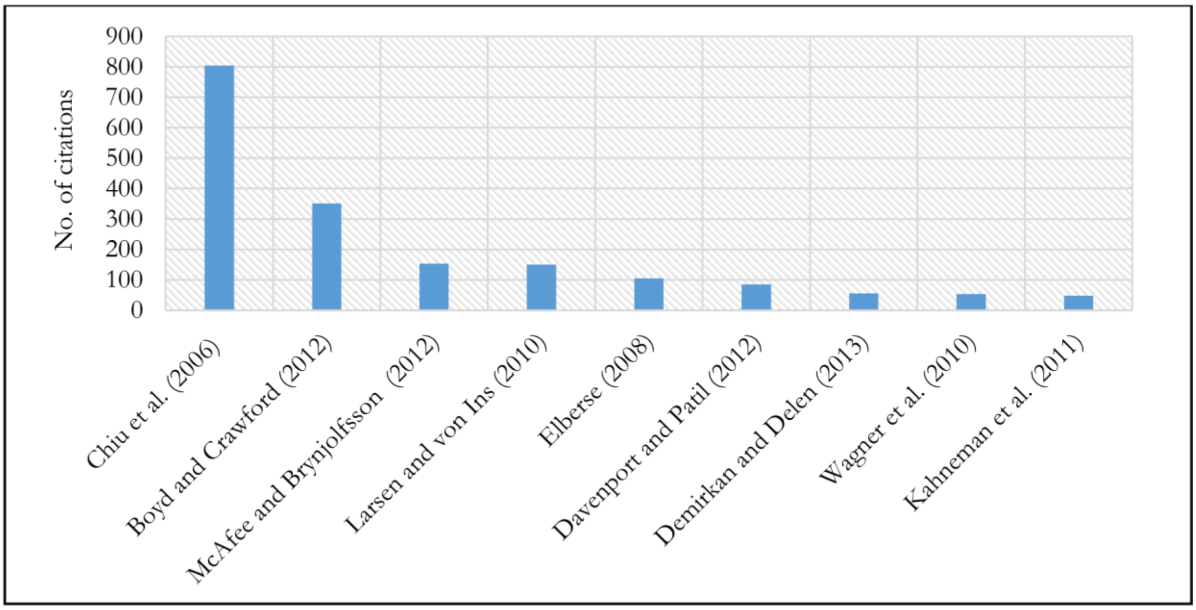

Figure 2: Frequency distribution of top 10 cited articles.

5.2 PageRank analysis

The most widely used method for measuring the importance of a paper is citation analysis (Cronin and Ding 2011), which has been discussed in the previous section. However, the popularity of an article can also be assessed by the number of times it is cited by other highly cited articles (Ding et al. 2009). To ensure that popularity and prestige are correlated, PageRank was introduced by Brin and Page (1998) to measure these concepts and prioritise the results of web keyword searches (Mishra et al. 2016a; 2016b).

Assume that paper A has been cited by papers $T_1, \ldots, T_n$. Define a parameter d as the damping factor, which represents the fraction of random walks that continue to propagate along the citations. The value of parameter d is fixed between 0 and 1. Now, define C (T_i) as the number of times paper T_i has cited other papers. The PageRank of paper A, denoted by PR (A), in a network of N papers is calculated as follows:

$$PR(A) = \frac{(1-d)}{N} + d\left(\frac{PR(T_1)}{C(T_1)} + \cdots + \frac{PR(T_n)}{C(T_n)}\right)$$

It is important to note that if C(T_i) = 0, then PR(T_i) will be divided to the number of papers instead of C(T_i). The value of parameter d has been the subject of debate, with scholars suggesting a value of 0.85 (Brin and Page 1998) while others a value of 0.5 (Chen et al. 2007).

The top 10 papers using PageRank analysis are shown in Table 6. When comparing Table 5 and Table 6, it is observed that the topmost paper based on citations, Chiu et al. (2006) is not present in this list whereas McAfee and Brynjolfsson (2012) which was at third position in Table 5 is at second position in Table 6. The second highly cited paper, Boyd and Crawford (2012), is in sixth position, and Jacobs (2009), not present in Table 5, dominates the list in Table 6.

Table 6: Top 10 articles based on PageRank

Author (year)	Page Rank	Citation
Jacobs (2009)	0.0142	76
McAfee and Brynjolfsson (2012)	0.0137	153
Manyika et al. (2011)	0.0127	1809
Chen et al. (2012)	0.0097	355
Barton and Court (2012)	0.0096	30
Lavalle et al. (2011)	0.0095	78
Boyd and Crawford (2012)	0.0092	351
Schadt et al. (2010)	0.0083	198
Acker et al. (2011)	0.0083	8
Demirkan and Delen (2013)	0.0080	52

5.3 Co-citation analysis

Co-citation analysis can be used in authors and/or publications order to track and study the relationship between authors, topics, journals or keywords (Small 1973; Pilkington and Liston Heyes 1999). If applied on authors, co-citation analysis reveals the structure of the social relationships between authors, while if applied on publications the intellectual structure of a field (Chen et al. 2010) as well as the evolution and variation of research over time (Pilkington and Meredith 2009) can be seen.

To conduct co-citation analysis: (i) We opened the .NET file for 286 articles in Gephi and a random map was generated in Gephi with no visible pattern. (ii) To restore visibility, we used the algorithm 'Force Atlas' provided by Gephi and created networks of co-cited articles. The structure of the network allowed strongly connected nodes to be centralised while loosely connected nodes were located in the boundaries of the network.

The Force Atlas layout node co-citation map is shown in Figure 3. The co-cited articles are connected with each other while the poorly connected nodes shift away from the centre. The nodes that are isolated from rest of the network, termed 'outliers', are excluded for the purpose of data clustering that takes place in the next section. On excluding these outliers we are left with a network having 233 nodes and 1096 edges.

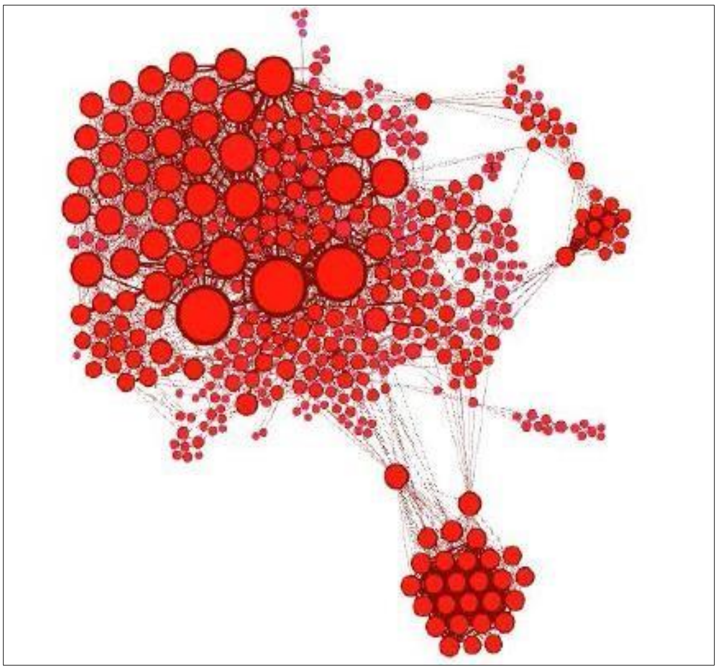

Figure 3: Force Atlas layout of 233 connected nodes

5.3.1 Data clustering

Data clustering aims at placing together sets of articles that share same characteristics (Radicchi et al. 2004). To conduct data clustering (i) we placed nodes so that links of nodes within the same cluster are dense compared to the nodes belonging in different clusters (Clauset et al. 2004; Leydesdorff 2011; Radicchi et al. 2004). To measure the density of the links, we used the concept of Modularity (Blondel et al. 2008) which was further measured in Gephi with the Louvain algorithm, where the value of modularity index varies between -1 and +1. The formula for modularity index was provided by Blondel et al. (2008) and used in other studies (e.g. Mishra et al. 2016a):

$$Q = \frac{1}{2m}\sum_{ij}\left[A_{ij} - \frac{k_i k_j}{2m}\right] \delta\left(c_i, c_j\right),$$

where A_{ij} represents the weight of the edge between nodes i and j, k_i is the sum of the weights of the edges attached to node i ($k_i = \sum_j A_{ij}$), c_i is the community to which vertex i is assigned, δ(u, v) is equal to 1 if u = v and 0 otherwise, and finally $m = (1/2)\sum_{ij} A_{ij}$.

In this study, this algorithm was applied to 233-node network (see previous section) thereby creating six major clusters; their positioning and interaction is extrapolated in Figure 4, and the value of the modularity index was calculated as 0.19. This means that within each cluster there exists a strong relationship among nodes. Furthermore, from Figure 4 we infer different levels of thickness between the nodes. This is because of the difference in the frequency for co-occurrence of any two papers in the reference list of other papers (Mishra et al. 2016a; 2016b).

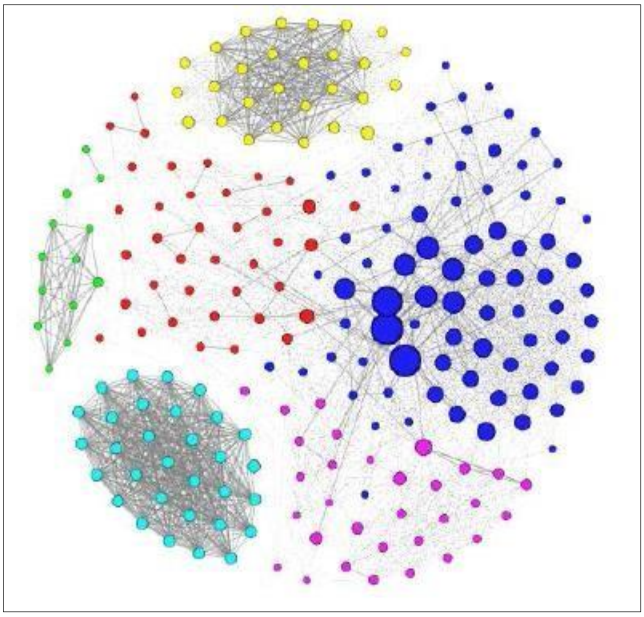

Figure 4: Structure of six clusters

Hjørland (2013) noted that the papers that are often cited together are more likely to share same area of interest. Therefore, the research area of a cluster can be identified by a thorough analysis of the papers belonging to that cluster. Since the number of papers in each cluster is high, we considered only the top publications of each cluster which were identified on the basis of their cocitation PageRank (Mishra et al. 2016a). Table 7 shows the top publications of each cluster based on PageRank.

The contents and research areas of the leading papers were carefully examined to find out the research focus area of each of the six clusters. It was found that researchers belonging to cluster 1 have contributed by giving theoretical and conceptual studies on Big Data. They suggest that the era of Big Data is growing rapidly, and more importantly, advanced analytic tools must be developed to operate on such data sets. Hence, cluster 1 was targeted to study the concept of Big Data and analytics. Research in cluster 2 mainly revolved around the application of Big Data in SCM, the problems associated with data quality and how this concept can be used to resolve the problems of supply chains. Studies in this cluster also analysed that how data generated from social networking sites and especially Twitter can be used for predicting stock markets, in mass convergence and emergency events; these studies have also proposed frameworks for social media analytics in political contexts. Next, cluster 3 mainly concentrated on developing architectures, algorithms and models for processing and generating large data sets, such as MapReduce, S4, among others, while researchers in cluster 4 investigated the utilization of data in hospitals, i.e. for studying productivity developments and for comparing quality of care measurements.

Table 7: Top 10 papers of each cluster: co-citation PageRank measure

Cluster 1	Cluster 2	Cluster 3
Jacobs, 2009	Bollen et al., 2011	Laney, 2001
Mcafee and Brynjolfsson, 2012	Anderson, 2008	Sakr et al., 2011
Manyika et al., 2011	Cecere, 2012	Bengio, 2009
Russom, 2011	Davenport and Harris, 2007	Neumeyer et al., 2010
Chen et al., 2012	Hazen et al., 2014	Leong, 2009
Barton and Court, 2012	Thelwall et al., 2011	Lynch, 2008
Lavalle et al., 2011	Hughes and Palen, 2009	Ishii and De Mello, 2009
Boyd and Crawford, 2012	Chae et al., 2014	Dean and Ghemawat, 2008
Schadt et al., 2010	Stieglitz and Dang-Xuan, 2013	Cohen et al., 2009
Acker et al., 2011	Boyd, 2010	Isard et al., 2007
Cluster 4	Cluster 5	Cluster 6
Fare et al., 1995	Flyvbjerg, 2013	Richardson and Domingos, 2002
Yu et al., 2009	Erikson and Wlezien, 2012	Rogers, 2003
Tieman, 2003	Larrick and Soll, 2006	Kleinberg, 2007
Reinsdorf et al., 2002	Graefe et al., 2015	Kempe et al., 2005
Roberts, 2004	Graefe, 2015	Kempe et al., 2003
Mekhjian et al., 2003	Fildes and Petropoulos, 2015	Domingos and Richardson, 2001
Lovis et al., 2007	Goodwin, 2015	Leskovec et al., 2007
Mcglynn, 2008	Gardner, 2006	Narayanam and Narahari, 2011
Stolle, 2010	Makridakis and Hibon, 2000	Zhu et al., 2014
Segal, and Heer, 2010	Tessier and Armstrong, 2015	Wang et al., 2014

Studies in cluster 5 were concerned with developing methods and models that can be used in forecasting election results or social science problems with a main focus on Bayes formula, and in improving these forecasts by using specific predictors. Lastly, data mining and its applications were the main topic of interest for researchers in cluster 6. In this cluster scholars were also interested in developing algorithms and models for maximizing the spread of influence through a social network. It can be observed that cluster 1 is the most popular one, clusters 2 to 5 have received considerable attention from researchers while there is a scope of future work in cluster 6.

The proposed six cluster classification is extrapolated in Table 8, where the different clusters, current research and suggestions for future research for each of the clusters are extrapolated. We note that the clusters would need to be seen in relation to each other. In particular, conceptualising Big Data (Cluster 1) is the first step to building applications for Big Data analytics focusing on SCM (in Cluster 2) and healthcare (Cluster 4). Furthermore, in order to build applications for SCM and healthcare, clusters 3, 5, and 6 are needed to build robust optimised applied tools for Big Data and data mining.

Table 8: Proposed cluster classification with current and future research per cluster

Cluster number and label	**Current research**	**Future research suggestions**
Cluster 1: Conceptualisation of Big Data and analytics	Theoretical and conceptual studies on Big Data.	Advanced analytic tools to operate on such data sets.
Cluster 2: Big Data and SCM	Data quality and related challenges and how Big Data can be used to resolve SCM challenges. Social media data analysis for predicting stock markets, in mass convergence and emergency events.	Frameworks for social media analytics and SCM.
Cluster 3: Big Data tools and algorithms	Architectures, algorithms and models for processing and generating large data sets.	Applied tools for Big Data analysis in SCM. Capacity building.
Cluster 4: Big Data applications in healthcare	Applications of Big Data in healthcare.	Big Data analytics for productivity and care quality provision.
Cluster 5: Big Data and forecasting	Forecasting election results or social science problems. Main focus on Bayes formula.	Improve forecasts by using predictors.
Cluster 6: Data mining and applications	Developing data mining techniques, algorithms and models.	Predictive science using large data sets. Optimisation of algorithms for faster analytics.

6. Discussion

6.1 Contributions to theory

The current study contributes to the literature on Big Data and extends current reviews (Sagiroglu and Sinanc 2013; Fosso-Wamba et al. 2015; Gandomi and Haider 2015; Wang et al. 2016; Khorheh et al. 2015; Mishra et al. 2016a; 2016b) in that: (i) it goes beyond a mere systematic literature review of the field since it proposes and applies the techniques of bibliometric and network analysis to obtain and compare the most influential works (based on citations, co-citations and PageRank), (ii) through the aforementioned it analyses, identifies and proposes six clusters ('Conceptualisation of Big Data and analytics', 'Big Data and SCM', 'Big Data tools and algorithms', 'Big Data applications in healthcare', 'Big Data and forecasting', and 'Data mining and applications') that focus on particular areas of Big Data, from conceptualisation to methods and tools and applications in SCM and healthcare; and (iii) it illustrates the relationships between the clusters and argues that better conceptualisation and consensus of Big Data and use of particular tools and techniques will result in better applications of Big Data in SCM (and healthcare), and therefore future research should include all these clusters, starting from Cluster 1 (conceptualisations) to 3,5, and 6 (tools and techniques), to 2 and 4 (applications in different fields and SCM).

We further argue that Big Data and SCM has attracted significant attention from scholars but the Big Data research is in nascent stage and there is urgent need for research to delineate high quality data sets from poor quality data sets (Hazen et al. 2014). Furthermore, while analysing Big Data and SCM related research using the perspective of Waller and Fawcett (2013a), we noted that there are gaps in the literature and in particular on machine learning techniques for SCM applications. Understanding the role of BDA in improving SCM is extremely valuable since integration of BDA in operations and supply chains aids firms in realizing their customers in a better way, minimizing cost to serve, managing risk efficiently, and in generating new and unexpected sources of revenue (Sanders and Ganesan, 2015). Thus, future research should assess the ability of BDA to improve intra- and inter- firm efficiency and effectiveness (e.g., identification of bottlenecks, improved predictive maintenance, and scenario building for improved quality control) (Fosso-Wamba and Akter, 2015). Therefore further research is required in this field.

6.2 Contributions to Managerial Practice

Our study offers multiple opportunities to the practitioners and consulting firms that are engaged in leveraging benefits from supply chains using Big Data. Our study can equip managers with different schools of thought that enable them to harness the benefits from using Big Data and analytics for SCM in their everyday work. Furthermore, through our proposed six cluster classification of the literature, managers can: (i) assess the current state of their Big Data in terms of conceptualisation, tools and techniques, and different applications and (ii) identify their future needs in the relevant 'clusters' in order to take appropriate decisions on whether to invest and improve current tools/techniques and/or further re-think the conceptualisation of Big Data, as well as the implications for the realisation of their business strategy through Big Data.

6.3 Limitations of the study

The limitations of the study are as follows:

1. Our review was based on the review of the literature on Big Data and SCM using 286 articles published in past 10 years. We have used particular keywords for this research, and it may be that the use of other keywords may have yielded different results.
2. We used the bibliometric and network analysis for reviewing the literature based on Pilkington and Meredith (2009). Other methods may be used for such an analysis.
3. We have used classified the literature in six research clusters. Other methods may result in other classifications.

7. Conclusion

Drawing on bibliometric and network analysis we presented an extensive review of literature on Big Data and SCM over the period of 10 years (2006-2016). We offered insights regarding the contributions of scientific journals towards advancing Big Data related research and the contributions of researchers to the emerging field of Big Data. To the best of our knowledge this is the first study attempting to identify the top contributing authors, countries and key research topics related to this field. Despite the limitations, we believe that our study provides food for thought and encouragement for researchers to further investigate the field of Big Data and SCM.

References

Acker, O., Gröne, F., Blockus, A., & Bange, C. (2011). In-memory analytics – strategies for real-time CRM. *Journal of Database Marketing & Customer Strategy Management, 18* (2), 129–136.

Anderson, C. (2008). The end of theory, will the data deluge makes the scientific method obsolete? Edge, http://www.edge.org/3rd_culture/anderson08/anderson08_index.html (25 July 2011).

Barton, D., & Court, D. (2012). Making advanced analytics work for you. *Harvard business review, 90*(10), 78-83, 128.

Batagelj, V., & Mrvar, A. (2011). Pajek: Program for Analysis and Visualization of Large Networks – *Reference Manual*, University of Ljubljana, Slovenia.

Bengio, Y. (2009). Learning deep architectures for AI. Foundations and Trends in Machine Learning, 2 (1), 1–127.

Bi, Z., & Cochran, D. (2014). Big Data Analytics with Applications. *Journal of Management Analytics, 1*(4), 249-265.

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of*

Computational Science, 2 (1), 1-8.

Boyd, D., & Crawford, K. (2012). Critical questions for Big Data. *Information, Communication and Society, 15*(5), 662–679.

Boyd, D. (2010). Social network sites as networked publics: Affordances, dynamics, and implications. In Z. Papacharissi (Ed.), *A networked self: Identity, community and culture on social network sites* (pp. 39–58). New York: Routledge.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hyper textual Web search engine. *Computer Networks and ISDN Systems, 30* (1-7), 107–117.

Brown B., Chul M. & Manyika J. (2011). Are you ready for the era of Big Data? *McKinsey Quarterly, 4*, 24-27+30-35.

Casemore, S. (2012). Social Media and the Coming Supply-Chain Revolution.

http://www3.cfo.com/article/2012/2/supply-chain_supply-chain-innovation-social-media casemore-ghg.

Cecere, L. (2012). Big Data: Go Big or Go Home? *Supply Chain Insights LLC* ⟨http:// supplychaininsights.com/wpcontent/uploads/2012/07/Big_Data_Report_16_JULY_2012.pdf⟩ (accessed 05.11.13).

Chae, B., & Olson, D. (2013). Business analytics for supply chain: a dynamic-capabilities framework. *International Journal of Information Technology & Decision Making, 12* (9), 9–26.

Chae, B., Olson, D., Sheu, C. (2014). The impact of supply chain analytics on operational performance: a resource based view. *International Journal of Production Research, 52,* 4695-4710.

Chen, P., Xie, H., Maslov, S., & Redner, S. (2007). Finding scientific gems with Google's PageRank algorithm. *Journal of Informetrics, 1*, 8–15.

Chen, C., Ibekwe- SanJuan, F, & Hou, J. (2010). The Structure and Dynamics of Co-Citation Clusters: A Multiple-Perspective Co-Citation Analysis. *Journal of the American Society for Information Science, 61* (7), 1386–1409.

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.

Chiu, C. M., Hsu, M. H., Wang, E., & T. G. (2006). Understanding knowledge sharing in virtual communities: An integration of social capital and social cognitive theories. *Decision Support Systems, 42* (3), 1872−1888.

Clauset, A., Newman, M.E.J., & Moore, C. (2004). Finding community structure in very large networks, *Physical Review E, 70* (6), 1-6.

Cohen, J., Dolan, B., Dunlap, M., Hellerstein, J. M., & Welton, C. (2009). MAD Skills: New Analysis Practices for Big Data. *Proceedings of the VLDB Endowment 2* (2), 1481-1492.is

Craighead, C.W., Hult, G.T.M., & Ketchen Jr., D.J. (2009). The effects of innovation—cost strategy, knowledge, and action in the supply chain on firm performance. *Journal of Operations Management, 27* (5), 405–421.

Cukier, K. (2010). The Economist, Data, data everywhere: A special report on managing information", February 25, Retrieved from http://www.economist.com/node/15557443.

Culnan, M. (1986). The intellectual development of management information systems. *Management Science, 32* (2), 156-172.

Davenport, T.H. (2006). Competing on analytics. *Harvard Business Review, 84* (1), 84–93.

Davenport, T.H., & Harris, J.G. (2007). Competing on analytics: the new science of winning. *Harvard Business School Press.*

Dean, J., & Ghemawat, S. (2004). MapReduce: Simplified data processing on large clusters, in: *Proc. 6th USENIX Symposium on Operating Systems Design and Implementation*, OSDI 2004, San Francisco, USA, Dec.

Diebold, F. X. (2012). A personal perspective on the origin(s) and development of "Big Data": The phenomenon, the term, and the discipline. Scholarly Paper No. ID 2202843, Social Science Research Network. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract id=2202843.

Ding, Y., & Cronin, B. (2011). Popular and/or prestigious? Measures of scholarly esteem. *Information Processing Management, 47* (1), 80–96.

Ding, Y., Yan, E., Frazho, A., & Caverlee, J. (2009). PageRank for ranking authors in co-citation networks. *Journal of the American Society for Information Science and Technology, 60* (11), 2229–2243.

Domingos, P., & Richardson, M. (2001). Mining the network value of customers. In *Proceedings of the 7th International Conference on Knowledge Discovery and Data Mining.* ACM, New York, 57–66.

Dubey, R., Gunasekaran, A., Childe, S. J., Wamba, S. F. & Papadopoulos, T. (2015). The impact of Big Data on world-class sustainable manufacturing. *The International Journal of Advanced Manufacturing Technology*, 1-15.

Erikson, R. S., & Wlezien, C. (2012b). Markets vs. polls as election predictors: an historical assessment. *Electoral Studies, 31,* 532–539.

Fahimnia, B., Sarkis, J., & Davarzani, H. (2015). Green supply chain management: A review and bibliometric analysis. *International Journal of Production Economics, 162,* 101–114.

Fare, R., Grosskopf, S., Lindgren, B., & Roos, P. (1994a). Productivity developments in Swedish hospitals: A Malmquist output index approach. In: Charnes, A., Cooper, W.W., Lewin, A.Y., Seiford, L.M. (Eds.), *Data Envelopment Analysis: Theory, Methodology and Applications*. Kluwer Academic Publishers, Boston, (pp. 253–272).

Feng, Z. Y., Guo, X. H. and Zeng, D. J. (2013). On the research frontiers of business management in the context of Big Data. *Journal of Management Sciences in China, 16* (1), 1–9.

Fildes, R., & Petropoulos, F. (2015). Simple versus complex selection rules for forecasting many time series. *Journal of Business Research, 68*(8), 1692–1701.

Flory, M.M. (2012). The big kahuna. *Marketing Research* 24, 2, 3.

Flyvbjerg, B. (2014). What you should know about megaprojects and why: An overview. *Project Management Journal* 45 (2), 6–19.

Forrester (2012), The Big Deal about Big Data for Customer Engagement Business: Leaders Must Lead Big Data Initiatives to Derive Value”, http://www.forrester.com/The+Big+Deal+About+Big+Data+For+Customer+Engagement/fulltext/-/E-RES72241>.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big Data concepts, methods, and analytics. *International Journal of Information Management, 35* (2), 137-144.

Gardner, E.S., Jr. (2006). Exponential smoothing: The state of the art—Part II. *International Journal of Forecasting, 22,* 637–666.

Garfield, E. (1972). Citation Analysis as a Tool in Journal Evaluation. *Science, 178*, 471-479.

Gephi (2013). Gephi – Makes Graphs Handy.

Gobble, M. M. (2013). Big Data: the next big thing in innovation. *Research Technology Management, 56* (1), 64–66.

Goodwin, P. (2015). When simple alternatives to Bayes formula work well: Reducing the cognitive load when updating probability forecasts. *Journal of Business Research, 68* (8), 1686–1691.

Graefe, A., Küchenhoff, H., Stierle, V., & Riedl, B. (2015). Limitations of ensemble Bayesian model averaging for forecasting social science problems. *International Journal of Forecasting, 31* (3), 943-951.

Graefe, A. (2015). Improving forecasts using equally weighted predictors. *Journal of Business Research, 68* (8), 1792–1799.

Hazen, B.T., Boone, C.A., Ezell, J.D., & Jones-Farmer, L.A. (2014). Data quality for data science, predictive analytics, and Big Data in supply chain management: an introduction to the problem and suggestions for research and applications. *International Journal of Production Economics, 154,* 72–80.

Hjørland, B. (2013). Citation analysis: a social and dynamic approach to knowledge organization. *Information Processing Management, 49* (6), 1313–1325.

Hopkins, M.S., Lavalle, S., & Balboni, F. (2010). The new intelligent enterprise: 10 insights: a first look at the new intelligent enterprise survey. *MIT Sloan Management Review, 52* (1), 22.

Huang, M-Ch., Yen, G-F., & Liu, T-Ch. (2014). Re-examining supply chain integration and the supplier's performance relationships under uncertainty. *Supply Chain Management: An International Journal, 19* (1), 64–78.

Hughes, A.L., & Palen, L. (2009). Twitter adoption and use in mass convergence and emergency events. *International Journal of Emergency Management, 6*, 248–260.

Isard, M., Budiu, M., Yu, Y., Birrell, A., & Fetterly, D. (2007). Dryad: distributed data-parallel programs from sequential building blocks, in: EuroSys'07 *Proceedings of the 2nd ACM SIGOPS/EuroSys European Conference on Computer Systems, 41*(3), 59–72.

Ishii, R.P., & de Mello, R. F. (2012). An online data access prediction and optimization approach for distributed systems. *IEEE Transactions on Parallel and Distributed System, 23* (6) (2012) 1017–1029.

Jacobs, A. (2009). The Pathologies of Big Data. *Magazine Queue – Data, 7* (6), 10.

Kempe, D., Kleinberg, J., & Tardos, E. (2003). Maximizing the spread of influence through a social network. In KDD '03: *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, 137–146, New York, NY, USA, ACM.

Kempe, D., Kleinberg, J., & Tardos, E. (2005). Influential nodes in a diffusion model for social networks. In ICALP 2005: *Proceedings of the 32nd International Colloquium on Automata*, Languages and Programming, 1127–1138.

KetchenJr, D.J., & Hult, G.T.M., (2007). Bridging organization theory and supply chain management: the case of best value supply chains. *Journal of Operations Management, 25* (2), 573-580.

Khorheh, M.A., Moisiadis, F., & Davarzani, H. (2015). Socio-environmental performance of transportation systems. *Management of Environmental Quality: An International Journal, 26* (6), 826 851.

Kleinberg, J., (2007). Cascading behavior in networks: algorithmic and economic issues. *Algorithmic*

Game Theory.

Laney, D. (2001). 3d Data management: controlling data volume, velocity and variety. *Technical report Application Delivery Strategies Meta Group.*

Larrick, R. P., & Soll, J. B. (2006). Intuitions about combining opinions: misappreciation of the averaging principle. *Management Science, 52*, 111–127.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M.S., & Kruschwitz, N. (2011). Big Data: analytics and the path from insights to value. *MIT Sloan Management Review, 52* (2), 21–32.

Leong, D. (2009). A new revolution in enterprise storage architecture, *IEEE Potentials, 28* (6), 32–33.

Leskovec, J., Kleinberg, J., & Faloutsos, C. (2007). Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data, 1*(1).

Leydesdorff, L. (2011). Bibliometrics/citation networks. In: Barnett, G.A. (Ed.), *Encyclopedia of Social Networks*, SAGE Publications, Inc, Thousand Oaks, CA.

Lovis, C., Spahni, S., Cassoni-Schoellhammer, N., & Geissbuhler, A. (2006). Comprehensive management of the access to a component-based healthcare information system, *Studies in Health Technology and Informatics, 124*, 251–256.

Lynch, C. (2008). Big Data: how do your data grow? *Nature, 455* (7209), 28–29.

MacRoberts, M.H., & MacRoberts, B.R. (1989). Problems of citation analysis: A critical review. *Journal of the American Society for Information Science, 40* (5), 342–349.

MacRoberts, M.H., & MacRoberts, B.R. (2010). Problems of citation analysis: A study of uncited and seldom-sited influences. *Journal of the American Society for Information Science and Technology, 61* (1), 1-12.

Makridakis, S., & Hibon, M. (2000). The M-3 competition: Results, conclusions and implications. *International Journal of Forecasting, 16*, 451–476.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big Data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt, New York.

McAfee A., & Brynjolfsson E. (2012). Big Data: The management revolution. *Harvard Business Review, 90* (10), 60-66, 68, 128.

McGlynn, E. (2008). The Case for Keeping Quality on the Health Reform Agenda. Testimony before the U.S. Senate Committee on Finance. *Technical report*, June 3.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A.H. (2011). Big Data: the Next Frontier for Innovation, Competition and Productivity. *McKinsey Global Institute*.

Mekhjian, H., Saltz, J., Rogers, P., & Kamal, J. (2003). Impact of CPOE order sets on lab orders. In: *AMIA Annual Symposium Proceedings*, 931.

Mervis, J. (2012). Agencies rally to tackle Big Data. *Science, 336* (6077), 22.

Mishra, D., Gunasekaran, A., Childe, S., Papadopoulos, T., & Dubey, R. (2016a). Vision, applications and future challenges of Internet of Things: A bibliometric study of the recent literature. *Industrial Management and Data Systems* (Accepted Paper).

Mishra, D., Luo, Z., Jiang, S., Papadopoulos, T., & Dubey, R. (2016b). A Bibliographic Study on Big Data: Concepts, Trends and Challenges. *Business Process Management Journal* (Accepted Paper).

Narayanam, R., & Narahari., Y. (2011). A shapley value based approach to discover influential nodes in social networks. *IEEE Transactions on Automation Science and Engineering.*

Neumeyer, L., Robbins, B., Nair, A., & Kesari, A. (2010). S4: distributed stream computing platform. In: *2010 IEEE Data Mining Workshops* (ICDMW), Sydney, Australia, 170–177.

Oh, L., Teo, H., & Sambamurthy, V. (2012). The effects of retail channel integration through the use of information technologies on firm performance. *Journal of Operations Management, 30* (1), 368–381.

O'Leary, D. E. (2011). Blog Mining – Review and Extensions: From each according to his opinion. *Decision Support Systems, 51* (4), 821-830.

Oracle (2012). Big Data for the Enterprise. *Redwood Shores*, CA: Oracle.

Oxford Economics (2011). The new digital economy: how it will transform business. ⟨http://www.google.co.in/url?sa=t&rct=j&q=&esrc=s&source=web&cd=3&ved= 0CD0QFjAC&url = http%3 A%2 F%2Fwww.businessandleadership.com%2F download%2Ffs%2Fdoc%2Freports%2Fthe-new-digital-economy-1.pdf&ei =DeHUUpuW M66QiQfjs4HIBA&usg=AFQjCNGQKXaQccqoEy8Na0ctStXgaJPH1Q&bvm=bv.593784

⟩(accessed 10.01.14.).
65,d.aGc

Paloviita, A. (2009). Stakeholder perceptions of alternative food entrepreneurs. World Review of Entrepreneurship. *Management and Sustainable Development, 5* (4), 395–406.

Patil, R. (2014). Supermarket Tesco pioneers big data. Available at: http://dataconomy.com/tesco-pioneers-big-data/ (Accessed 30/04/2016).

Persson, O., Danell, R., & Schneider, J.W. (2009). *How to use Bibexcel for various types of bibliometric analysis*. In: Åstrom, F., Danell, R., Larsen, B., Schneider, J.W. (Eds.), Celebrating Scholarly Communication Studies.

Pilkington, A., & Meredith, J. (2009). The evolution of the intellectual structure of operations management—1980–2006: a citation/co-citation analysis. *Journal of Operations Management, 27* (3), 185–202.

Pilkington, A., & Liston-Heyes, C. (1999). Is production and operations management a discipline? A citation/co- citation study. *International Journal of Operations and Production Management, 19* (1), 7-20.

Radicchi, F., Castellano, C., Cecconi, F., Loreto, V., & Parisi, D. (2004). Defining and identifying communities in networks. *Proceedings of the National Academy of Sciences of the United States of America*, *101* (9), 2658–2663.

Ramos-Rodríguez, A.R., & Ruiz-Navarro, J. (2004). Changes in the intellectual structure of strategic management research: A bibliometric study of the Strategic Management Journal, 1980-2000. *Strategic Management Journal, 25* (10), 981-1004.

Records, R. L., & Fisher, Q. K. (2014). Manufacturers connect the dots with Big Data and analytics. *Computer Science Corporation*, 1–6 (accessed 31.05.14).

Reinsdorf, M.B., Diewert, W.E., & Ehemann, C. (2002). Additive decompositions for Fisher, Törnqvist and geometric mean indexes. *Journal of Economic and Social Measurement, 28*, 51–61.

Rowley, J., & Slack, F. (2004). Conducting a literature review. *Management Research News, 27*, 31–39

Richardson, M., & Domingos, P. (2002). Mining knowledge-sharing sites for viral marketing. *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton*, Canada: ACM Press, 61–70.

Roberts, S. (2004). Self-experimentation as a source of new ideas: Sleep, mood, health and weight. *Behavioral and Brain Sciences, 27*, 227-262.

Rogers, E.M. (2003). *Diffusion of Innovations*. Fifth edition. New York: Free Press.

Russom, P. (2011a). *The Three Vs of Big Data Analytics: TDWI.*

Russom, P. (2011b). Big Data Analytics, Best Practices Report, Fourth Quarter, The Data Warehouse Institute, Renton, WA, September 18 2011 (available at tdwi.org).

Sagiroglu, S., & Sinanc, D. (2013). Big Data: A Review. *International Conference on Collaboration*

Technologies and Systems (CTS), 42-47.

Sakr, S., Liu, A., Batista, D.M., & Alomari, M. (2011). A survey of large scale data management approaches in cloud environments. *IEEE Communications Surveys & Tutorials, 13* (3), 311–336.

Sanders, N.R., & Ganeshan, R. (2015). Special Issue of Production and Operations Management on "Big Data in Supply Chain Management. *Production and Operations Management, 24* (5), 852–853.

Schadt, E.E., Linderman, M.D., Sorenson, J., Lee, L., & Nolan, G.P. (2010). Computational solutions to large-scale data management and analysis. *Nature Reviews Genetics, 11* (9), 647–657.

SC Digital (2014). Impact of Big Data and Analytics in supply chain execution. Available at: http://www.supplychaindigital.com/logistics/3382/Impact-of-Big-Data-and-Analytics-in-supply-chain-execution (Accessed 30/04/2016)

Schonberger, M. V., & Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work and Think,* Canada: Eamon Dolan/Houghton Mifflin Harcourt, pp. 242.

Schoenherr, T., & Speier-Pero, C. (2015). Data science, predictive analytics, and Big Data in supply chain management: Current state and future potential. *Journal of Business Logistics, 36*(1), 120-132.

Scott, S.G., & Lane, V.R. (2000). A stakeholder approach to organizational identity. *Academy of Management Review, 25* (1), 43–62.

Segel E., & Heer, J. (2010). Narrative visualization: Telling stories with data. *TVCG: Transactions on Visualization and Computer Graphics*, *16*(6):1139–1148.

Sharplin, A., & Mabry, R. (1985). The relative importance of journals used in management research: an alternative ranking. *Human Relations, 38* (2), 139-149.

Sidorova, A., Evangelopoulos, N., Valacich, J.S., & Ramakrishnan, T. (2008). Uncovering the intellectual core of the information systems discipline. *MIS Quarterly, 32* (3), 467–482.

Slone, R.E. (2004). Leading a supply chain turn around. *Harvard Business Review, 82* (10), 114–121.

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between

two documents. *Journal of the American Society for Information Science, 24*.

Song, M., Fisher, R., Wang, J-L., & Cui, L-B. (2016). Environmental performance evaluation with

Big Data: theories and methods. *Annals of Operations Research*. doi: 10.1007/s10479-016-2158-8.

Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and information diffusion in social media – sentiment of microblogs and sharing behavior. *Journal of Management Information Systems, 29,* 217–248.

Strawn, G.O. (2012). Scientific research: how many paradigms? *Educause Review, 47* (3), 26-34.

Tan, K.H., Zhan, Y., Ji, G., Ye, F., & Chang, C. (2015). Harvesting big data to enhance supply chain innovation capabilities: an analytic infrastructure based on deduction graph. *International Journal of Production Economics, 165,* 223–233.

Terziovski, M. (2010). Innovation practice and its performance implications in small and medium enterprises (SMEs) in the manufacturing sector: a resource‐based view. *Strategic Management Journal, 31*(8), 892-902.

Tessier, T.H., & Armstrong, J.S. (2015). Decomposition of time-series by level and change. *Journal of Business Research, 68*(8), 1755–1758.

Thelwall, M., Buckley, K., & Paltoglou, G. (2011). Sentiment in Twitter events. *Journal of the American Society for Information Science and Technology, 62*, 406–418.

Thibeault, J., & Wadsworth, K. (2014). *Delivering Digital Experiences That People Want to Share*. Wiley, Hoboken.

Tieman, J. (2003). Experimenting with quality. CMS-premier initiative to reward best, punish worst. *Modern Healthcare, 33*(28), 6.

Trkman, P., McCormack, K., Oliveira, M., & Ladeira, M. (2010). The impact of business analytics on supply chain performance. *Decision Support Systems, 49* (3), 318–327.

Tsai, J., Raghu, T. S., & Shao, B. B. M. (2013). Information systems and technology sourcing strategies of e-Retailers for value chain enablement. *Journal of Operations Management, 31*(6), 345–362.

Yu, F. B., Menachemi, N., Berner, E. S., Allison, J. J., Weissman, N. W., & Houston, T. K. (2009). Full implementation of computerized physician entry and medication-related quality outcomes: A study of 3364 hospitals. *American Journal of Medical Quality, 24*(4), 278–286.

van Eck, N.J., & Waltman, L. (2013). *Manual for VOSviewer Version 1.5.4*, Universiteit Leiden and

Erasmus Universiteit Rotterdam.

Vokurka, R.J. (1996). The relative importance of journals used in Operations Management Research: A citation analysis. *Journal of Operations Management, 14* (4), 345–355.

Waller, M.A., and Fawcett, S.E. (2013a). Click Here for a Data Scientist: Big Data, Predictive Analytics, and Theory Development in the Era of a Maker Movement Supply Chain. *Journal of Business Logistics, 34*(4), 249–52.

Waller, M.A., & Fawcett, S.E. (2013b). Data science, predictive analytics, and Big Data: a revolution that will transform supply chain design and management. *Journal of Business Logistics, 34* (2), 77–84.

Wamba, S. F., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How 'Big Data' can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics, 165*, 234-246.

Wamba, S.F., & Akter, S. (2015). Big data analytics for supply chain management: A literature review and research agenda, *Lecture Notes in Business Information Processing, 231*, 61-72

Wang, G., Gunasekaran, A., Ngai, E.W.T., & Papadopoulos, T. (2016). Big data analytics in logistics and supply chain management: Certain investigations for research and applications. *International Journal of Production Economics, 176*, 98-110.

Wang, L., Zhan, J., Luo, C., Zhu, Y., Yang, Q., He, Y., Gao, W., Jia, Z., Shi, Y., Zhang, S., Zheng, C., Lu, G., Zhan, K., Li, X., & Qiu, B. (2014). BigDataBench: a Big Data Benchmark Suite from Internet Services. IEEE Xplore.

Werdigier, J. (2009). *Tesco, British grocer, uses weather to predict sales*, New York Times, p.1.

Whipple, J.M., & Frankel, R. (2000). Strategic alliance success factors. *Journal of Supply Chain Management, 36* (3), 21–28.

White, M. (2012). Digital workplaces: vision and reality. *Business Information Review, 29* (4), 205–214.

Wong, D. (2012). *Data is the Next Frontier, Analytics the New Tool: Five Trends in Big Data and Analytics, and Their Implications for Innovation and Organisations*. Big Innovation Centre, London.

Yong-Hak, J. (2013). Web of Science. Thomson Reuters. /http://wokinfo.com/media/ pdf/ WoSFS _08_7050.pdf 2013.

Zacharia, Z.G., Sanders, N.R., & Nix, N.W. (2011). The emerging role of the third-party logistics provider (3PL) as an orchestrator. *Journal of Business Logistics, 32* (1), 40–54.

Zhu, H., Madnick, S.E., Lee, Y.W., Wang, R.Y. (2014). *Data and information quality research: Its evolution and future.* Taylor and Francis Group, LLC (pp. 16-20).

Appendix

Table 1A: Journal titles and their abbreviations

Abbreviation	Journal title
JCP	Journal of Cleaner Production
BDR	Big Data Research
TRC-ET	Transportation Research Part C: Emerging Technologies An International Journal
IS	Information Sciences
Scientometrics	Scientometrics
JICS	Journal of Information and Computational Science
IJPR	International Journal of Production Research
ICS	Information Communication and Society
CLSR	Computer Law and Security Review
CFS	Computer Fraud and Security
IMDS	Industrial Management and Data Systems
IEEE-SP	IEEE Security and Privacy
IJPE	International Journal of Production Economics
HBR	Harvard Business Review
DSS	Decision Support Systems
JBR	Journal of Business Research
McKQ	McKinsey Quarterly
IJIM	International Journal of Information Management
JBR	Journal of Business Research
MS	Management Science