Stakeholder sentiment in service supply chains: big data meets agenda-setting theory

Ray Qing Cao¹ · Dara G. Schniederjans² · Vicky Ching Gu³

Received: 12 December 2020 / Accepted: 11 January 2021 / Published online: 5 February 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH, DE part of Springer Nature 2021

Abstract

With growing reluctance to store and disseminate sensitive data throughout a supply chain network, there is a need to understand sentiment of big data and ways of control to achieve greater economic viability in the service-oriented supply chain which reflect a greater focus on knowledge sharing from traditional supply chains. Social network data were collected after referencing a focal corporate media (CM) document. This study provides causal inference by first conducting a CM document search and then a social network post web scrape of postings that reference the CM document while controlling for time and other demographic variables. This study finds salience of the big data topic positively impacts stakeholder sentiment but not when future applications are discussed.

Keywords Text analytics · Sentiment analysis · Digital technologies · Corporate media · Regression

1 Introduction

In recent years, supply chain networks have evolved from simple sequential and linear networks to dynamic processes that require information sharing and visibility (Orenstein 2020). Current trends point to a rebrand of the traditional

✉ Ray Qing Cao
caoq@uhd.edu

Dara G. Schniederjans
schniederjans@uri.edu

Vicky Ching Gu
guvicky@uhcl.edu

¹ Marilyn Davies College of Business, University of Houston Downtown, One Main Street, Houston, TX 77002, USA

² College of Business, University of Rhode Island, 7 Lippitt Road, Kingston, RI 02881, USA

³ College of Business, University of Houston Clear Lake, 2700 Bay Area Boulevard, Houston, TX 77058, USA

“product-oriented” supply chains whose primary focus is on productivity and asset utilization to “service-oriented” supply chains which are more intangible in nature and focus on management of expertise, capacity and knowledge. Arguably, the purpose of supply chain management is to better serve upstream and downstream entities while simultaneously increasing service levels (Chang et al. 2013). This requires seamless integration amongst entities to reduce data transmission time thereby reducing supply lead time and improving service performance (Chang et al. 2013).

The shift of focus from traditional product-oriented supply chains to service-oriented supply chains (Lo 2016) has prompted a focus on the utilization of big data applications for greater information dissemination. Characterized by seamless connectivity among members of the value chain, digital connectivity has given rise to service dominant logic that allows for the rapid customization of both goods and services (Lee and Lee 2019). However, stakeholder reaction to big data applications appear to be mixed, impacting not only the use but the economic value added to the supply chain network. While big data applications in service-oriented supply chains have shown improvements in tracking and tracing shipments, the security and the lack of skills in analytics and management of the data continue to foster mixed sentiment (Neuman 2014). Understanding ways to control this sentiment is paramount to the application and subsequent benefit for the management of service and reduction of variability in supply chain management. One unexplored avenue in which to control for sentiment is through corporate media (CM).

In recent years a proliferation of CM that reference big data in supply chain management provides paramount evidence of both the benefits and drawbacks. Traditionally, data were captured through business transactions involving various stakeholders and processes. In the new era of big data digital transformation, the combination of size, complexity, and advanced analytics has improved applications for national security, marketing, credit risk, medical research, and urban planning applications (Martin 2015). The use of big data analytics in corporations has led to higher financial, market and customer satisfaction performance (Vitari and Raguseo 2019). From a supply chain perspective, products are tagged with sensing devices for tracking and tracing purposes. This includes leveraging large volumes for data to enhance coordination within the supply chain network leading to more efficient communication, agile response to change, and improved operations (Kuo and Kusiak 2019). Digital transformation has also allowed for customer input to delivery and customization thereby enhancing their role as no longer recipients of services but also contributors toward improved service quality (Ahn and Rho 2016). The application of both rapid and large-scale data collection and inter-organizational dissemination of data promises to provide supply chain partners with competitive business intelligence for strategic decision-making and enhance consumer service experience through greater responsiveness (Lau et al. 2017).

The potential to harness the power of big data for supply chain service value is enhancing the “hype” surrounding big data in the media. The continuous growth of data and organizational need to leverage have caused stress on the traditional information systems and techniques, thereby leading to insufficient utilization (Gupta et al. 2019). Thus, the reaction toward big data appear to be “mixed”. Big data have been criticized as a breach of privacy. Both Amazon and Orbitz have been flagged

with price discrimination. In addition, the tracking of movements and shopping habits of stadium spectators utilizing Verizon's Precision Marketing Insight program has been assessed as a "questionable use" of big data (Martin 2015). While collaboration and inter-company communication have never been easier, the platforms that give rise to the processing of information, documents, and key data continue to perpetuate security concerns, especially in pharmaceutical, healthcare, and financial supply chains where data stored can be sensitive. Any perceived unethical use of big data continues to impact sentiment through a lack of understanding of how data will be obtained and shared among supply chain partners. CM on hackers seeking insider trading information from insurance companies, like Anthem Inc. and the unethical use of data from Facebook partners, continue to impact "stakeholder sentiment" (i.e., the cumulative sentiment of stakeholders responding to a particular document).

Calls for research underpin the importance of behavioral research in examining incentive structures, change mechanisms, and social expectations to motivate the use of big data applications (Motiwalla et al. 2019). The organizational change affects organizations internally (and externally) through increasing uncertainty, anxiety, stress, and ultimately resistance, which are critical to the success or failure of change programs (Shah et al. 2017). Organizations have the capability to adjust for changes through altering culture which impacts attitudinal characteristics of internal and external stakeholders, thus impacting readiness or resistance (Shah et al. 2017). One method of impacting culture is through the use of media. Specific mentions of message framing have been suggested as potential impactors of individual attitude (Kohli and Tan 2016). Understanding sentiment surrounding big data in supply chain management, and how to control for sentiment through CM, will aid in achieving supply chain value from such applications. Social media have been discussed as a strategic approach for the further enhancement of knowledge between individuals and organizations, not only impacting organizational learning and dialogue, but in return yielding competitive advantages for the firm utilizing big data to share information throughout the supply chain (Lam et al. 2016). Yet organizations still lack a general understanding of how to craft content such that they can effectively connect and enhance stakeholder sentiment to reduce concerns, thereby mobilizing effective use of big data (Williams et al. 2018).

Thus, this study focuses on CM documents and their respective impact on stakeholder sentiment regarding big data in service-oriented supply chains. Currently, there exists minimal insight into stakeholder sentiment in this area. This study utilizes the collection of textual data from two sources of media (video and practitioner media) which reference corporate communications through a company representative. Through the utilization of a web-scraper, social network post data are collected specifically to analyze stakeholder sentiment of CM on big data in supply chain management. Using sentiment analysis and controlling for various demographic factors, this study determines the associated impact of CM documents on stakeholder sentiment. Unlike previous studies that utilize social network posts to examine cumulative sentiment from media using keyword searches, this study provides causal inference by first conducting a search of CM including news and video media and then a social network post web scrape to collect social network posts that reference the CM document while controlling for time.

This study finds that CM documents that mention big data are positively associated with stakeholder sentiment while CM documents that do not mention big data are not significantly associated with stakeholder sentiment. Additionally, CM documents that discuss future applications of big data are negatively associated with stakeholder sentiment.

The paper is organized as follows: Sect. 2 will present a literature review of big data in supply chain management and how it is projected through media and reciprocal stakeholder sentiments. Section 3 will discuss how the roles of technology adoption theory coupled with agenda setting and a review of psychological bias address how CM documents impact stakeholder sentiment of big data. Section 4 presents the methodology and results of our findings from a regression analysis utilizing demographic variables, topic discussions, and other control variables to analyze the research question as well as post-hoc analyses which discuss actual results from a review of social network posts. Finally, Sect. 5 provides a discussion of the findings as they relate to managerial implications on CM and theoretical implications to inform corporate discussions of big data. Following this are concluding remarks and future directions of this research.

2 Literature review

In this study, we analyze the “stakeholder” perception of big data in CM documents. Borrowing from Freeman (1984), stakeholders can be defined and categorized into three broad groups with equity stakes, economic (or market stakes), and influencer stakes. This model takes into account the majority of individuals who will respond to CM documents in media. They might either be members of the economic or market stakes in that there is an economic interest but no ownership interest in the organizations (i.e., employees, customers, suppliers, and competitors). They might also be an influencer who are individuals who do not have an ownership or economic interest but who have interests as consumer advocates, environmental groups, etc. They might also belong to the group that comprises of individuals with direct ownership (i.e., shareholders, stockholders, minority interest owners).

2.1 Big data

Distinctive characteristics of the service focus (i.e. intangibility, simultaneity, heterogeneity and perishability) in supply chain management suggest a vulnerability toward greater variability (Lo 2016) requiring organizations to restructure accordingly. Big data mark the rise of exploration and utilization of large data sets through the application of advanced statistics to stored electronic communication (Gupta et al. 2012; McAfee et al. 2012; Kache and Seuring 2017). Typically characterized by the four V's: *volume*, *velocity*, *variety* and *veracity* (Goes 2014), recent research has added a fifth “v”, *value*, which incorporates not only the integration of different large data sources and the validity of such data but also the understanding and appreciation of the dual nature of prediction and causality in big data (Baesens et al.

2016). Velocity, as defined by a firm's ability to process data at high speeds, bears the greatest potential for performance improvements (Hofmann 2017) particularly useful for service productivity (Ahn and Rho 2016).

Recent events including a global pandemic (COVID-19) have prompted organizations to rethink current strategies focused on developing robust supply chain networks. With a new need to develop and understand the optimal restocking of critical products, organizations seek to utilize advanced information communication technologies and digital systems. Previous research elucidates this through applications and performance implications of big data adoption from a global perspective (Gawankar et al. 2020). But the adoption of big data begins with the individual and/or user perceptions of the technology.

2.2 Application of big data in firms

Organizations have faced both opportunities and challenges when seeking to enhance individual value through big data in supply chain management (Setia and Patel 2013). Big Data analytics are increasingly used to improve business outcomes as the collaboration resulting from technology has been found to be positively related to financial performance (Hwang and Kim 2019). Several new sources of data over the past two decades has resulted in the availability of large amounts of user data in the digitization of commerce (Motiwalla et al. 2019). Big data sources have helped to enhance supplier insight into user behavior. For example, big data have been used to personalize product offerings, prices, and display assortments through the utilization of downstream attributes including geo-localization, demographic information, and the time of day (Cohen 2017; Zhan and Tan 2018). Other benefits involve, customization in electronic retailing (Thirumalai and Sinha 2009), the use of metadata for enterprise document searches (Schymik et al. 2015) targeted promotions and campaigns which reduce the costs involved in procurement, the use of service robots in the hotel industry impacting assurance and reliability (Chiang and Trimi, 2020), as well as predictive modeling and analytics aiding product and process modeling (Cohen 2017; Duan et al. 2018). These capabilities have prompted service inseparability allowing for downstream entities to actively participate in service production and delivery which can lead to increased customer satisfaction and service productivity (Ahn and Rho 2016).

From a logistics service perspective, various research has incorporated and emphasized the role of technology and analytical support as a core competency providing performance outcomes (Hwang and Kim 2019). Logistics applications include vehicle equipment with on-board diagnostic devices for monitoring driver behaviors in real time (Stankovic 2014; Baesens et al. 2016), which allows for vehicle managing companies and insurance companies to make decisions through predictive modeling. Utilizing a Delphi technique, Kache and Seuring (2017) identify other various opportunities and challenges for big data including information management, operations efficiency and maintenance, supply chain visibility and transparency, product and market strategy, product traceability, supply chain inventory

optimization, risk evaluation, etc. all to enhance the consumer service level, trust in the organizations through transparent, ethical practices, and experience.

2.3 Implications of big data

With a focus on service to satisfy downstream members' needs for financial, strategic and marketing benefits, firms that proactively design service systems to country variability can open up market opportunities and decrease costs incurred with transactions (Lo 2016). However, these improvements can come at a cost. Information privacy has been referred to as one of the most important issues facing individuals since the advent of e-commerce (Smith et al. 1996) and is now becoming more important as digital technologies give rise to big data applications. Challenges including a lack of governance to control big data efforts across supply chain networks have led to organizational reluctance from many (Kache and Seuring 2017). More prevalent in recent literature is cybersecurity issues. Rapid growth of concerns about revealing proprietary information has, in some cases, led to a reluctance by retailers to share information across the supply chain despite the advantages big data bring to the networked economy (Menon and Sarkar 2016).

Understandably both positive and negative aspects to big data applications exist. However, literature suggests that it is likely no area of business activity will remain unaltered by this paradigm (Kache and Seuring 2017). Ways in which corporations can project and communicate with stakeholders on big data to reduce concerns and optimize sentiment is still elusive in literature. Recent literature focuses on applications of big data including social media use in supply chain setups (Chae 2015), allocation of resource decisions (Saboo et al. 2016), and early warning systems or anomaly detection (Breuker et al. 2016). Other literature discusses the implications of big data. For example, articles dating back several years ago discuss leveraging analytics to optimize a company's supply chain network (Davenport 2006). Trkman et al. (2010) and De Oliveira et al. (2012) examine analytics on supply chain performance. Hazen et al. (2014) look at the impact of data complexity on the importance of monitoring and controlling for data quality. Other implication studies focus on optimizing freight in the transport supply chain (Harris et al. 2015), value portfolio optimization in manufacturing (Opresnik and Taisch 2015), supply chain innovation competencies (Tan et al. 2015), and logistics management in the automotive industry (Zhong et al. 2015).

While the research on applications and implications of big data provide various avenues for progression in this domain, there is little insight regarding stakeholder sentiment of big data in supply chain management. Data leakage both within and outside a supply network are sometimes unforeseeable. Corporate transparency has been shown to impact investor sentiment (Firth et al. 2015). Additionally, media publicity has been shown to have an impact on individual perceptions on the importance of corporate social responsibility behavior (Elias 2004). This study explores if CM can be utilized as tools to enhance stakeholder sentiment for promoting big data applications in supply chain management.

2.4 Agenda-setting theory

Before technology adoption takes place, individuals need to first understand the function as well as develop an attitude toward the technology (Sun 2018). Communication through channels is fundamental for technology perception which will lead to either the decision to adopt or not to adopt. This paper deals with CM as one fundamental channel among members of a social system. Agenda setting was popularized in the early 1970s when McCombs and Shaw (1972) found that issue “salience” (defined as the amount of attention granted to a particular issue) in media is strongly associated with stakeholder perceptual importance of specific topics. These findings reiterated previous scholarly undertones that media might not necessarily tell stakeholders how to think about a topic but rather what topics to think about. That is, the basic premise of agenda setting is that media transfer the salience of topics from the media agenda to the public agenda. Media lead the public to assign importance to public issues (Zhu and Blood 1997), but does not accomplish this by telling the public about the importance of one issue over another. Instead media give certain issues more prominence and preferential treatment through frequency of coverage and prominent positions.

The connections between technology adoption and agenda setting can be situation specific. Both have been tied to organizational behavior theory. Specifically, previous research in technology adoption identified five determinants of rate of adoption: relative advantage, compatibility, complexity, trialability and observability (Rogers 2003). Agenda setting encompasses concepts from individual perception of “obtrusiveness” (i.e. personal experience with an issue), “need for orientation” (relating to a person attending to an agenda) and “salience” (Chernov et al. 2011). Both media and technology can generate individual perceptions as they relate to subjects and objects. The ties between agenda setting and technology adoption theory can be expanded upon utilizing psychological concepts thus building value for organizational communication and subsequent adoption and use of technology. Chew et al. (2012) discuss “familiarity bias”, in which individuals tend to focus on adverse scenarios in evaluating any defection from the status quo. The fear that propagates from the unfamiliar tends to impact how we react during times of decision including our basic stances in regard to CM. At times the primary emotions can manifest into secondary emotions, including anger from powerlessness in the face of uncertainty (Seltzer and Mahmoudi 2013). Individuals tend to compensate for this by utilizing their own communication mechanisms to rebuild self-esteem that is hindered due to the primary emotions perpetuated by the unknown. Related to the concept of “complexity” is agenda setting’s “obtrusiveness. The difference between the two concepts lies in the experience of the individual. Complexity is purely based upon perception regardless of experience vs. obtrusiveness which incorporates actual experience. Research as it relates to complexity clearly demonstrates the negative impact on technology adoption (Sun et al. 2018). However, mixed results appear regarding the impact of obtrusiveness on agenda setting effects. Some research has found that media agenda setting effects are stronger for unobtrusive issues since audiences rely on the media for information on these issues versus other sources for those issues that are obtrusive (Zucker 1978). Alternative literature reports otherwise. For

example, Yagade and Dozier (1990) found that the media agenda setting is enhanced by “concreteness” and reduced by “abstraction”. While the connection between obtrusiveness and agenda setting has been and is currently being discussed, limited knowledge exists on the reason behind these findings and specifically understanding the psychological confounding factors that may play into the findings.

3 Hypotheses and research model

Technology adoption is defined as the process where a technology is communicated through channels among members of a social system (Rogers 2003). Before adoption is to take place, individuals need to first understand the function and develop an attitude toward the technology (Sun 2018). Communication through channels is fundamental for technology perception, which will lead to either the decision to adopt or not to adopt. “Salience” in media is strongly associated with stakeholder perceptual importance of specific topics relevant to technology. These findings reiterated previous scholarly undertones that media might not necessarily tell stakeholders how to think about a topic but rather what topics to think about. Media transfers the salience of topics from the media agenda to the public agenda. This leads the public to assign importance to public issues (Zhu and Blood 1997), but does not accomplish this by telling the public about the importance of one issue over another. Instead, media give certain issues more prominence and preferential treatment through the frequency of coverage and prominent positions.

Individuals develop opinions, beliefs, and views on technology utility based on messages received externally. Considered as one of the most consistently critical predictors of information technology adoption (Sun et al. 2018), the relative advantage is a perceptual measure that captures an individual’s perceptions of the benefit in terms of value and competitiveness of the technology. This can be with or without experience but must be based upon a conceptualization of the technology in comparison with its predecessors.

Corporate communications in the media can enhance the salience of relative advantage, thereby impacting stakeholder sentiment brought about by corporate communication on a topic. *Cumulative stakeholder sentiment* can thus be impacted and is defined here as the mean sentiment score of all stakeholders responding to a CM document. As such, the salience of a technology within a CM document is purported to impact sentiment as outlined in hypothesis 1.

H1. Salience of “big data” topic within the CM document will have a significant positive association with cumulative stakeholder sentiment.

Content of CM can be either topic, or the syntax utilized. Specifically, sentiment can be relayed throughout the article that indicates either neutral, positive, or negative opinions on a topic. Previous research has assessed the effect of social and conventional media sentiment on firm performance. Specifically, Yu et al. (2013), identify significant positive impacts of the blog and Twitter sentiments on firm performance. Ecommerce research also suggests that the way information is presented impacts e-customer satisfaction (Pieria et al. 2017). Indicative of a shareholder sentiment, the positive impacts of media sentiment indicate document sentiment

(i.e. associated sentiment of the CM document) will likely impact stakeholder sentiment. Clapman et al. (2019) indicates the sentiment of messages and news which are highly popular can achieve higher investor attention. As such, we hypothesize:

H2. Document sentiment will have a significant positive association with cumulative stakeholder sentiment.

Yu et al. (2013) indicates that while social media have a stronger impact on financial performance than conventional media, both will have a joint effect on the market. That is, the more discussed a particular topic is, the more likely the influence will be more significant. This indicates that the volume of influence is more with salience than with sentiment (Yu et al. 2013). Additionally, Waters (2013) found that when organizations were able to place spokespeople in the newscast to talk about efforts, the donations increased substantially. This alludes to the implication that when corporations make themselves more salient in media this can help to enhance their performance through agenda setting. Bridging this with technology adoption theory, to make a technology topic or organization more concrete, it suggests a potential impact on the stakeholder's sentiment towards that technology.

Of course, big data are far different from humanitarian relief donation efforts. Intuitively, stakeholder attitudes will likely differ between perceptions of relief efforts over corporate efforts to enhance perceptions of relative advantage with big data. Yet seeking an understanding of what impacts stakeholder sentiment, this study relies on the empirical findings of Waters (2013) as well as the basic tenets of technology adoption theory and agenda-setting to formulate the following:

H3. The salience of "big data" topic within the CM document will have a stronger positive association with cumulative stakeholder sentiment than the document sentiment.

In the age of digital media, where individuals are being overrun by vast amounts of information from organizations, current studies suggest high levels of skepticism (Elving 2013) despite the service benefits. While salience may enhance the likelihood of a stakeholder thinking about a specific topic, the sentiment toward that topic is likely to be controlled by the stakeholder. Ultimately, this drives the question of how stakeholder psychology may impact the sentiment of big data through corporate communications in media.

Chew et al. (2012) discuss "familiarity bias", in which individuals tend to focus on adverse scenarios in evaluating any defection from the status quo. The fear that propagates from the unfamiliar tends to impact how we react during times of decision including our basic stances in regard to CM. At times the primary emotions can manifest into secondary emotions, including anger from powerlessness in the face of uncertainty (Seltzer and Mahmoudi 2013). Individuals tend to compensate for this by utilizing their own communication mechanisms to rebuild self-esteem that is hindered due to the primary emotions perpetuated by the unknown.

But how do organizations control for the psychological bias against the unknown? An emergence from agenda setting is the topic of "framing" in which a topic can be framed in such a way that issue attributes are emphasized in order to control for the sentiment (Neuman et al. 2014). In basic terms, framing is the selection of a restricted number of related attributions when a particular topic is discussed (McCombs et al. 2014). A topic such as big data in service-oriented supply chains

needs to be framed in such a way that it reduces the unknown and resulting fear and anger associated with the topic due to a general lack of understanding of its implications. Utilization of concepts from technology adoption theory including the clarifying stage of diffusion (i.e., Miles 2012) involves helping individuals understand the complexity through correcting any misunderstanding and clarifying the purpose, meaning and function of the technology. Specifically, instead of focusing on potential avenues for the technology and how the organization can fit the purpose of the technology (i.e. a focus on potential Future Applications), this stage seeks to suggest “current applications” for how a technology can fit the needs of the organization (Miles 2012). Based on this reasoning, this study hypothesizes the following.

H4. Discussion of “Future Applications of Big Data” in a CM document will be significantly negatively associated with cumulative stakeholder sentiment.

H5. Discussion of “Current Applications of Big Data” in a CM document will be significantly positively associated with cumulative stakeholder sentiment.

4 Research methodology

Researchers have typically used surveys and experiments to understand the role of CM in influencing stakeholder reactions to communication dynamics. Unfortunately, these often confront distortion from self-reported data (Neuman et al. 2014). In an effort to obtain current mainstream data and to gather from a variety of sources, this study uses news sources (i.e., practitioner magazines including *Supply Chain Quarterly*, *Supply Chain Management Review*, *DC Velocity*, *Inbound Logistics*, *Logistics Management*, *Supply Chain Dive*, *Modern Materials Handling*) as well as video sources (i.e., YouTube and Google Videos including company respondent interviews, company discussions on big data applications, panel discussions, and company representative speeches).

4.1 CM document collection and data cleansing

This paper refers to the practitioner and video media as “CM documents”, specifically referencing corporate communications through media on two platforms (i.e., practitioner publications and video media). While not all media are reflective of “corporate communications” in this study, any CM document that did not have a company spokesperson interviewed or quoted was removed.

Data acquisition and subsequent cleansing of data occurred over a four-month period from January 2018 through April 2018. Data were collected and cleansed by one professor and two graduate student assistants whose expertise lies in social media and supply chain management.

The research question in this paper addresses whether organizations can enhance the salience of big data in supply chain management through discussions in CM documents and what organizations can discuss in these documents to enhance stakeholders’ sentiment on big data in supply chain management. This question required obtaining CM from differing sources. Practitioner magazines as

well as video sources (i.e., YouTube and Google Video) were utilized to obtain the CM documents. Article textual data were initially collected through a cloud web scraper through Google Chrome. This included accessing the practitioner magazine websites and signing in through a personal or university subscription and subsequently running the web scraper code to collect data including: “title”, “date”, “author”, “affiliation”, “main text”, and “abstract text”. A different code was used for each publication based on the layout.

Video data were manually collected by a co-author and two hired graduate assistants. This involved manually collecting “title”, “date”, “author”, “affiliation”, “text” and “summary” data when available through the YouTube and Google Video search. Textual data were collected through closed captioning of the data provided via YouTube or Google Video. This involved collecting closed captions from the website as well as having two graduate assistants review and, when necessary, adjust the closed captioning to ensure accuracy. Some of the videos obtained required graduate assistants to re-write the closed captions due to a lack of accuracy.

Table 1 presents the search string as well as the specific periodicals referenced and exclusion criteria. The search string was created from our research question based on current research. Our definition of “big data” can be operationalized in various ways and is a direct result of digital technologies. Thus, instead of focusing on CM that just reference “big data”, this study utilizes a broader term to account for the technologies that produce “big data”.

Following the work of Durach et al. (2015), the search string was initially developed by the co-authors of the paper and then revised by a subject librarian at one of the co-author(s) institutions who had knowledge of digital technologies and was able to adjust based on the needs of the database. An initial period of January 1 2011 to January 1 2018 was chosen based on the emergence of the term “big data” (Gartner 2011). However, since the authors were also collecting social network post data referencing the CM documents, a period of two years from the social network collection, was utilized to help control for relevancy of the topic as well as current CM document coverage. That is, CM documents from the previous year were utilized for our final analysis. Following this, social network data referencing the specific CM documents were collected. Relevancy was ensured through a title search as well as an overview of the summary or abstract textual data. After the initial collection, graduate assistants manually went through each title and summary textual data to ensure its relevance to the research question. Where summary or abstract textual data were not made available, two co-authors read through the textual data to ensure its relevance to the research question.

Through this process, articles were coded based on the technology topic discussed. This was done through a search of key terms both in the title, summary and main text for terms like “big data”, “big-data”, “data analytics”, “data analysis (and process*ing)”, “data processing”, “data mining”, “big analytics”, “data science”, “extreme info*rmination”. These keywords were developed based on an initial literature search referencing academic articles and news articles. Other categorizations of technology-specific terminology to digital technologies were also coded but not used in this analysis. The initial results as well as the frequency of results from the

Table 1 Textual sources and exclusion criteria

Type	Source	Search string	Periodicals/videos referenced	Exclusion criteria	Frequencies
Video	YouTube; Google Video	"supply chains" OR "supply chain" OR "supply chain management" OR "suppliers" OR "supply network" OR "logistical network" OR "demand chain" OR "supply management" OR "reverse logistics" AND "Fourth industrial revolution" OR "smart factory" OR "Connected Enterprise" OR "Industrie 4.0" OR "Industry 4.0" OR "smart industry" OR "smart manufacturing" OR "Industrial internet" OR "re-industrialisation" OR "Digitalisation" OR "Digital" OR "Digitisation" OR "Internet of things" OR "IoT" OR "location detection technologies" OR "location detection technology" OR "Advanced human-machine interfaces" OR "fraud detection" OR "3D print" OR "additive manufacturing" OR "smart sensors" OR "big data" OR "augmented reality" OR "cloud computing"	YouTube, Vimeo, Company	Advertisements, Non-English (without English subtitles), No subtitles	Initial: 11,200 Company use and focus on digital technologies: 360 Date Range: 340 Repeats on similar platforms: 120 English or Subtitles: 74
News	Practitioner Magazines		<i>Supply Chain Quarterly</i> , <i>Supply Chain Management Review</i> , <i>DC Velocity</i> , <i>Inbound Logistics</i> , <i>Logistics Management</i> , <i>Supply Chain Dive</i> , <i>Modern Materials Handling</i> , <i>TechCrunch</i> , <i>CIO Magazine</i> , <i>Infosecurity Group Magazine</i> , <i>Wired</i> , <i>Fast Company</i> , <i>ComputerWorld</i>	Advertisements, New Company Announcements, Interviews (Removed text not related to industry 4.0)	Initial: 143,222 Company use and focus on digital technologies: 10,347 (title summary search) Date Range: 538 Repeats on similar platforms, advertisements, new company announcements, interviews: 353

use of our exclusion criteria are depicted in Table 1. These numbers reflect the initial search conducted in January 2018.

4.2 Social network collection and data cleansing

Social network data collection occurred during January 2018 through March 2018. As mentioned previously, to control for the time effect on sentiment, we utilized social network data only during January 2018. Similarly, we only used CM that happened just prior (i.e. January 2016 through December 2017). Controlling for the number of days since the publication versus the post helped to control for the impact that time has on big data application sentiment.

RapidMiner, a machine learning, text mining, and predictive analytics software, was used to collect social network data. Utilizing the Twitter connector, Twitter data were directly collected through RapidMiner Studio. Phrases, tweets, and user profile information were collected through an authentication mechanism (OAuth 2.0). OAuth 2.0 enables applications to obtain access to user accounts through an HTTP service. This is done through a streaming API that fulfills both the resource and authorization server roles. To receive more complete coverage of social network data obtained from each CM document, several queries were utilized per CM document. This included queries pertaining to the “title”, “author”, “affiliation” and when necessary elements from the “abstract text” or “summary” were utilized to gather social network data. After the initial collection of the social network data, the co-authors manually went through and deleted social network posts that did not specifically reference the article. Additionally, social network posts that included URLs not related to the CM document were removed. The deletions were later reviewed by graduate assistants to ensure accuracy.

In totality, sixty-seven CM documents (as defined by the inclusion of a firm representative) had 1408 social network posts. Data cleansing involved the elimination of non-English social network posts, non-letter characters, URLs, mentions and re-tweet identifiers (Pournarakis et al. 2017). Overall a total of 508 social network posts were utilized for further analysis. Following this demographic data were also collected from each of the social network stakeholders. This included the utilization of RapidMiner’s “Get Twitter User Details” operator through a Twitter connection to specify the Twitter account for the API access. Additionally, other demographic characteristics were captured manually through a social network account by utilizing a username and/or ID searches. Initial characteristics included: Twitter Username, ID number, screen name, description, URL, location, followers, friend, favorites, tweet, language, time zone, and profile image. All twitter data and CM document data were stored in multiple Excel Spreadsheets.

4.3 Sentiment analysis

Sentiment analysis has been used to effectively determine customer responses by extracting sentiment in text data deriving valuable information (Kim and Hong 2020). Data preparation for subsequent sentiment analysis for both the

CM documents and the social network posts was based on the work of Sul et al. (2017), who developed a model that combines topic and sentiment classification to elicit influential subjects from consumer perceptions in social media. Additionally, this study references Pournarakis et al. (2017) who empirically apply a model to conduct an analysis of over 280,000 tweets related to a specific topic (i.e., Uber transportation network) with the goal of providing insights into awareness and meaning in brands. The authors introduce a model that evokes subjects from consumer perceptions in social media. As the purpose of this study examines a similar context (stakeholder perceptions of big data), without creating topic classifications, this study utilizes the data preparation and subsequent sentiment classification steps that Pournarakis et al. (2017) define.

The first step involved data cleansing and pre-processing. This step aims to prepare data for various data mining tasks. First, text tokenization (i.e. text segmentation and removal of stop words), and stemming (i.e. obtaining root forms of derived words) (Pournarakis et al. 2017; Swain and Cao 2019) were applied. Both the raw CM documents and the social network data are separated into databases and the final version of each was formed as a set of purified documents (i.e. involving the deletion of blanks, duplicated records and stored as a refined corpus in a computable format) (Swain and Cao 2017). Each document (i.e. represented as either a complete CM document or a social network response) is represented through a collection of words. After the first stage of data clearing pre-processing, sixty-seven CM documents were used for the classification stage as well as 508 social network posts. This was conducted through a lexicon (WordNet-SentiWordNet) (Miller 1995). This is a sentiment lexicon that is derived from WordNet where words are clustered into groups of synsets. Each CM document and social network post were represented by a collection of words and then utilized for sentiment classification.

The second step involved sentiment classification through the utilization of support vector machines (SVM) which are non-linear classifiers operating in higher dimensional vector spaces than the original feature space of a given data set. SVM's training process involves a quadratic minimization problem in the context of a binary classification. Through the utilization of the Gaussian kernel function, performance evaluation of the SVM classifier was measured through a tenfold cross-validation process on an equally balanced set of 50 social network posts and 20 documents. Each fold split the pre-labeled samples into a 95% training-data/5% testing data ratio. The study utilized this ratio based on the size of the data set. This split allows for greater accuracy and reliability (Pournarakis et al. 2017). The first subset of data instances was utilized to build a classifier and the latter was used to assess its ability to suggest sentiment polarity. Mean accuracy per fold was 0.7092 (± 0.179), mean recall per fold was 0.7083 (± 0.194).

The sentiment of the rest of the data were conducted through an exploitation of the complete set of pre-labeled data instances. Each CM Document (*Document Sentiment*), and the social network posts or (*Individual Sentiment*), Y_d , was given a soft decision value $\in \{-1 \leq d \leq 1\}$ and $\in \{-1 \leq Y_d \leq 1\}$ where

$$\begin{cases} d = 1, \text{for positive sentiment classification.} \\ d = -1, \text{for negative sentiment classification.} \end{cases}$$

4.4 Variables

Since analysis occurred at the CM Document level, and one document would have numerous social network posts, further data preparation for “*Stakeholder Sentiment*” or Y_d was necessary. Initial values of “STAKEHOLDER_SENTIMENT”, Y_d , were computed by taking the average sentiment score for each social network post (or document). If a stakeholder posted more than once about a CM document, the average was used as that stakeholder’s cumulative score. No stakeholder repeats were noted for different CM documents, so this was not factored into the research model. In the case of repeated social network posts by the same stakeholder per one CM document, $i_{1,1}$ represents stakeholder 1’s first social network post for sd_n and $i_{1,2}$ represents stakeholder 1’s second social network post for sd_n and so on. This computation of γ_d was only utilized for three of the CM documents in which there was a stakeholder that posted more than once. Table 2 presents a summary of our variables utilized in this study. The independent variable, “*Big Data*”, was represented as $x_{1d} \in \{0,1\}$ for every CM document d . The terms utilized for a classification of $\{1\}$ included: “big data”, “big-data”, “data analytics”, “data analysis (and process*ing)”, “data processing”, “data mining”, “big analytics”, “data science”, “extreme info*rmation”.

In order to determine the strength in relation to “*Big Data*”, “*Document Sentiment*” is another variable that we input into the model, which is represented as

$$x_{2d} \in \{-1 \leq x_{2d} \leq 1\} \quad (1)$$

where x_{2d} is the associated sentiment of the CM document d . While decision values close to $\{0\}$ can be indicative of a neutral sentiment, this is not explicitly presented in the SVM classifier during training. The grouping variable, “*Current Applications*”, is represented as $x_{3d} \in \{0,1\}$ for every Big Data CM document. The other grouping variable, “*Future Applications*”, is represented as $x_{4d} \in \{0,1\}$ for every Big Data CM document. Variables x_{1d} , x_{3d} , x_{4d} were coded by two independent coders (one co-author and a graduate student assistant). All dummy-coded variables were mean-centered for interpretation in the interaction model. Several control variables were utilized based on previous agenda-setting literature and adjusted to fit the document level analysis in this study. Specifically, agenda-setting purports that the type of media used can impact stakeholder perception and attitude (Singer 2018). Since this study utilizes both video and print media, dummy variables were created for the categorical variable of “*Type of Media*” ($\mathcal{C}_{1dj}$) with video and print documents. Another variable this study controls for is location, specifically the country of the stakeholder. Demographic data were collected from the stakeholders in seven different country locations, all of which differed in their economic development and national conflicts. Thus, these analyses called for a dummy coding of each country into a variable labeled “*Location*” ($\mathcal{C}_{2dj}$) and regression onto the dependent variable.

Table 2 Summary of variables and definitions

Variable	Description	Symbol	Calculation
Stakeholder sentiment	Cumulative sentiment of stakeholders responding to CM document	Y_d	$\bar{x} \{sd_{i1}, sd_{21}, \dots, sd_{n1}\}$ Or $\bar{x} \left\{ \left[\bar{x} (sd_{i1,1}, sd_{i1,2}, \dots, sd_{i1,n}) \right] \dots \left[\bar{x} (sd_{n1,1}, sd_{n1,2}, \dots, sd_{n1,n}) \right] \right\}$
Document sentiment	Associated sentiment of CM Document	x_{2d}	$\in \{-1 \leq x_{2d} \leq 1\}$. Where x_{2d} is the associated sentiment of the CM document d
Big data	Use of big data in CM document	x_{1d}	$\in \{0,1\}$ for every CM document d. Where $\begin{cases} x_{1d} = 1, \text{when big data terms are referenced.} \\ x_{1d} = 0, \text{when big data terms are not referenced.} \end{cases}$ $\in \{0,1\}$ for every Big Data CM document. Where
Current applications	Inclusion of current applications in Big Data CM document. Text in the CM document is reflective of current oriented activities	x_{3d}	$\begin{cases} x_{3d} = 1, \text{when current applications of big data are referenced.} \\ x_{3d} = 0, \text{when current applications of big data are not referenced.} \end{cases}$ $\in \{0,1\}$ for every Big Data CM document. Where
Future applications	Inclusion of future applications in Big Data CM document. Text in the CM document is reflective of future oriented activities	x_{4d}	$\begin{cases} x_{4d} = 1, \text{when future applications of big data are referenced.} \\ x_{4d} = 0, \text{when future applications of big data are not referenced.} \end{cases}$ $\in \{0,1\}$ for every CM document. Where
Type of media	Specification of video or print CM document	C_{dj}	$\begin{cases} 1, \text{for video CM document.} \\ 0, \text{for print CM document.} \end{cases}$
Location	Country of the stakeholder	C_{ij}	$\in \{0000n, 0000n0, 000n00, 00n000, 0n0000, n00000\}$ for every CM document. Where n = number of individuals from country, c.
Experience	Ratio of the stakeholders who had job experience within operations, supply chain management or information systems fields	C_{ij}	$\in \{0.00 - 1.00\}$, ratio of stakeholders who had experience in operations, supply chain management or information systems (1), versus those that did not (1)
Amount of coverage	Total number of media (i.e. big data) occurring in the month of the social network post	C_{dj}	$\in \{0 \leq n\}$, count of big data media η occurring in month i of the social network post
Friends	Average number of friends of the cumulative group of stakeholders for every CM document	C_{ij}	$\in \{0 \leq n\}$, average number of friends for each stakeholder, for every CM document, d $\bar{x} \{count(friends_i), count(friends_2), \dots, count(friends_n)\}$
Specificity	Specificity of digital technology topic in article	C_{dj}	$\in \{0 \leq n\}$, sum of the number of digital technologies discussed in CM document

This study also controlled for experience within operations, supply chain, or information systems fields. This was done by assessing demographic data collected by a public API. In some cases, the career title was not made clear based on the initial data collection. In this case, we conducted an external search through primarily LinkedIn but also the affiliated social network. The timing of the social network post was accounted for by assessing the time of the social network post with the time of the job provided on the social network platform. Dummy codes were created for “*Experience*” if the stakeholder indicated a career in Operations, Supply Chain or Information Systems fields. Approximately 1.2% of stakeholders did not have any employment information accessible to the researchers. In this situation, the data point was removed. The ratio of experience to no-experience was then utilized as the cumulative score for “*Experience*” (C_{3dj}) for the document, d . Further, this study controlled for coverage. To control for this and timing effects, this study assessed the popularity of the digital technology topic covered in that month. Specifically, we counted the number of big data CM documents η occurring in month i of the social network post. For each CM document, the average of the total number of social network posts was used as the variable “*Amount of Coverage*” (C_{4dj}).

This study also controlled for the number of friends a social network poster has. Agenda setting is particularly popular among stakeholders with a large number of “followers” or “friends”. A continuous variable, “*Friends*” (C_{5dj}), was created and defined as the average number of friends utilized for the cumulative score for the document, d .

Finally, “*Specificity*” (C_{6dj}) of the CM document on a particular technology was controlled for. This included utilizing a keyword search of the CM document based on synonyms of digital technologies utilized in the initial literature review. Technologies, including cloud computing, sensors, autonomous vehicles, 3D-printing, etc., were counted and reflected in a continuous variable indicating the specificity of the topic to the article.

4.5 Bayesian regression

To assess the significance of the hypotheses presented in the research model, this study employs Bayesian regression with Gibbs sampling. The Bayesian regression for the hypotheses testing helps handle more complex models and facilitates decision making when additional information is gathered. Additionally, Bayesian analyses can be useful in smaller samples (Hahn et al. 2020). Utilizing the Bayesian regression, we form a likelihood function based on an asymmetric Laplace distribution. Given the uniqueness of the dependent variable’s distribution, we are able to model within the specified distribution and estimate a set of percentiles indicating the direction of the effect.

This study first examines the direct relationship between “*Document Sentiment*” and “*Big Data*” on “*Stakeholder Sentiment*” after controlling for six variables as outlined in Table 4. That is, this study tests the following equation:

$$Y_{dj} = \alpha + \beta_0 X_{1dj} + \beta_1 X_{2dj} + \beta_2 X_{1dj} * F_{1dj} + \beta_3 X_{1dj} * A_{1dj} + \gamma C_{dj} + \delta_j \quad (2)$$

where Y_{dj} refers to the cumulative stakeholder sentiment for document d at level j . The variable, X_{1dj} refers to the dummy coded “BIG_DATA” variable for document

d , X_{2dj} refers to the sentiment of document d . The variable F_{1dj} refers to the future applications for document d at level j and A_{1dj} refers to the future applications for document d at level j . $\mathcal{C}_{dj}$ represents the six control variables for the document, d . Tables 3 and 4 shows the summary statistics as well as the correlation table.

Table 5 shows the posterior summaries of the estimated parameters β_0 , β_1 and γ . The sample mean of the output chain for the intercept β_2 is 0.0514 with the 95% credible not containing 0 and instead of being positive, indicating a high probability that the direction of the effect of *Big Data* are positive. This indicates support for hypothesis 1. The results do not support hypothesis 2, as the coefficient of the variable *CM Document Sentiment* is 0.0190 with the 95% credible interval containing 0, indicating the direction of the effect of *Document Sentiment* is ambiguous. The analysis conducted in this study did find more support for the association between the inclusion of a *Big Data* topic in the document versus the *Document Sentiment*. However, since hypothesis 2 was not found to be significant the analysis does not provide further support for hypothesis 3.

Bayesian analyses were conducted to determine the impacts of both *Future Application* and *Current Application* communications in *Big Data* documents. Results indicated support for hypothesis 4, where the coefficient variable of *Big Data* with *Future Applications* is -0.0835 (-0.1422 , -0.0128 95% credible interval) indicating the direction of the effect is negative. Results do not show support for hypothesis 5 with the coefficient variable of *Big Data* with *Current Applications* is 0.000886 (-0.2674 , 0.2225 95% credible interval) indicating the direction of the effect is ambiguous.

Table 3 Summary statistics

Variable	Reference	Mean	Standard Deviation
Big data	BD	0.10769	0.31240
Type of media (v)	TMV	0.27692	0.45096
Type of media (print 1)	TMP1	0.03077	0.17404
Specificity	S	1.27692	1.37509
Document sentiment	DS	0.50769	0.75256
Experience	EXP	3.58462	6.92796
Friends	FR	15,685	35,873
Demo_NA (Location)	L0	0.1538	0.1454
Demo_SA (Location)	L1	0.12308	0.54508
Demo_EUR (Location)	L2	0.52308	0.23935
Demo_ME (Location)	L3	0.26154	0.66795
Demo_AFR (Location)	L4	0.01538	0.12403
Demo_ASIA (Location)	L5	0.12308	0.33108
Demo_AUST (Location)	L6	0.10769	0.35895
Amount of Coverage	COV	10.92308	11.69710

Table 4 Correlations

<i>BD</i>	<i>TMV</i>	<i>TMP1</i>	<i>TMP2</i>	<i>S</i>	<i>DS</i>	<i>EXP</i>	<i>FR</i>	<i>L0</i>	<i>L1</i>	<i>L2</i>	<i>L3</i>	<i>L4</i>	<i>L5</i>	<i>L6</i>	<i>COV</i>
<i>BD</i>	1														
<i>TMV</i>	0.11773	1													
<i>TMP1</i>	-0.06190	-0.11026	1												
<i>TMP2</i>	-0.09099	-0.92828**	-0.26726*	1											
<i>S</i>	0.25685*	-0.07520	-0.10145	0.11087	1										
<i>DS</i>	0.09611	0.26987*	-0.00184	-0.26096*	0.01301	1									
<i>EXP</i>	-0.15227	0.28745*	-0.00219	-0.27788*	0.18448	-0.07880	1								
<i>FR</i>	0.04157	0.30728*	-0.05083	-0.27890*	0.11799	0.08322	0.75189***	1							
<i>L0</i>	-0.12631	0.02121	-0.03288	-0.00826	0.12000	-0.01850	0.58209***	0.35885*	1						
<i>L1</i>	-0.07905	0.17700	0.12417	-0.21807†	0.01636	-0.00234	0.67577***	0.74773**	0.11078	1					
<i>L2</i>	-0.06679	0.34804*	-0.02216	-0.32915*	0.08466	0.07708	0.81776**	0.88719**	0.22848†	0.81692***	1				
<i>L3</i>	-0.06221	0.32640*	0.06410	-0.34044*	0.26015*	-0.01961	0.65863***	0.45450**	0.33859*	0.42519**	0.51376**	1			
<i>L4</i>	-0.04343	0.20199	-0.02227	-0.18750	0.34107*	-0.25238*	0.53487***	0.12362	0.44689**	-0.02844	0.10332	0.51617**	1		
<i>L5</i>	-0.13015	-0.02254	0.20442	-0.05463	-0.04171	0.18427	0.28150*	0.28945*	0.31653*	0.34766*	0.26512*	0.06413	-0.04683	1	
<i>L6</i>	-0.10504	-0.09059	-0.05387	0.10798	-0.09302	-0.03204	0.26331*	0.26328*	0.02927	0.41035**	0.29098*	0.01103	-0.03779	0.14968	1
<i>COV</i>	0.09210	-0.01071	-0.15233	0.06738	0.91255**	0.04178	0.18721	0.12566	0.17279	0.03092	0.09378	0.24060†	0.32392*	0.04686	-0.07987

† < 0.10, * < 0.05, ** < 0.001

Table 5 Posterior summaries and intervals

Variable	Mean	Standard deviation	95% HPD interval	
$\beta_0 X_{1dj}$	0.2765	0.1405	0.0109	0.5589
$\beta_1 X_{2dj}$	0.0514	0.0199	0.0203	0.0907
$\beta_2 X_{1dj} * F_{1dj}$	-0.0835	0.0350	-0.1422	-0.0128
$\beta_3 X_{1dj} * A_{1dj}$	0.000886	0.1236	-0.2674	0.2225
$\gamma_1 C_{1dj}$	0.0112	0.0296	-0.0248	0.0707
$\gamma_2 C_{2dj}$	0.0963	0.0428	0.0105	0.1721
$\gamma_3 C_{3dj}$	-0.0428	0.0271	-0.1018	0.00456
$\gamma_4 C_{4dj}$	0.0461	0.0176	0.0171	0.0756
$\gamma_5 C_{5dj}$	0.000025	0.000033	-0.00003	0.000093
$\gamma_6 C_{6dj}$	0.0177	0.0449	-0.0582	0.0795

5 Conclusion

As supply chains move from product-orientation to service-orientation, there will inevitably be a greater need for organizations to understand not only asset utilization, but also the management of knowledge, talent and resources. This requires a network perspective focused on seamless integration amongst various entities and stakeholders. Previous research particularly focuses on the importance of the downstream entities in service-oriented firms (Ahn and Rho 2016). Customers drive service networks to adopt both diverse and flexible process designs additionally internal members are empowered to consistently deal with changing needs in service production and delivery processes. Yet, service-oriented supply chains require seamless integration amongst all members in a network. Big data applications allow for ease of data transfer but the sentiment toward data transfer can vary amongst entities leading to complexities in use from a network perspective.

Very minimal work has been done on stakeholder sentiment of big data use in service-oriented supply chains. Methodologically, this paper employs both textual analysis techniques while also controlling for common problems present in textual analysis by utilizing stakeholders to verify sentiment analyses of textual data. Further, previous studies utilizing social media to examine cumulative sentiment from media typically only utilize keyword searches to assess both news media and social network posts and thus do not control for whether the stakeholders have read the media. This study first conducted a CM document search and then a social network web scrape via Google to collect social network posts that reference the CM document found through the CM document search, thus establishing a causal inference while controlling for time.

Secondly, from a theoretical standpoint, this study contributes by examining what an organization can control to impact stakeholder sentiment, thereby eliciting insight on how corporate communications through media can alter stakeholder sentiment. The findings support the role of the salience of a topic having a more significant impact on stakeholder sentiment than *CM Document Sentiment*. Specifically, this study finds the use of CM Documents discussing *Big Data* in service-oriented

supply chains is more positively associated with stakeholder sentiment than the actual sentiment of the CM Document. This provides support that what is discussed in CM documents is more important in influencing *Stakeholder Sentiment*, than how it is discussed. Further, this finding suggests a positive outlook of big data in the supply chain.

Third, we find the discussion of certain subtopics can elicit stakeholder sentiment through a focus on familiarity bias in technology adoption. Specifically, the subtopic of *Future Applications* is negatively associated with *Stakeholder Sentiment*. This elicits insights regarding stakeholder psychological bias toward proven applications of big data in supply chain management versus potential Future Applications of big data in supply chain management. Possible reasons for this might be derived from a technology adoption theory perspective where immediate reactions to Future Application discussions may trigger stakeholder skepticism thereby reducing sentiment of stakeholders. While the emotional element was not studied in this analysis, providing proof of applications is vital when eliciting enhanced stakeholder sentiment.

The use of big data in supply chain management is progressively impacting how supply chains store, disseminate, and utilize data to enhance knowledge to value creation both internally and to the consumer. In today's environment, organizations are becoming more accommodating in their business operations in order to deliver services according to downstream preferences (Ahn and Rho 2016). Big data applications allow for the quick dissemination of information throughout the network to reduce variability required for greater service performance. However, stakeholder sentiment of big data vary due in part to perceptions relating to its innate connection with privacy risk and vulnerability (Lowry et al. 2017). The use of big data is, ultimately, required for service-oriented supply chains to keep pace with the rapid enhancements caused by digital technologies. The use of corporate communications through media have been utilized to help inform stakeholders of the benefits and to calm the fear imposed by this transition in supply chain management. Nevertheless, understanding how corporate communications can be utilized to reduce the concerns while also enhancing stakeholder sentiment is still in its infancy.

There are, however, several opportunities for expanding on this research that will not only elucidate the various complexities of big data perception from a supply chain perspective, but also enhance organizational knowledge of how corporate communications can enhance the perspectives of stakeholders on big data. For example, this research focuses on one-way communication from the organization to the stakeholder. In the age of digital media, communications are happening rapidly between organizations and stakeholders as well as stakeholder to stakeholder. Previous research has even found a potential link between community concerns about environmental issues and organizational responses to disclosures of environmental information (Brown and Deegan 1998). Future research may entail enhancing insight of stakeholder sentiment on corporate communications through a time-series analysis of social network data. Additionally, technology maturity and regulation can have various effects on manufacturing companies servitization paths (Finne et al. 2013). A time series analysis and larger sample of the various stakeholders of an organization can provide an understanding of how big data analytics can impact

service-oriented procedures and perceptions of those procedures from an employee perspective.

Future research might also address topics discussed on big data in more depth. Specifically, while future applications and current applications were assessed in this study, other topics might also be relevant. Some topics might include feelings expressed, subjectivity of the text, and/or actual applications of big data analyses. Although the latter would require a substantial amount of data, the results would provide further insights into ways in which big data application might yield more positive sentiment.

Aside from social media data, comments from the CM might also be another method of analyzing the impact of CM on reader sentiment. This can also be used to assess the results in this paper with another set of data. However, the user data would need to contain demographics of the individual poster. This is sometimes provided by the website, but often times difficult to find. Future research is encouraged to collect a larger sample and utilize differing forms of user/reader data.

One limitation of this paper was the determination of the cumulative sentiment of response to CM document. An average was utilized in order to determine an average cumulative sentiment of response to CM document. The use of an averaging approach can be problematic when data do not follow a normal distribution. Future research may find other methods to determine the cumulative sentiment of response to CM document including but not limited to a quantile approach.

Another limitation of this research is a focus on stakeholder reaction through social networks. It is problematic to equate online comments to “public opinion” as users are not necessarily demographically representative of local cultures (Neuman et al. 2014). Of course, this is not adjusted for only by the use of in-person, telephone, and online surveys. Collecting a larger, more representative sample of stakeholders through a multi-method analysis may help aid in reducing this bias in future research.

Funding This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

- Ahn J, Rho T (2016) Influence of customer-firm relationships on customer participation in the service industry. *Serv Bus* 10:113–133
- Baesens B, Bapna R, Marsden JR, Vanthienen J, Zhao JL (2016) Transformational issues of big data and analytics in networked business. *Manag Inf Syst Quart* 40(4):807–818
- Breuker D, Matzner M, Delfmann P, Becker J (2016) Comprehensible predictive models for business processes. *Manag Inf Syst Quart* 40(4):1009–1034
- Brown N, Deegan C (1998) The public disclosure of environmental performance information—a dual test of media agenda setting theory and legitimacy theory. *Account Busin Res* 29(1):21–41
- Chae BK (2015) Insights from hashtag# supplychain and Twitter analytics: considering Twitter and Twitter data for supply chain practice and research. *Int J Prod Econ* 165:247–259
- Chang HH, Hung CJ, Wong KH, Lee CH (2013) Using the balanced scorecard on supply chain integration performance—a case study of service businesses. *Serv Bus* 7:539–561

- Chernov G, Valenzuela S, McCombs M (2011) An experimental comparison of two perspectives on the concept of need for orientation in agenda-setting theory. *Journal Mass Commun Quart* 88(1):142–155
- Chew SH, Ebstein RP, Zhong S (2012) Ambiguity aversion and familiarity bias: evidence from behavioral and gene association studies. *J Risk and Uncert* 44(1):1–18
- Chiang AH, Trimi S (2020) Impacts of service robots on service quality. *Serv Bus* 14:439–459
- Cohen MC (2017) Big data and service operations. *Prod Oper Manag* 20(2):112–123
- Davenport TH (2006) Competing on analytics. *Harvard Busin Rev* 84(1):98
- De Oliveira MPV, McCormack K, Trkman P (2012) Business analytics in supply chains—the contingent effect of business process maturity. *Expert Syst Appl* 39(5):5488–5498
- Duan Y, Cao G, Edwards J (2018) Understanding the impact of business analytics on innovation. *Eur J Oper Res* 32:128–140
- Durach CF, Wieland A, Machuca JAD (2015) Antecedents and dimensions of supply chain robustness: a systematic literature review. *Int J Phys Dist Logist Manag* 45(1/2):118–137
- Elias RZ (2004) An examination of business students' perception of corporate social responsibilities before and after bankruptcies. *J Bus Ethics* 52:267–281
- Elving WJ (2013) Scepticism and corporate social responsibility communications: the influence of fit and reputation. *J Market Commun* 19(4):277–292
- Finne M, Brax S, Holmström J (2013) Reversed servitization paths: a case analysis of two manufacturers. *Serv Bus* 7:513–537
- Firth M, Wang KP, Wong SML (2015) Corporate transparency and the impact of investor sentiment on stock prices. *Manag Sci* 61(7):1473–1740
- Freeman RE (1984) *Stakeholder management: framework and philosophy*. Pitman Press, Mansfield, Massachusetts
- Gartner (2011) Gartner's 2011 Hype cycle special report evaluates the maturity of 1900 technologies. <https://www.gartner.com/newsroom/id/1763814>. Accessed Jan 1 2018
- Gawankar SA, Gunasekaran A, Kamble S (2020) A study on investments in the big data-driven supply chain, performance measures and organizational performance in Indian retail 4.0 context. *Int J Prod Res* 58(5):1574–1593
- Goes PB (2014) Editor's comments: big data and IS research. *Manag Inf Syst Quart* 38(3): iii–viii.
- Gupta R, Gupta H, Mohania M (2012) Cloud computing and big data analytics: what is new from data-bases perspective? *International conference on big data analytics*. Springer, Berlin, Heidelberg, pp 42–61
- Gupta S, Modgil S, Gunasekaran A (2019) Big data in lean six sigma: a review and further research directions. *Int J Prod Res* 58(3):947–969
- Hahn PR, Murray JS, Carvalho CM (2020) Bayesian regression tree models for causal inference: regularization, confounding, and heterogeneous effects. *Bayesian Analy* 42:20–34
- Harris I, Wang Y, Wang H (2015) ICT in multimodal transport and technological trends: unleashing potential for the future. *Int J Prod Econ* 159:88–103
- Hazen BT, Boone CA, Ezell JD, Jones-Farmer LA (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. *Int J Prod Econ* 154:72–80
- Hofmann E (2017) Big data and supply chain decisions: the impact of volume, variety and velocity properties on the bullwhip effect. *Int J Prod Res* 55(17):5108–5126
- Hwang T, Kim ST (2019) Balancing in-house and outsourced logistics services: effects on supply chain agility and firm performance. *Serv Busin* 13:531–556
- Kache F, Seuring S (2017) Challenges and opportunities of digital information at the intersection of big data analytics and supply chain management. *Int J Oper Prod Manag* 37(1):10–36
- Kim NR, Hong SG (2020) Text mining for the evaluation of public services: the case of a public bike-sharing system. *Serv Bus* 14:315–331
- Kohli R, Tan SSL (2016) Electronic health records: how can IS researchers contribute to transforming healthcare? *Manag Inf Syst Quart* 40(3):553–573
- Kuo YH, Kusiak A (2019) From data to big data in production research: the past and future trends. *Int J Prod Res* 57(15–16):4828–4853
- Lam HK, Yeung AC, Cheng TE (2016) The impact of firms' social media initiatives on operational efficiency and innovativeness. *J Oper Manag* 47:28–43
- Lau RYK, Zhang W, Xu W (2017) Parallel aspect-oriented sentiment analysis for sales forecasting with big data. *Prod Oper Manag* 20(2):10–20

- Lee SM, Lee DH (2019) “Untact”: a new customer service strategy in the digital age. *Serv Bus* 14:1–22
- Lo SM (2016) The influence of variability and strategy of service supply chains on performance. *Serv Bus* 10:393–421
- Lowry PB, Dinev T, Willison R (2017) Why security and privacy research lies at the centre of the information systems (IS) artefact: proposing a bold research agenda. *Europ J Inf Syst* 26(6):546–563
- Martin KE (2015) Ethical issues in the big data industry. *Manag Inf Syst Quart Exec* 14:2–12
- McAfee A, Brynjolfsson E, Davenport TH, Patil DJ, Barton D (2012) Big data: the management revolution. *Harvard Busin Rev* 90(10):60–68
- McCombs ME, Shaw DL (1972) The agenda-setting function of mass media. *Public Opin Quart* 36(2):176–187
- McCombs ME, Shaw DL, Weaver DH (2014) New directions in agenda-setting theory and research. *Mass Commun Society* 17(6):781–802
- Menon S, Sarkar S (2016) Privacy and big data: scalable approaches to sanitize large transactional databases for sharing. *Manag Inf Syst Quart* 40(4):1–12
- Miles JA (2012) Management and organization theory: a Jossey-Bass reader. John Wiley Sons, New York
- Miller GA (1995) WordNet: a lexical database for English. *Commun ACM* 38(11):39–41
- Motiwalla L, Deokar AV, Sarnikar S, Dimoka A (2019) Leveraging data analytics for behavioral research. *Inf Syst Front* 21:735–742
- Opresnik D, Taisch M (2015) The value of big data in servitization. *Int J Prod Econ* 165:174–184
- Orenstein P (2020) The changing landscape of supply chain networks: an empirical analysis of topological structure. *Inform Syst Oper Res* 10:1–21
- Pereira HG, Salgueiro MDF, Rita P (2017) Online determinants of e-customer satisfaction: application to website purchases in tourism. *Serv Bus* 11:375–403
- Pournarakis DE, Sotiropoulos DN, Giaglis GM (2017) A computational model for mining consumer perceptions in social media. *Decis Support Syst* 93:98–110
- Rogers EM (2003) Elements of diffusion. Diffusion of innovations, 5th edn. Free Press, New York
- Saboo AR, Kumar V, Park I (2016) Using big data to model time-varying effects for marketing resource (Re) allocation. *Manag Inf Syst Quart* 40(4):15–30
- Schymik G, Corral K, Schuff D, Louis RS (2015) The benefits and costs of using metadata to improve enterprise document search. *Decis Sci* 46(6):1049–1075
- Seltzer E, Mahmoudi D (2013) Citizen participation, open innovation, and crowdsourcing: challenges and opportunities for planning. *J Planning Lit* 28(1):3–18
- Setia P, Patel PC (2013) How information systems help create OM capabilities: consequents and antecedents of operational absorptive capacity. *J Oper Manag* 31(6):409–431
- Shah N, Irani Z, Sharif AM (2017) Big data in an HR context: exploring organizational change readiness, employee attitudes and behaviors. *J Busin Res* 70:366–378
- Singer JB (2018) Transmission creep: media effects theories and journalism studies in a digital era. *Journalism Stud* 19(2):209–226
- Smith HJ, Milberg SJ, Burke SJ (1996) Information privacy: measuring individuals’ concerns about organizational practices. *Manag Inf Syst Quart* 20(2):167–196
- Stankovic JA (2014) Research directions for the internet of things. *IEEE Internet of Things J* 1(1):3–9
- Sun S, Cegielski CG, Jia L, Hall DJ (2018) Understanding the factors affecting the organizational adoption of big data. *J Comput Inf Syst* 58(3):193–203
- Swain AK, Cao RQ (2019) Using sentiment analysis to improve supply chain intelligence. *Inf Syst Front* 12(2):12–24
- Tan KH, 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. *Int J Prod Econ* 165:223–233
- Thirumalai S, Sinha KK (2009) Customization strategies in electronic retailing: implications of customer purchase behavior. *Decis Sci* 40(1):5–36
- Trkman P, McCormack K, De Oliveira MPV, Ladeira MB (2010) The impact of business analytics on supply chain performance. *Decis Supp Syst* 49(3):318–327
- Vitari C, Raguseo E (2019) Big data analytics business value and firm performance: linking with environmental context. *Int J Prod Res* 60:395–410
- Waters RD (2013) Tracing the impact of media relations and television coverage on US charitable relief fundraising: an application of agenda-setting theory across three natural disasters. *J Public Relat Res* 25(4):329–346
- Williams CB, Fedorowicz J, Kavanaugh A, Mentzer K, Thatcher JB, Xu J (2018) Leveraging social media to achieve a community policing agenda. *Gov Inf Quart* 35(2):210–222

- Yagade A, Dozier DM (1990) The media agenda-setting effect of concrete versus abstract issues. *Journalism Quart* 67(1):3–10
- Yu Y, Duan W, Cao Q (2013) The impact of social and conventional media on firm equity value: a sentiment analysis approach. *Decis Supp Syst* 55(4):919–926
- Zhong RY, Huang GQ, Lan S, Dai QY, Chen X, Zhang T (2015) A big data approach for logistics trajectory discovery from RFID-enabled production data. *Int J Prod Econ* 165:260–272
- Zhu JH, Blood D (1997) Media agenda-setting theory: telling the public what to think about. *Emerg Theor Human Commun* 1(1):88–114
- Zucker GH (1978) The variable nature of news media influence. *Annal Int Commun Assoc* 2(1):225–240

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.