Unleashing the Power of Social Media Data in Business Decision Making: An exploratory study

Xinwei Li[a], Ying Kei Tse[a*], Fernando Fastoso[b]

[a] *Cardiff Business School, Cardiff University, Cardiff, UK*
[b] *Pforzheim University, Germany*

*Corresponding author: Ying Kei Tse, Email: tsem1@cardiff.ac.uk

Xinwei (Leo) Li (lix154@cardiff.ac.uk) is a PhD candidate at the department of Logistics and Operations Management, Cardiff Business School, Cardiff University, UK. Leo holds an MSc degree in Management with Business Finance from the University of York in the UK. He has published in the International Journal of Production Economics (IJPE), Journal of Retailing and Consumer Services (JRCS), and Sustainability. His main research interests lie in the areas of Operation Management, Supply Chain Management, Corporate Social Irresponsibility, Fake News and Global Relocation Strategies. His work adopts experimental design, eye-tracking techniques and social media big-data analytics.

Prof. Ying Kei (Mike) Tse (tsem1@cardiff.ac.uk; corresponding author) is a Professor in Operations Management at the Cardiff University (Cardiff Business School). His research crosses over different disciplines, including empirical research in risk management and supply chain management, data mining of big social data, decision support in supply chain management, and development of OM educational simulation platform. He has published more than 40 academic articles, including high quality journals such as International Journal of Operations and Production Management (IJOPM), British Journal of Management, IEEE Transactions on Engineering Management, International Journal of Production Economics (IJPE), International Journal of Production Research (IJPR), Supply Chain Management: an International Journal (SCMIJ), Journal of Business Research (JBR), Production Planning and Control (PPC), Annals of Operations Research (AOR), among others.

Prof. Fernando Fastoso (fernando.fastoso@hs-pforzheim.de) is a Professor and Endowed Chair in Brand Management for Luxury and High-Value Brands at Pforzheim University, Pforzheim, Germany. His research interests lie in the areas of global and luxury branding and consumer psychology.

Unleashing the power of Social Media Data in Business Decision Making: An exploratory study

This study systematically reviews the research on applying social media data (SMD) in business decision making. We applied bibliometric mapping and a Latent Dirichlet Allocation topic modelling approach to conduct a systematic literature review. Our analysis shows that research to date has uncovered that sentiment analysis and opinion mining supported businesses in observing, analysing, and predicting customer behaviour in various sectors, such as tourism and online shopping decision making. However, by applying big data analytical methods on paper classification, we found that descriptive and predictive analyses are prevalent while prescriptive analyses on SMD are rare. For example, whereas most research focuses on examining the application of SMD on communication and marketing, predicting customer behaviour and decision making, few studies focus on guiding companies on what to do next. Our analysis highlights the need for future research to shed light onto newly discovered forms of SMD such as bullet-commenting and audio-chat data, increasing the accuracy of sentiment detection with concept-level analysis, and in leveraging SMD on electronic word of mouth (E-WOM) development.

Keywords: social media data; business decision making; text mining; topic modelling, systematic literature review

1. Introduction

With the upsurge in social media use over the past decade, enterprises have become more active in using social media. As far back as 2011, Burberry spent more than 60% of its marketing budget on digital media (Barrett and Bradshaw 2011). Today, Facebook ads are used by 75% of marketers (Stelzner 2021) and there were around 10 million active advertisers on Facebook in the third quarter of 2020 (Deborah 2020). The prevalence of social media in consumers' lives allows companies to generate enormous social media data (SMD) that can be transformed into business advantages, such as enhancing marketing performance, developing customer engagement, motivating user participation and enabling social commerce activities (Wamba et al. 2017; Bu, Huang, and Zhao 2021). According to McAfee and colleagues (McAfee et al. 2012), data-driven companies achieve 5% more productivity and 6% more profits than their competitors.

SMD belongs to the category of "big data" (Davenport, Barth, and Bean 2012). SMD refers to different forms of data generated by social media users and is distinct from other big data like sensor and biometric data (Verma et al. 2016). Research on SMD in the context of business decision making has investigated a wide range of issues such as customer communication, marketing, and organisational management (Nayak, Nayak, and Jena, 2020; Palalic et al., 2021). In light of the plethora of work on SMD in business decision making, some existing works have reviewed the application of SMD amongst different topics, such as supply chain resilience (Li et al., 2023) and operations management (Choi, Wallace, and Wang 2018; Zhang et al., 2020), however, it is surprising that a review of the research body on the SMD and business decision making is still lacking, this study delivers such a review. Specifically, we apply bibliometric mapping and the Latent Dirichlet Allocation topic modelling approach (Blei and Lafferty 2009; Van Eck and Waltman 2010) to conduct a systematic literature review (Tranfield, Denyer, and Smart 2003) of the SMD field and business decision making. This review is timely and important on several ground. First, the plethora of work in the field makes it challenging for businesses to understand how to deal with the overwhelming amount of social media research to inform decision making (Stieglitz et al. 2018). This challenge is exacerbated by the nature of the data itself (e.g., volume, variety, velocity, and veracity), the processes needed to analyse it (e.g., techniques on analysing unstructured data), and the management of such processes (e.g., privacy and security) (Sivarajah et al. 2017). Second, the lack of a review of the field makes it hard for researchers to identify important research gaps to advance this research field, and a critical benefit of using analytical techniques like text mining is to explore the large scale of articles (Li et al., 2023; Zhang et al., 2020).

This study has the following aims: (1) to provide an in-depth understanding of the SMD research contributions in supporting business decision making; (2) to explore the current trends in research on social media aided decision making, and (3) to identify future research directions in the field of SMD and business decision making. To achieve these, we combine bibliometric mapping and topic modelling approach (Blei and Lafferty 2009; Van Eck and Waltman 2010), which allows us to systematically review hidden relationships and the co-occurrence of crucial terms and concepts from prior publications. Moreover, these tools help us reveal constructs that are difficult to identify from unstructured data, such as connecting future research opportunities from disparate research topics.

The remaining sections of this paper are structured as follows: First, we introduce the approached used in our systematic literature review, that is, bibliometric mapping and the Latent Dirichlet Allocation (LDA) topic modelling approach. Next, we present a bibliometric analysis together with the result of descriptive and thematic analysis from a systematic literature review. Lastly, we discuss exploratory insights with topic modelling and propose future directions from the perspective of research methods and practical research topics.

2. Methodology

2.1 The systematic literature review

The systematic literature review approach is used to evaluate the state of knowledge on a particular research topic. It was first introduced to the management discipline by Tranfield, Denyer, and Smart (2003). This method can be adapted to create research agendas or identify research gaps. We follow a similar systematic literature review approach introduced by prior researchers (Tranfield, Denyer, and Smart 2003), wherein there are three stages: planning (identify the need, make a proposal, develop a review protocol); conducting (select studies, study quality assessment, data extraction and monitoring progress, data synthesis), and reporting (report and recommendations, getting evidence into practice).

The research protocol begins with the article searching process (Figure 1). Since the focus of our research is SMD in business decision making, we employed the following search string to ensure data consistency and relevance: [“social media” AND “business” AND “decision making”]. This string was used within the title, abstract and keyword search fields on Scopus, the largest abstract and citation database of peer-reviewed literature, which in 2020 had over 24,000 titles from more than 7,000 international publishers worldwide. First, the keywords of “social media” allow researchers to collect social media research papers, the keyword of “business” ensure the sample articles are business context related, and “decision making” ensures the captured data are related to different decision making. Second, we limited the search string to appear in the search fields of the title, abstract and keywords, this helps researchers to effectively narrow down the sample articles to the target scope and avoid losing focus.

An initial search returned 529 results. Next, several inclusion criteria were implemented to filter the results. We selected papers published between 2013 and 2021, given that the number of articles published before 2013 on this specific topic was very low. By 2013, there had been remarkable growth and a threefold increase in the number of previous publications (9 papers) which has continually grown since then. Further, only journal articles in their final stage and in English were considered, as these journal articles are likely to have the highest impact on the field (Keupp, Palmié, and Gassmann 2012). Performing this filtering step resulted in a set of 210 articles. We then used this dataset for text mining; the detailed mining process is explained in section 2.2.

Next, to improve the quality of the paper being reviewed, we excluded journals that possessed fewer than 2 stars in the ABS Journal Guide and that had a 5-year impact factor of less than 1.0 (Gong et al. 2017). Additionally, we browsed the abstracts of all remaining papers and removed articles that did not particularly focus on social media and business topics. Finally, 89 unique articles met the pre-design filtering criteria and were then used for descriptive and thematic analysis. This was followed by annotation of the articles with their research scope and key themes. Finally, four key themes were identified from the articles and we then integrated the big data analysis method to determine the final theme classification.

[Insert Figure 1 here]

2.2 Systematic literature review with text-mining approach

The current systematic literature review approach is widely used (Tranfield, Denyer, and Smart 2003). However, the explosive growth of publications makes identifying relevant studies in an unbiased way increasingly complex (Bastian, Glasziou, and Chalmers 2010), which has prompted an increased interest in applying text mining in systematic literature reviews to address information overload. Text mining is a data-mining approach used to discover unknown information and structure from textual data, and uncover hidden relationships and the co-occurrence of terms and concepts which reviewers may overlook in traditional content analysis (Feldman and Sanger 2007). Therefore, text mining can help researchers reveal some interesting “patterns” in the literature, and different text-mining techniques have been applied to facilitate systematic literature reviews, such as topic modelling, clustering and bibliometric analysis (Moro, Cortez, and Rita 2015; Demeter, Szász, and Kő 2019; Kumar, Sharma, and Salo 2019).

This research applies bibliometric mapping and topic modelling, each for a different purpose. The former focuses on a graphical representation of bibliometric maps, which is especially useful for displaying bibliometric maps in an easy-to-interpret way (Van Eck and Waltman 2010). Topic modelling is used to attain a deeper understanding by creating term-document matrices that can be used for more fine-grained analysis. Despite our pre-defined criteria resulting in a selection of 89 articles, this was too few for a meaningful text-mining analysis. Therefore, we decided to use the 210 articles previously selected. First, this dataset went through a pre-designed filter process (Figure 1; and see 2.1), which covers the major journal articles on our research topic. Second, text mining is a Natural Language Processing tool to discover linguistic similarities in text texts, although the textual data that captured from the 89 articles well representing the existing works in our research topic, there is insufficient topics and information that can be extracted from a small dataset; therefore, we applied text mining to the 210 articles dataset to obtain sufficient and valuable textual data, and used the 89-article set for a fine-grained descriptive analysis and thematic analysis.

Bibliometric mapping is a statistical and quantitative approach that provides network analysis of literature, using a built in clustering and mapping algorithm to automatically construct maps of authors, journals, and institutions based on co-citation data or co-occurrence of terms (Van Eck and Waltman 2010). Hence, we followed a similar approach to construct a journal co-citation network using keyword co-occurrence mapping and term co-occurrence mapping, based on co-citation information, author keywords and abstracts harvested from the 210 articles.

Topic modelling is an effective technique for the identification of co-occurring words in qualitative data. An important aspect is that the model assumes that documents hold a hidden distribution of latent topics (Griffiths and Steyvers 2004). We adopted the LDA algorithm to detect underlying thematic information, which is a flexible and efficient unsupervised machine learning approach to extract meaningful topics from textual documents (Blei and Lafferty 2009; DiMaggio, Nag, and Blei 2013). There are several reasons that this approach is suitable here. First, discussion of future research directions is typically located in a specific section or in the last couple of sentences of an article, yet they are often relatively unstructured in content (i.e., not in table format). However, LDA topic modelling is efficient in extracting meaningful latent topics from unstructured data, and therefore, can help reveal valuable information that is difficult to uncover from unstructured data. Moreover, conventional qualitative analysis of text requires scholars to identify topics manually, and the topic generating process is heavily reliant on the interpretation of individual researchers. Conversely, topic modelling does not require prior annotations of the corpus, and it can help researchers reduce bias in the topic interpretation process by detecting topics based on the co-occurrence of words and their probability distributions. Additionally, this technique has been demonstrated effectively in a variety of short-text settings, such as online chatter and Twitter data (Ibrahim and Wang 2019; Tirunillai and Tellis 2014).

First, we started data collection by capturing content that mentioned the direction of future research using the 210 articles as the corpus for the LDA model. The detailed steps used to generate this corpus is presented in Appendix A. This was followed by data pre-processing to remove noise and transform unstructured data into an analytics-ready structure. This process includes tokenisation (i.e., breaking down text into separate words), removing stop words (e.g. “a” or “and”), transforming text to lower case, and reducing words to their stems (e.g. competition to compet). Determining an optimal number of topics (K value) is important as it can significantly impact the model result. Therefore, we tested model stability by experimenting with a variety of K settings to graph the average coherence score (see in Appendix B) to narrow in on stable sets of topics. Average coherence score is a quantitative metric for topic quality which reveals a plateau value (Fligstein, Stuart Brundage, and Schultz 2017). The peak of the curve represents the optimal coherence value, K, for the model. We then assigned the maximum value in our model as the number of topics. This process yielded nine topics extracted from the corpus, and each topic had 10 keywords. Next, we manually annotated the extracted topics into meaningful themes with the combined judgement (Guo et al. 2016). Two researchers labelled the topic independently with a pre-defined rule (examine the keywords with a descending word weight order), which can best explain the combination of keywords in each topic. Then, we compared the results and determined the final topic name by consensus. Finally, seven highly related topics among nine topics were selected as a result of topic modelling.

3. Findings

3.1 Bibliometric mapping

This section presents the results of the bibliometric analysis of the 210 articles. In Figure 2, clusters of keywords with high co-occurrence are depicted in the same colour, important keywords are presented in larger fonts, and links indicate co-citation of keywords. “Big data” and “sentiment analysis” were the most frequently cited keywords and both are positioned in the centre, suggesting that SMD research is strongly associated with the big data concept as well as with the popularity of sentiment analysis in SMD research. As can be seen from Figure 2, four key clusters were identified in the tag cloud. The green cluster belongs to the organisation management field. The red cluster includes important keywords like “big data”, “decision making” and “data mining”, which is predominantly focused on big data analytics. The green cluster predominantly consists of social media and marketing keywords. The remaining clusters are in blue (consumer behaviour) and purple (social media analytics). It is worth mentioning that “text mining”, “data mining” and “natural language processing” are positioned around the centre of different clusters and next to “social media” and “decision making”, representing that these methods play essential roles.

[Insert Figure 2 here]

Next, we created a co-citation network using the cited sources (Figure 3). The relatedness of journals was determined by the frequency with which they are cited together. The size of the dot represents the number of citations of different journals, and the lines between dots represent the co-citation relationship among journals. Journal sources in the same colour indicate a higher co-citation count. Figure 3 shows that MIS Quarterly was the most influential journal due to the number of times it was cited. Additionally, there are five clusters in the co-citation network with different colours: sector domain (purple), information system and decision science (red & yellow), organisational management (blue), and operations management (green). This result indicates that the information system and decision science journals (e.g., MIS quarterly and Decision Support System) showed the most interest in using SMD in business decision making. The sector journals paid less attention to SMD, whereas the tourism-related journals (e.g., Tourism Management) were the only journal type in the sector interested in SMD decision-making. Journals in the field of organisational management (e.g., Management Science) and operations management (e.g., International Journal of Operation Management) had a similar interest in SMD decision making.

[Insert Figure 3 here]

3.2 Descriptive analysis

This section provides a descriptive analysis of our set of 89 articles, including journal distribution, year distribution, classification of publications, and research method type. Overall, SMD research in business decision making was rather spread out in different fields and in relatively new areas as there were only 1–2 papers published in some journals, suggesting that this topic was rather interdisciplinary. Specifically, the International Journal of Information Management had the most papers (C=5) with this content, and there was a high percentage of publications in information-system-related journals, suggesting that data-driven decision making in the information system field was increasingly important. Moreover, there were a few publications from the field of operation management, showing the potential of incorporating SMD into business operations. Figure 4 shows that the largest number of publications were recorded for the year 2020 (C=20). The rapid increase of publications since 2013 highlights the awareness and importance of SMD in the business decision-making process. Although the number of publications dropped back to only three publications in 2016, the following years represent a relatively high output in this field. Overall, the distribution of publications in this domain represents real growth from 2013 to 2021, which indicates that there will be more publications in this research field in the future.

[Insert Figure 4 here]

Next, we employed an analogous list of publishers and previous scholars (Dwivedi and Mustafee 2010), then categorised the 89 papers into six types of publication. Table 1 shows that most publications were research papers (C=42), which accounted for 47.19% of publications overall, followed by conceptual papers (C=17, 19.10%).

[Table 1 near here]

Furthermore, following the methodological classification suggested in previous work (Sivarajah et al. 2017; Dwivedi and Mustafee 2010), we coded the 89 articles into nine categories: (1) conceptual/ descriptive/ theoretical; (2) survey; (3) case study; (4) experiment; (5) interview; (6) design research; (7) literature analysis; (8) viewpoint/ commentary, and (9) mixed. Overall, design research (C=26, 29.21%) and conceptual/ descriptive/ theoretical (C=22, 24.72%) were the two most popular research methods. Other methods with their associated counts and percentage are presented in Table 2. Moreover, there were few papers that used interviews, which is consistent with the findings of Sivarajah et al. (2017) on big data research, due to the discipline generally requiring technical and methodical analysis of the big data involved. As a subtype of big data, SMD has characteristics such as volume, velocity, variety, and veracity (McAfee et al. 2012). Hence, interviews are little used in SMD research.

[Table 2 near here]

Next, we present a summary of selected articles, which include parameters such as research focus, methods, data type and sampling, theories/ models, and findings. As can be seen from the selected paper in Table 3, research topics developed from SMD are diverse, ranging from customer issues, marketing strategies innovation and supply chain purchasing problems. Moreover, many studies apply the text-mining approach, or combine it with other approaches (e.g., surveys, case studies, and interviews) in SMD research. Regarding the data and sampling, social media comments, reviews and ratings from customers, social media platforms were frequently used, and some research collected data from industry practitioners.

Furthermore, different theories were applied in the 89 SMD articles. For example, in the Uses and Gratifications Theory (West, Turner, and Zhao 2010), people actively seek out specific media to satisfy their specific demands. This theory was adopted to support the investigation of marketing aggressiveness level and how to obtain competitive advantages from SMD (He, Zha, and Li 2013; Wang, Baesens, and Zhu 2019). Building upon the Theory of Reasoned Action, Itani, Agnihotri, and Dingus (2017) explain the antecedents and consequences of a B2B salesperson's social media use; the Theory of Reasoned Action suggests that a person's behaviour is determined by their intention to perform a specific behaviour (Fishbein, Ajzen, and Belief 1975). Moreover, Dual Process Theory suggests the notion that individuals have two modes of cognition when processing information; a system 1 that refers to a fast, impulsive mind which is employed initially, and a system 2 that refers to a slow, deliberate mode which may be additionally invoked later (Evans 2003; Kahneman 2011). This theory was incorporated with Processing Theory to explain B2B purchasing and the attitudes of B2B salespeople towards social media use (Krijestorac, Garg, and Konana 2021). Notwithstanding the different theories that have been used in SMD research, we also found that it may be difficult for researchers to make theoretical contributions when the research is simply applying the text-mining techniques to textual data without engaging other constructs or strategies, which can only be useful for exploratory purposes on a certain topic. For example, by only mining customers' opinions or sentiments from SMD.

[Table 3 near here]

3.3 Thematic analysis

This section first introduces the classification of big data analytical methods and then discusses key themes from the perspective of descriptive, predictive, and prescriptive methods.

3.3.1 Big data analytical methods

Descriptive analysis answers the question of what happened, presenting a summary of the past (Lustig et al. 2010). It adopts basic statistical operations like reporting, visualisation, creating reports, and presenting information (Rehman et al. 2016). A predictive analysis informs what will happen next, carrying on from more complex methods like probabilistic models, machine learning, and statistical analysis (Lepenioti et al. 2020). It determines future possibilities by forecasting and statistical modelling (Waller and Fawcett 2013). Then there is prescriptive analysis, which recommends what the business should do next, covering methods like mathematical programming, evolutionary computation, simulation, and logic-based models (Lepenioti et al. 2020; LaValle et al. 2011). Prescriptive analytics help businesses manage shifting information and the evolution of business models (Rehman et al. 2016), providing the most valuable support compared to predictive and descriptive analysis in big data analytics.

[Insert Figure 5 here]

The term “big data” covers a great range of digitized information, including SMD. This research follows the definition generated by Lustig et al. (2010) to classify social media research. We conducted thematic coding to identify four key themes, i.e., communication and marketing, organisational management and crisis management, customer behaviour and decision making, and big data and business intelligence. Then, we classified them into either descriptive, prescriptive, or predictive, to indicate their value and complexity (Figure 5).

3.3.2 Descriptive analytics perspective

Communication and marketing

Online interactions have become a dominant vector of communication between companies and customers. Applying sentiment analysis to SMD can help companies to assimilate and understand the torrent of communications about their products and services, such as managing customer relationships (Gu and Ye 2014), enhancing reputation and brand image (Dijkmans, Kerkhof, and Beukeboom 2015), and reaching potential customers (Wang, Baesens, and Zhu 2019).

The continuous stream of user-generated content contributes to a huge amount of SMD, such as messages, comments, replies and social tagging networks. Many attempts have been made to use SMD to support businesses. For example, analysing SMD about company responses to consumer criticism can provide a guide for companies to better interact with customers (Xia 2013), collecting social tagging from

SMD as a proxy measure for brand performance (Nam and Kannan 2014), and managing marketing decision making with online reviews (Kauffmann et al. 2019). Moreover, companies can gather market insights from SMD to assist with production, such as acquiring knowledge from SMD to drive brand innovations (Nguyen et al. 2015) or support new product development (Du, Yalcinkaya, and Bstieler 2016). However, using SMD to support marketing is not always beneficial. Hardey (2011) argues that with the overwhelming amount of user-generated content being collected, the information can be very noisy, contributing to infoglut. There is also research which suggests that the relationship between seller marketing aggressiveness and marketing popularity follows an inverted U-shaped curve, as businesses may be penalised in terms of popularity by pursuing too aggressive a marketing strategy on social media (Wang, Baesens, and Zhu 2019).

Organisational management and crisis management

With the increased application of social media in business, SMD can be used to benefit organisation management. For example, SMD can be used to manage the communication of co-workers and address inappropriate social media contacts, or manage supplier attractiveness in business-to-business markets (Tóth et al. 2019). Furthermore, social networking applications can be used to facilitate change by addressing employee concerns and managing knowledge sharing (Naeem 2020). Moreover, social media allows employees to voice their thoughts to improve communication and inform corporate decision making. Miles and Mangold (2014) suggest that the SMD contributed by employees can be an untapped resource for enhancing an organisation's public image when delivering positive information. However, when the employee voice is negative, there can be a devastating impact on the firm's reputation.

Technology allows social media to provide instant delivery of a message during a crisis. Meanwhile, SMD generated by users is also valuable for guiding future crisis management (Roshan, Warren, and Carr 2016). For example, SMD can be used to support decision making and influence after-emergency training (Pohl, Bouchachia, and Hellwagner 2015), prevent or mitigate the impact of future product recalls (Tse et al. 2018), or estimate the post-emergency recovery status of small businesses in urban areas (Eyre, De Luca, and Simini 2020).

3.3.3 Predictive analytics perspective

Customer behaviour and decision making

Predictive analytics provides forecasting insights from a massive influx of data, supporting businesses to predict market development. Prior studies use different SMD to unveil trends and customer behaviour patterns and decision making, for example, developing social computing algorithms based on social network analysis for anticipatory computing (Liao and Chen 2019), using topic sentiment mining to predict e-commerce sales performance (Yuan et al. 2018), applying deep learning models on tweet data to predict interest rates (Yasir et al. 2020), and identifying online sellers by classifying user connections through social media avatars and users' social activities

(Mao et al. 2020).

Moreover, the explosion of social media usage also facilitates the self-reporting of products and services, allowing customers to gather relevant information from SMD before making a purchase. Particularly, SMD has significant implications for the tourism industry, such as affecting tourism destination choices (Munar and Jacobsen 2014). Therefore, SMD also offers great value to predictive analysis for businesses, for example, by detecting influential online reviews of tourist attractions to help attraction management (Fang et al. 2016), analysing tourist dining preferences (Vu et al. 2019), and using online reviews with historical sales data to measure restaurant performance (Fernandes et al. 2021). Overall, SMD has emerged as an important component of the tourism domain, which can help marketers understand current tourists better and better predict how to reach potential customers.

3.3.4 Prescriptive analytics perspective

Big data and business intelligence

Prescriptive analytics can help businesses find the best course of action for the future, including proactive decision making and action implementation. This type of analytics can optimise business processes and allow the pursuit of the best outcome, such as arranging transportation capacities and varying ticket prices to achieve maximum profitability (Soltanpoor and Sellis 2016). With the emergence of the big data concept, it is trendy to extract business intelligence from big data to support decision making. Business intelligence is defined as a decision support system that combines data gathering, data storage and knowledge management (Negash and Gray 2008), aiming to assess risks and explore opportunities in a competitive environment. As one type of big data, SMD can support business competition and intelligence gathering from different perspectives (Yang et al., 2022), for example, gaining competitive intelligence from tweets (He et al. 2015) or detecting anomalous user emotions (Sun et al. 2018), and investigating how participating in online review sites influences collective decisions (Cho and Chan 2021). Prescriptive analytics is the most complex big data analysis method, yet extracting business intelligence from SMD and effectively engaging it in prescriptive analytics remains a field full of potential.

3.4 Topic modelling

This section presents the results of future direction LDA topic model analysis of the corpus of 210 articles. We identified seven topics, have listed associated keywords and ranked them by topic weight in Table 4.

[Table 4 near here]

(1) Marketing Group: The development of social media has revolutionised communications and, consequently, marketing strategies. Social media has a strong impact on customer purchase decision making, such as online shopping, hostility consumption and healthcare decision making. Companies are making great efforts on social media marketing to boost their business and build brand awareness.

(2) Sample & Data Selection Group: A limitation of SMD research is the sampling biases caused by using a single SMD source. SMD involves information from different cultural backgrounds, user groups, generations and gender differences. Moreover, the selection of SMD had various sources; Twitter and Facebook are well-known social media platforms but are also joined by video sharing platforms such as Tiktok, YouTube and Bilibili. Additionally, when researchers focus on particular industries or areas, specialised data channels like user forums (e.g., Reddit and GitHub) and expert blogs provide precious information. Therefore, narrowing down the data source from a general platform to a smaller social media platform may be a wise option.

(3) Sentiment Group: Sentiment analysis is a subfield of text mining, which extracts user sentiment from SMD and distinguishes their polarity, the progressive increase of social media usage, and the popularity of exchanging thoughts online. There are four sentiment analysis levels: document-level, sentence-level, aspect-level, and concept-level (Hemmatian and Sohrabi 2019). The four levels represent the lowest accuracy to the highest accuracy of sentiment evaluation. The concept-level delivers the most precise detection, and it also requires the most demanding machine learning techniques.

(4) Method Group: The text-mining method combines statistics, linguistics, and machine learning, which covers extracting features from documents, detecting sentiment from sentences and training data by machine to obtain deep information. Text-mining methods range from basic methods (e.g., word frequency, collocation

and concordance) to advanced methods (e.g., text classification and extraction). Different techniques have different levels of complexity and accuracy for mining opinions.

(5) Practical Group: SMD has provided tremendous social function and commercial values, and many studies have demonstrated how to mine SMD and perform sentiment analysis. Furthermore, extracting business intelligence from SMD and transforming the data into practical applications provides more value. For example, SMD has a significant impact on health surveillance (e.g., tracking and tracing the spread of coronavirus through Twitter), disaster management (e.g., helping to issue flash flood watches or warnings) and practical business applications (e.g., product development based on empirical evidence on social media).

(6) E-WOM Group: E-WOM has taken on a significant role in promoting online shopping. Individual purchasing decisions tend to be impacted by the attitude of others and social media reputations are becoming increasingly important. When customers share their thoughts about service and product quality online, these SMD are beneficial for other individual customers and transform how corporations perform digital marketing. Furthermore, as SMD involves a vast proportion of the younger generation, it is not surprising that some new enterprises perform only digital marketing to improve marketing efficiency.

(7) Management Group: The adoption of SMD related to management issues covers a wide range, from fostering and managing relationships with organisational stakeholders (e.g., employee management and customer management) to crisis

management (e.g., crisis communication and management during natural crisis events, and corporate reputation management during public crisis events).

4. Future research directions

This section discusses two streams of future research opportunities: (i) research methods and (ii) research topics from topic modelling.

4.1 Research methods for future research

There are numerous qualitative SMD pouring through the internet, but few qualitative studies use SMD in the information system field (Müller et al. 2016). Social media research is a relatively new field, and the social media environment continues change with technological development. Specific research topics such as studying online communities or social media experiences can be better addressed with qualitative research methods. Examples include investigating the role of Facebook in romantic relationship development using a focus group methodology (Fox, Warber, and Makstaller 2013) or studying the LGBT movement through SMD in the game of World of Warcraft (McKenna, Myers, and Newman 2017). Qualitative research methods in the information systems field have not fully embraced the opportunities that SMD offers, and specific research questions such as community studies can be addressed with qualitative research methods in the future.

Moreover, using experimental methods in business research is trendy, for example, adopting experimental methods on word-of-mouth investigation (Wang et al. 2021). According to Tim (2014), experimental methods have great potential to generate impact.

Gong et al. (2017) propose recommendations for businesses on managing Twitter use as a marketing tool through experiments. Bapna et al. (2017) applied experimental methods to explore how online social networks are linked to the economic measure of trust. Experimental methods enable researchers to manipulate different variables and establish causality rather than mere correlation (Harrison and List 2004). For example, an examination of social media users and their information-sharing behaviour (Shi, Rui, and Whinston 2014) heavily relied on cross-sectional designs to study this relationship, but was inconclusive with regard to how message sender attributes impact other people's sharing behaviour. Ert, Fleischer, and Magen (2016) then adopted a vignette-based experimental method to isolate the impact of message sender attributes, such as avatars. Therefore, we encourage future research to consider applying experimental research methods to overcome the limits of more traditional methods, for example, study questions that better suit establishing causality or studies which require precise control and manipulation of variables in the research design.

4.2 Future directions using topic modelling

(1) Expanding data selection

SMD is massive, noisy, incomplete, and unstructured. There is population bias from a single SMD. For example, Instagram is especially appealing to a specific group of people and is dominated by females (Ruths and Pfeffer 2014). SMD research using opinion mining and sentiment analysis can select different data sources (e.g., Reddit, TikTok, YouTube) to achieve a more accurate representation of a population. Moreover, there are platforms (e.g., user forums and expert blogs) that serve a specific group of users; for example, LinkedIn is an employment-oriented social networking platform with rich information about professional networking. Hence, when researchers study some niche topics that focus on certain areas and industries, collecting data by tapping into their specialised channels is fruitful. However, some platforms view privacy issues as a priority (e.g., MeWe and Telegram), so it is important to be cautious due to ethical concerns when engaging with SMD from those platforms.

Moreover, different types of SMD can be used together. SMD composes multiple formats, including text, image, video, and networking information. Moreover, new SMD such as bullet-commenting technology on Bilibili fosters an alternative form of online commenting, which is another excellent source of SMD (Zhang, Wang, and Chen 2020). It is worthwhile combining opinions from different platforms to provide a comprehensive view and understanding of different user groups. Besides, images and videos are more informative and sometimes “speak” louder than words. Other data can be used as a complement to textual data. Therefore, future research can take advantage of SMD’s format diversity and richness rather than only focusing on textual data.

(2) Concept-level approach to opinion mining and sentiment analysis

The prominent appearance of text mining in social media research draws another future direction. Opinion mining and sentiment analysis are two emerging fields that aim to help users find opinionated information and detect sentiment polarity, which is considered as a subfield of text mining. Among the four levels of text mining, the document-level approach (i.e., lexicon-based), is relatively naive compared to other advanced methods like sentence-aspect, aspect-level or concept-level (Hemmatian and Sohrabi 2019). Simultaneously, the machine learning method provides a concept level of opinion mining and sentiment analysis, which has higher accuracy in sentiment detection and text classification. Therefore, researchers can consider developing machine learning in concept-level text-mining research to obtain a higher accuracy of public sentiment, facilitating business decision-making in a smarter, more precise way.

(3) E-WOM in SMD

From a resource-based view to a knowledge-based view, the way for firms to gain competitive advantages has shifted from tangible to more intangible resources. For example, developing good E-WOM is crucial for business development. Social media creates a digital environment for communication and marketing, which also harbours a wealth of user-generated content to inform purchase decisions, especially in the hospitality industry. E-WOM is becoming a powerful force in the marketplace, affecting brand image and customer purchase intention (Jalilvand and Samiei 2012), influencing travel destination choices and average restaurant meal prices (Yim, Lee, and Kim 2014). With the booming development of social media, there are more opportunities for both academics and practitioners to capture valuable information from SMD, helping businesses with data-driven marketing decision making.

(4) SMD generated by the remote working environment

SMD was previously used widely for corporate management. Working from home has become a new trend due to the government officials' response to the pandemic, and it is likely to remain common post-COVID. Moreover, various social media tools such as online meetings and file-sharing software play remarkable roles in supporting remote working and team cooperation (e.g., Zoom and Dropbox). As most work and communication rely on social media tools, there is a higher density of online interaction among staff, which provides a high volume of SMD. Therefore, future researchers should consider the potential use of these new SMD to improve corporate management and organisational decision-making.

(5) SMD in crisis management and communication

Social media is an efficient form of public participation and backchannel communication during crisis events. SMD in crisis management refers to using information and communication technology to respond to and cope with disaster and uncertainty (Palen and Anderson 2016). For example, exploring consumer opinions on Twitter in the food fraud incident of the horsemeat scandal (Tse et al. 2016; Tse et al. 2018), and examining the impact of social media opinions about COVID-19 (Eachempati, Srivastava, and Zhang 2021). By scrutinising customer responses on Twitter, they provide guidance for businesses to either prevent or mitigate the impact of future crises. Increasing engagement of social media in business operations leads to a greater role of SMD in crisis management. It is important for companies to adjust their social media marketing and communication strategies in response to unexpected events such as product-harm crises. However, interactions between customers and firms, corporate response mechanisms, customers and public perception, are still not well investigated. Therefore, we suggest that future research further explores how businesses should establish different social media management and communication strategies before, during and after a business crisis occurs, therefore scrutinising customer engagement strategy best practices and their impact on firm performance.

(6) Misinformation in SMD

There are many warnings about SMD's reliability (Oh, Agrawal, and Rao 2013). Social media allows the easy exchange of information, and this also facilitates the spread of online misinformation, which may bring harmful consequences to individuals and organisations, such as loss of reputation or reduced trust (Bordia et al. 2005) and affect customer knowledge and attitudes towards brands (Rapp and Salovich 2018; Visentin, Pizzi, and Pichierri 2019). Vosoughi, Roy, and Aral (2018) investigated the diffusion of 126,000 news stories distributed on Twitter between 2006 and 2017 and suggest that falsehood diffused significantly farther, faster, deeper, and more broadly than the truth. Moreover, health-related rumours generally have a wide appeal and can impact healthcare decision making. For example, misguided information about the Zika virus was far more popular than accurate public health information on Facebook, suggesting an urgent need to better curate health-related posts on social media during health crises/ pandemics (Sharma et al. 2017). Hence, several opportunities can be found, such as studying the characteristics of business-related rumours, exploring the drivers and patterns of online rumours, and examining health-related SMD about COVID-19 during different periods (beginning, mid-term, post-COVID). Moreover, an important field is developing techniques for detecting online misinformation from SMD to reduce its impact early.

(7) Real-time SMD

Real-Time SMD holds significant business value in supporting decision making during developing events. Owing to the advantage of information delivery speed, social media provides a live update to the public about everything that concerns people. Li et al. (2018) suggest that accurate real-time observations during incidents are essential for crisis forecasting and nowcasting. Real-time SMD is a great representation of what is happening, which serves as a great reference for individuals and businesses to make predictions and assumptions when emergencies occur. Disaster response is a great example of real-time SMD. For example, SMD can be used as an early-warning system for detecting earthquakes (Li et al. 2018). Another prominent case is the outbreak of Covid-19; the trending topic on social media platforms contributes to a massive amount of SMD with a real-time observation of the effects of coronavirus (e.g., NHS tweets).

5. Conclusion

Overall, this study has explored how organisations and enterprises in different sectors benefit from SMD in practice and how they mine valuable information from the massive pool of SMD. We identified and delineated the main research topics in the literature and classified them using different big data analytic method categorisation (i.e., descriptive, predictive and prescriptive). Furthermore, we found that SMD has been often used for performing descriptive and predictive analysis, answering the question of what happened (e.g., opinion mining from reviews) and what will happen (e.g., using SMD to predict the market). However, there has been less prescriptive analysis of SMD to suggest what businesses should do next. Meanwhile, we applied text-mining techniques to examine hidden information behind the clustered groups. First, the bibliometric analysis provides a pre-visual mapping to identify current research topics by generating keyword co-occurrence mapping and a journal co-citation network. Next, LDA topic modelling revealed the constructs of future directions that are often difficult to identify from unstructured data by extracting meaningful topics and keywords from the articles' arguments about future opportunities and limitations.

This study makes important practical and theoretical implications. First, we present a novel approach by combining LDA topic modelling with bibliometric analysis in an SLR research. A similar approach can be applied to effectively explore other research topics. Moreover, drawing on big data analytic methods and thematic analysis, we propose the future direction of methodology and suggest several practical future research topics. Furthermore, by summarising the theoretical applications of SMD research, we found that simply applying text-mining tools to SMD may make an insufficient theoretical contribution. Therefore, future SMD research should consider engaging more constructs, conceptual measurements or strategies when conducting social media research. In addition, we present a summary and comparison of selected studies related to using SMD in supporting business decision making. We hope that this can help future researchers in terms of identifying theories and methodologies to answer their research questions more effectively.

These findings should be viewed in light of some limitations. Notwithstanding the fact that we apply text mining to minimise subjective bias in the thematic analysis, the process of capturing arguments from articles about future direction awaits to be improved; this process should not just rely on authors' judgements, but a stricter and unbiased approach could be developed to determine the choice of mining objective to improve the capture process.

Acknowledgements

N/A.

Declaration of interest statement

To the best of our knowledge, the named authors have no conflict of interest, financial or otherwise.

Reference

Bapna, Ravi, Alok Gupta, Sarah Rice, and Arun Sundararajan. 2017. "Trust and the Strength of Ties in Online Social Networks: An Exploratory Field Experiment." *MIS Q.* 41 (1):115-30.

Barrett, C, and T Bradshaw. 2011. "Burberry in step with digital age." *Financial Times* 31.

Bastian, Hilda, Paul Glasziou, and Iain Chalmers. 2010. "Seventy-five trials and eleven systematic reviews a day: how will we ever keep up?" *PLoS medicine* 7 (9):e1000326.

Blei, David M, and John D Lafferty. 2009. "Topic models." In *Text mining*, 101-24. Chapman and Hall/CRC.

Bordia, Prashant, Nicholas DiFonzo, Robin Haines, and Elizabeth Chaseling. 2005. "Rumors denials as persuasive messages: Effects of personal relevance, source, and message characteristics 1." *Journal of applied social psychology* 35 (6):1301-31.

Bu, Xiangzhi, Zhoucheng Huang, and Quanwu Zhao. 2021. "Mining analysis of customer perceived value of online customisation experience under social commerce." *Enterprise Information Systems* 15 (10):1658-82.

Chang, Yung-Chun, Chih-Hao Ku, and Chun-Hung Chen. 2019. "Social media analytics: Extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor." *International Journal of Information Management* 48:263-79.

Chen, Wen-Kuo, Dalianus Riantama, and Long-Sheng Chen. 2020. "Using a text mining approach to hear voices of customers from social media toward the fast-food restaurant industry." *Sustainability* 13 (1):268.

Cho, Vincent, and Desmond Chan. 2021. "How social influence through information adoption from online review sites affects collective decision making." *Enterprise Information Systems* 15 (10):1562-86.

Choi, Tsan‐Ming, Stein W Wallace, and Yulan Wang. 2018. "Big data analytics in operations management." Review of. Production and Operations Management

27 (10):1868-83.

Davenport, Thomas H, Paul Barth, and Randy Bean. 2012. "How'big data'is different."

Deborah, C. . 2020. "Fourth Quarter 2020 Results Conference Call, page 10. Facebook." In.

Del Vecchio, Pasquale, Gioconda Mele, Valentina Ndou, and Giustina Secundo. 2018. "Creating value from social big data: Implications for smart tourism destinations." *Information Processing & Management* 54 (5):847-60.

Demeter, Krisztina, Levente Szász, and Andrea Kő. 2019. "A text mining based overview of inventory research in the ISIR special issues 1994–2016." *International Journal of Production Economics* 209:134-46.

Dijkmans, Corné, Peter Kerkhof, and Camiel J Beukeboom. 2015. "A stage to engage: Social media use and corporate reputation." *Tourism management* 47:58-67.

DiMaggio, Paul, Manish Nag, and David Blei. 2013. "Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of US government arts funding." *Poetics* 41 (6):570-606.

Du, Shuili, Goksel Yalcinkaya, and Ludwig Bstieler. 2016. "Sustainability, social media driven open innovation, and new product development performance." *Journal of product innovation management* 33:55-71.

Dwivedi, Yogesh K, and Navonil Mustafee. 2010. "Profiling research published in the Journal of Enterprise Information Management (JEIM)." *Journal of Enterprise Information Management*.

Eachempati, Prajwal, Praveen Ranjan Srivastava, and Zuopeng Justin Zhang. 2021.

"Gauging opinions about the COVID-19: a multi-channel social media approach." *Enterprise Information Systems* 15 (6):794-828.

Ert, Eyal, Aliza Fleischer, and Nathan Magen. 2016. "Trust and reputation in the sharing economy: The role of personal photos in Airbnb." *Tourism management* 55:62-73.

Evans, Jonathan St BT. 2003. "In two minds: dual-process accounts of reasoning." *Trends in cognitive sciences* 7 (10):454-9.

Eyre, Robert, Flavia De Luca, and Filippo Simini. 2020. "Social media usage reveals recovery of small businesses after natural hazard events." *Nature communications* 11 (1):1-10.

Fang, Bin, Qiang Ye, Deniz Kucukusta, and Rob Law. 2016. "Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics." *Tourism management* 52:498-506.

Feldman, Ronen, and James Sanger. 2007. *The text mining handbook: advanced approaches in analyzing unstructured data*: Cambridge university press.

Fernandes, Elizabeth, Sergio Moro, Paulo Cortez, Fernando Batista, and Ricardo Ribeiro. 2021. "A data-driven approach to measure restaurant performance by combining online reviews with historical sales data." *International Journal of Hospitality Management* 94:102830.

Fishbein, Martin, Icek Ajzen, and Attitude Belief. 1975. "Intention and Behavior: An introduction to theory and research." In.: Addison-Wesley, Reading, MA.

Fligstein, Neil, Jonah Stuart Brundage, and Michael Schultz. 2017. "Seeing like the Fed:

Culture, cognition, and framing in the failure to anticipate the financial crisis of 2008." *American Sociological Review* 82 (5):879-909.

Fox, Jesse, Katie M Warber, and Dana C Makstaller. 2013. "The role of Facebook in romantic relationship development: An exploration of Knapp's relational stage model." *Journal of Social and Personal Relationships* 30 (6):771-94.

Gal-Tzur, Ayelet, Susan M Grant-Muller, Tsvi Kuflik, Einat Minkov, Silvio Nocera, and Itay Shoor. 2014. "The potential of social media in delivering transport policy goals." *Transport Policy* 32:115-23.

Gong, Shiyang, Juanjuan Zhang, Ping Zhao, and Xuping Jiang. 2017. "Tweeting as a marketing tool: A field experiment in the TV industry." *Journal of Marketing Research* 54 (6):833-50.

Griffiths, Thomas L, and Mark Steyvers. 2004. "Finding scientific topics." *Proceedings of the National academy of Sciences* 101 (suppl 1):5228-35.

Gu, Bin, and Qiang Ye. 2014. "First step in social media: Measuring the influence of online management responses on customer satisfaction." *Production and Operations Management* 23 (4):570-82.

Guo, Lei, Chris J Vargo, Zixuan Pan, Weicong Ding, and Prakash Ishwar. 2016. "Big social data analytics in journalism and mass communication: Comparing dictionary-based text analysis and unsupervised topic modeling." *Journalism & Mass Communication Quarterly* 93 (2):332-59.

Hardey, Mariann. 2011. "To spin straw into gold? New lessons from consumer-generated content." *International Journal of Market Research* 53 (1):13-5.

Harrison, Glenn W, and John A List. 2004. "Field experiments." *Journal of Economic literature* 42 (4):1009-55.

He, Wu, Jiancheng Shen, Xin Tian, Yaohang Li, Vasudeva Akula, Gongjun Yan, and Ran Tao. 2015. "Gaining competitive intelligence from social media data: Evidence from two largest retail chains in the world." *Industrial management & data systems* 115 (9):1622.

He, Wu, Shenghua Zha, and Ling Li. 2013. "Social media competitive analysis and text mining: A case study in the pizza industry." *International Journal of Information Management* 33 (3):464-72.

Hemmatian, Fatemeh, and Mohammad Karim Sohrabi. 2019. "A survey on classification techniques for opinion mining and sentiment analysis." *Artificial Intelligence Review* 52 (3):1495-545.

Ibrahim, Noor Farizah, and Xiaojun Wang. 2019. "Decoding the sentiment dynamics of online retailing customers: Time series analysis of social media." *Computers in Human Behavior* 96:32-45.

Itani, Omar S, Raj Agnihotri, and Rebecca Dingus. 2017. "Social media use in B2b sales and its impact on competitive intelligence collection and adaptive selling: Examining the role of learning orientation as an enabler." *Industrial Marketing Management* 66:64-79.

Jalilvand, Mohammad Reza, and Neda Samiei. 2012. "The effect of electronic word of mouth on brand image and purchase intention: An empirical study in the automobile industry in Iran." *Marketing Intelligence & Planning*.

Kahneman, Daniel. 2011. *Thinking, fast and slow*: Macmillan.

Kauffmann, Erick, Jesús Peral, David Gil, Antonio Ferrández, Ricardo Sellers, and Higinio Mora. 2019. "Managing marketing decision-making with sentiment analysis: An evaluation of the main product features using text data mining." *Sustainability* 11 (15):4235.

Keupp, Marcus Matthias, Maximilian Palmié, and Oliver Gassmann. 2012. "The strategic management of innovation: A systematic review and paths for future research." *International journal of management reviews* 14 (4):367-90.

Krijestorac, Haris, Rajiv Garg, and Prabhudev Konana. 2021. "Decisions Under the Illusion of Objectivity: Digital Embeddedness and B2B Purchasing." *Production and Operations Management* 30 (7):2232-51.

Kumar, Prashant, Arun Sharma, and Jari Salo. 2019. "A bibliometric analysis of extended key account management literature." *Industrial Marketing Management* 82:276-92.

LaValle, Steve, Eric Lesser, Rebecca Shockley, Michael S Hopkins, and Nina Kruschwitz. 2011. "Big data, analytics and the path from insights to value." *MIT sloan management review* 52 (2):21-32.

Lepenioti, Katerina, Alexandros Bousdekis, Dimitris Apostolou, and Gregoris Mentzas. 2020. "Prescriptive analytics: Literature review and research challenges." *International Journal of Information Management* 50:57-70.

Li, Dun, Bangdong Zhi, Tobias Schoenherr, and Xiaojun Wang. 2023. "Developing Capabilities for Supply Chain Resilience in a Post-COVID World: A Machine

Learning based Thematic Analysis." Review of. *IISE Transactions* (just-accepted):1-34.

Li, Zhenlong, Cuizhen Wang, Christopher T Emrich, and Diansheng Guo. 2018. "A novel approach to leveraging social media for rapid flood mapping: a case study of the 2015 South Carolina floods." *Cartography and Geographic Information Science* 45 (2):97-110.

Liao, Chien-Hsiang, and Mu-Yen Chen. 2019. "Building social computing system in big data: From the perspective of social network analysis." *Computers in Human Behavior* 101:457-65.

Lustig, Irv, Brenda Dietrich, Christer Johnson, and Christopher Dziekan. 2010. "The analytics journey." *Analytics Magazine* 3 (6):11-3.

Mao, Yu, Yifan Zhu, Yiping Liu, Qika Lin, Hao Lu, and Fuquan Zhang. 2020. "Classifying user connections through social media avatars and users social activities: a case study in identifying sellers on social media." *Enterprise Information Systems*:1-20.

McAfee, Andrew, Erik Brynjolfsson, Thomas H Davenport, DJ Patil, and Dominic Barton. 2012. "Big data: the management revolution." *Harvard business review* 90 (10):60-8.

McKenna, Brad, Michael D Myers, and Michael Newman. 2017. "Social media in qualitative research: Challenges and recommendations." *Information and Organization* 27 (2):87-99.

Miles, Sandra Jeanquart, and W Glynn Mangold. 2014. "Employee voice: Untapped resource or social media time bomb?" *Business horizons* 57 (3):401-11.

Moro, Sérgio, Paulo Cortez, and Paulo Rita. 2015. "Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation." *Expert Systems with Applications* 42 (3):1314-24.

Müller, Oliver, Iris Junglas, Jan Vom Brocke, and Stefan Debortoli. 2016. "Utilizing big data analytics for information systems research: challenges, promises and guidelines." *European Journal of Information Systems* 25 (4):289-302.

Munar, Ana María, and Jens Kr Steen Jacobsen. 2014. "Motivations for sharing tourism experiences through social media." *Tourism management* 43:46-54.

Naeem, Muhammad. 2020. "Using social networking applications to facilitate change implementation processes: insights from organizational change stakeholders." *Business Process Management Journal*.

Nayak, Bimal Chandra, Gopal Krishna Nayak, and Debasish Jena. 2020. "Social recognition and employee engagement: The effect of social media in organizations." Review of. *International Journal of Engineering Business Management* 12:1847979020975109.

Nam, Hyoryung, and Pallassana Krishnan Kannan. 2014. "The informational value of social tagging networks." *Journal of marketing* 78 (4):21-40.

Negash, Solomon, and Paul Gray. 2008. "Business intelligence." In *Handbook on decision support systems 2*, 175-93. Springer.

Nguyen, Bang, Xiaoyu Yu, TC Melewar, and Junsong Chen. 2015. "Brand innovation and social media: Knowledge acquisition from social media, market orientation, and the moderating role of social media strategic capability." *Industrial Marketing Management* 51:11-25.

Oh, Onook, Manish Agrawal, and H Raghav Rao. 2013. "Community intelligence and social media services: A rumor theoretic analysis of tweets during social crises." *MIS quarterly*:407-26.

Palalic, Ramo, Veland Ramadani, Syedda Mariam Gilani, Shqipe Gërguri-Rashiti, and Leo–Paul Dana. 2021. "Social media and consumer buying behavior decision: what entrepreneurs should know?" Review of. *Management Decision* 59 (6):1249-70.

Palen, Leysia, and Kenneth M Anderson. 2016. "Crisis informatics—New data for extraordinary times." *Science* 353 (6296):224-5.

Pohl, Daniela, Abdelhamid Bouchachia, and Hermann Hellwagner. 2015. "Social media for crisis management: clustering approaches for sub-event detection." *Multimedia tools and applications* 74 (11):3901-32.

Rapp, David N, and Nikita A Salovich. 2018. "Can't we just disregard fake news? The consequences of exposure to inaccurate information." *Policy Insights from the Behavioral and Brain Sciences* 5 (2):232-9.

Roshan, Mina, Matthew Warren, and Rodney Carr. 2016. "Understanding the use of social media by organisations for crisis communication." *Computers in Human Behavior* 63:350-61.

Ruths, Derek, and Jürgen Pfeffer. 2014. "Social media for large studies of behavior." *Science* 346 (6213):1063-4.

Senadheera, Vindaya, Matthew Warren, and Shona Leitch. 2017. "Social media as an information system: improving the technological agility." *Enterprise Information Systems* 11 (4):512-33.

Sharma, Megha, Kapil Yadav, Nitika Yadav, and Keith C Ferdinand. 2017. "Zika virus pandemic—analysis of Facebook as a social media health information platform." *American journal of infection control* 45 (3):301-2.

Shi, Zhan, Huaxia Rui, and Andrew B Whinston. 2014. "Content sharing in a social broadcasting environment: evidence from twitter." *MIS quarterly* 38 (1):123-42.

Sivarajah, Uthayasankar, Muhammad Mustafa Kamal, Zahir Irani, and Vishanth Weerakkody. 2017. "Critical analysis of Big Data challenges and analytical methods." *Journal of Business Research* 70:263-86.

Soltanpoor, Reza, and Timos Sellis. 2016. Prescriptive analytics for big data. Paper presented at the Australasian Database Conference.

Steinhoff, Lena, Denni Arli, Scott Weaven, and Irina V Kozlenkova. 2019. "Online relationship marketing." *Journal of the Academy of Marketing Science* 47 (3):369-93.

Stelzner, M.A. 2021. "How marketers are using social media to grow their businesses." In *Social Media Marketing Industry Report*.

Stieglitz, Stefan, Milad Mirbabaie, Björn Ross, and Christoph Neuberger. 2018. "Social media analytics–Challenges in topic discovery, data collection, and data preparation." *International Journal of Information Management* 39:156-68.

Sun, Xiao, Chen Zhang, Guoqiang Li, Daniel Sun, Fuji Ren, Albert Zomaya, and Rajiv Ranjan. 2018. "Detecting users' anomalous emotion using social media for business intelligence." *Journal of Computational Science* 25:193-200.

Tim, H. . 2014. *Behavioural economics and public policy*. Financial Times.

Tirunillai, Seshadri, and Gerard J Tellis. 2014. "Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation." *Journal of Marketing Research* 51 (4):463-79.

Tóth, Zsófia, Martin Liu, Jun Luo, and Christos Braziotis. 2019. "The role of social media in managing supplier attractiveness: An investigation of business-to-business markets." *International Journal of Operations & Production Management*.

Tranfield, David, David Denyer, and Palminder Smart. 2003. "Towards a methodology for developing evidence‐informed management knowledge by means of systematic review." *British journal of management* 14 (3):207-22.

Tse, Ying Kei, Hanlin Loh, Juling Ding, and Minhao Zhang. 2018. "An investigation of social media data during a product recall scandal." *Enterprise Information Systems* 12 (6):733-51.

Tse, Ying Kei, Minhao Zhang, Bob Doherty, Paul Chappell, and Philip Garnett. 2016. "Insight from the horsemeat scandal: Exploring the consumers' opinion of tweets toward Tesco." *Industrial management & data systems*.

Rehman, Muhammad Habib, Victor Chang, Aisha Batool, and Teh Ying Wah. 2016. "Big data reduction framework for value creation in sustainable enterprises."

International Journal of Information Management 36 (6):917-28.

Van Eck, Nees Jan, and Ludo Waltman. 2010. "Software survey: VOSviewer, a computer program for bibliometric mapping." *scientometrics* 84 (2):523-38.

Verma, Jai Prakash, Smita Agrawal, Bankim Patel, and Atul Patel. 2016. "Big data analytics: challenges and applications for text, audio, video, and social media data”." *International Journal on Soft Computing, Artificial Intelligence and Applications (IJSCAI)* 5 (1):41-51.

Visentin, Marco, Gabriele Pizzi, and Marco Pichierri. 2019. "Fake news, real problems for brands: The impact of content truthfulness and source credibility on consumers' behavioral intentions toward the advertised brands." *Journal of Interactive Marketing* 45:99-112.

Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. "The spread of true and false news online." *Science* 359 (6380):1146-51.

Vu, Huy Quan, Gang Li, Rob Law, and Yanchun Zhang. 2019. "Exploring tourist dining preferences based on restaurant reviews." *Journal of Travel Research* 58 (1):149-67.

Waller, Matthew A, and Stanley E Fawcett. 2013. "Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management." In.: Wiley Online Library.

Wamba, Samuel Fosso, Angappa Gunasekaran, Shahriar Akter, Steven Ji-fan Ren, Rameshwar Dubey, and Stephen J Childe. 2017. "Big data analytics and firm performance: Effects of dynamic capabilities." *Journal of Business Research*

70:356-65.

Wang, Xu, Bart Baesens, and Zhen Zhu. 2019. "On the optimal marketing aggressiveness level of C2C sellers in social media: Evidence from china." *Omega* 85:83-93.

Wang, Yichuan, and Chiahui Yu. 2017. "Social interaction-based consumer decision-making model in social commerce: The role of word of mouth and observational learning." *International Journal of Information Management* 37 (3):179-89.

Wang, Yichuan, Minhao Zhang, Shuyang Li, Fraser McLeay, and Suraksha Gupta. 2021. "Corporate Responses to the Coronavirus Crisis and their Impact on Electronic - Word - of - Mouth and Trust Recovery: Evidence from Social Media." *British journal of management*.

West, Richard L, Lynn H Turner, and Gang Zhao. 2010. *Introducing communication theory: Analysis and application*. Vol. 2: McGraw-Hill New York, NY.

Xia, Lan. 2013. "Effects of companies' responses to consumer criticism in social media." *International Journal of Electronic Commerce* 17 (4):73-100.

Yang, Yefei, Ciwei Dong, Xin Yao, Peter KC Lee, and TCE Cheng. 2020. "Improving the effectiveness of social media-based crowdsourcing innovations: roles of assurance mechanism and innovator's behaviour." *Industrial Management & Data Systems*.

Yang, Jie, Pishi Xiu, Lipeng Sun, Limeng Ying, and Blaanand Muthu. 2022. "Social media data analytics for business decision making system to competitive analysis." *Review of. Information Processing & Management* 59 (1):102751.

Yasir, Muhammad, Sitara Afzal, Khalid Latif, Ghulam Mujtaba Chaudhary, Nazish Yameen Malik, Farhan Shahzad, and Oh-young Song. 2020. "An efficient deep learning based model to predict interest rate using twitter sentiment." *Sustainability* 12 (4):1660.

Yim, Eun Soon, Suna Lee, and Woo Gon Kim. 2014. "Determinants of a restaurant average meal price: An application of the hedonic pricing model." *International Journal of Hospitality Management* 39:11-20.

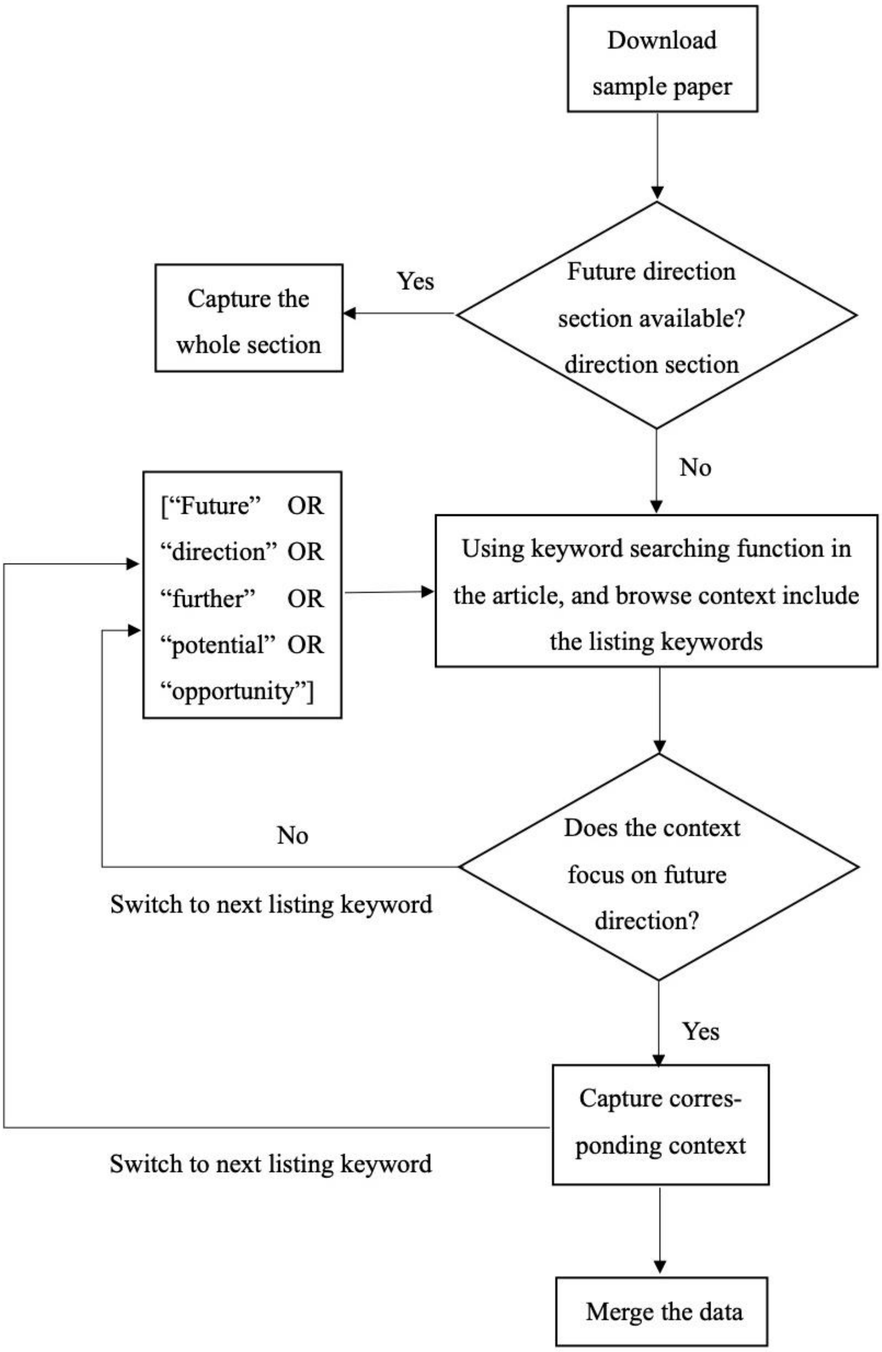
Yuan, Hui, Wei Xu, Qian Li, and Raymond Lau. 2018. "Topic sentiment mining for sales performance prediction in e-commerce." *Annals of Operations Research* 270 (1):553-76.

Zhang, Fuqiang, Xiaole Wu, Christopher S Tang, Tianjun Feng, and Yue Dai. 2020. "Evolution of operations management research: From managing flows to building capabilities." Review of. *Production and Operations Management* 29 (10):2219-29.

Zhang, Qiang, Wenbo Wang, and Yuxin Chen. 2020. "Frontiers: In-Consumption Social Listening with Moment-to-Moment Unstructured Data: The Case of Movie Appreciation and Live Comments." *Marketing Science* 39 (2):285-95.

Appendix

Appendix A. Corpora capture process about the future direction for LDA topic modelling



Appendix B. Average coherence score at different K settings.

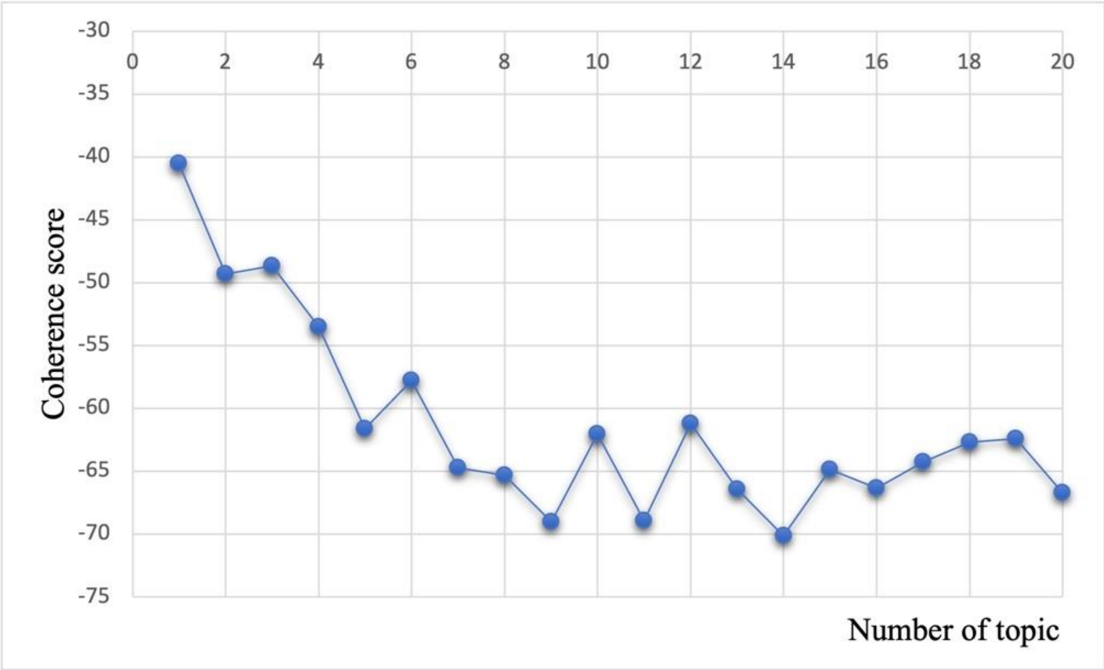


Table 1. Classification of publications

Type of publication	Freq	Percent
Research paper	42	47.19%
Conceptual paper	17	19.10%
Technical Paper	18	20.22%
Case study	9	10.11%
Literature review	2	2.25%
Viewpoint	1	1.12%
Total	89	100%

Table 2. Classification of research methods.

Research Methods	Count	Percent
Design research	26	29.21%
Conceptual/descriptive/theoretical	22	24.72%
Case study	12	13.48%
Survey	12	13.48%
Mixed	7	7.87%
Experiment	5	5.62%
Interview	2	2.25%
Literature analysis	2	2.25%
Viewpoint/ commentary	1	1.12%
Total	89	100%

Table 3. Summary and comparison of selected studies related to using SMD in supporting business decisions.

Author	Research focus	Method	Data types	Sampling	Theory/ model	Findings
Chang, Ku, and Chen (2019)	An integrated framework to visualise TripAdvisor and Google Trends data	Sentiment analysis, visual analytics	Hotel ratings and reviews	TripAdvisor	N/A	Authors propose an integrated framework to visualise data
Chen, Riantama, and Chen (2020)	Customers voice in fast-food restaurant industry	Text-mining	Social media comments	Trustpilot.com	N/A	Authors apply text-mining on restaurant reviews to help companies adapt to future similar epidemic situations.
Del Vecchio et al. (2018)	Online reviews in healthcare industry	Sentiment analysis and multi-criteria decision-making technique	Online reviews and comments	Twitter	Sentiment analysis model	Authors propose a sentiment analysis model that enables B2C and C2C commerce.
Fernandes et al. (2021)	Restaurant performance measurement	Text-mining, sentiment analysis	Online review of six restaurants	TripAdvisor	Bass model	Authors propose a sales forecast model to simplify the decision-making process of restaurant managers
Gal-Tzur et al. (2014)	Transport planning	Text-mining, hierarchical approach	Text data	Twitter	N/A	Authors report the significance and novelty of the text-mining technique in improving transport planning
He, Zha, and Li (2013)	Social media competitive analysis and text mining	Text-mining, case study	Social media data	Facebook, Twitter	Uses and gratifications theory	The value of social media competitive analysis.
Krijestorac, Garg, and Konana (2021)	B2B Purchasing	Structural Equation Model	Survey data	Purchasing managers	Information processing theory, dual process theory	Authors found that more digitally embedded buyers are more willing to adopt innovations.

Munar and Jacobsen (2014)	Tourism experiences sharing through social media	Chi-square statistic	Survey data	Leisure travellers	Social cognitive theory	Sharing practices through social media have lesser relevance as information sources for holiday decision-making.
Senadheera, Warren, and Leitch (2017)	social media strategy for improving business agility	Mixed methods	Longitudinal data from social media	Australian banks	Honeycomb model	Authors empirically test the Honeycomb model as a tool that enhances the technological agility of social media.
Steinhoff et al. (2019)	Online relationship marketing	Conceptual research	N/A	N/A	Evolving theory	Authors propose an evolving theory of online relationship marketing.
Tóth et al. (2019)	Managing supplier attractiveness by social media	Thematic approach with open coding	Interview data	Senior managers	N/A	There is an inverse U-shaped relationship between the intensity of the supplier's SM activity and its attractiveness
Wang and Yu (2017)	Social commerce	Structural Equation Model	Survey data	Recruited research participants	word-of-mouth theory, observational learning theory	WOM and WOM content, and observing other consumers' purchases significantly affect consumers' purchase intention
Wang, Baesens, and Zhu (2019)	Marketing aggressiveness level of C2C sellers	Text classification, statics test	Social media data	Microblogs posted by Taobao sellers	Uses and gratifications theory	There is an empirical relationship between the marketing aggressiveness level and the marketing popularity following an inverted U-shape curve
Yang et al. (2020)	Crowdsourcing innovations	Hierarchical regression	Solvers and trade records	zbj.com (crowdsourcing business platform)	Uncertainty reduction theory	There is a positive relationship between the effectiveness of assurance mechanism and solver's behaviour.
Yuan et al. (2018)	Sale prediction	Topic sentiment mining, case study	Reviews	E-commerce companies in China	Joint sentiment topic	Authors propose a novel methodology to extract consumers' sentiments from product reviews to enhance sales predicting performance.

Table 4. Topic modelling results.

Topic ID	Topic	Topic weight	Keywords
1	Marketing Group	835	busi, research, process, decis, market, make, technolog, inform, data, approach
2	Sample & Data Selection Group	751	studi, research, limit, effect, futur, sampl, data, find, consum, review
3	Sentiment Group	362	sentiment, analysi, featur, model, propos, perform, text, product, topic, review
4	Method Group	320	base, method, user, improv, propos, qualiti, analysi, algorithm, develop, learn
5	Practical Group	245	studi, model, data, predict, variabl, complex, price, mood, product, research
6	E-WOM Group	133	project, onlin, platform, content, environ, custom, union, right, geograph, emot
7	Management Group	93	risk, enterpris, wast, worker, polici, recycl, hotel, state, case, narr

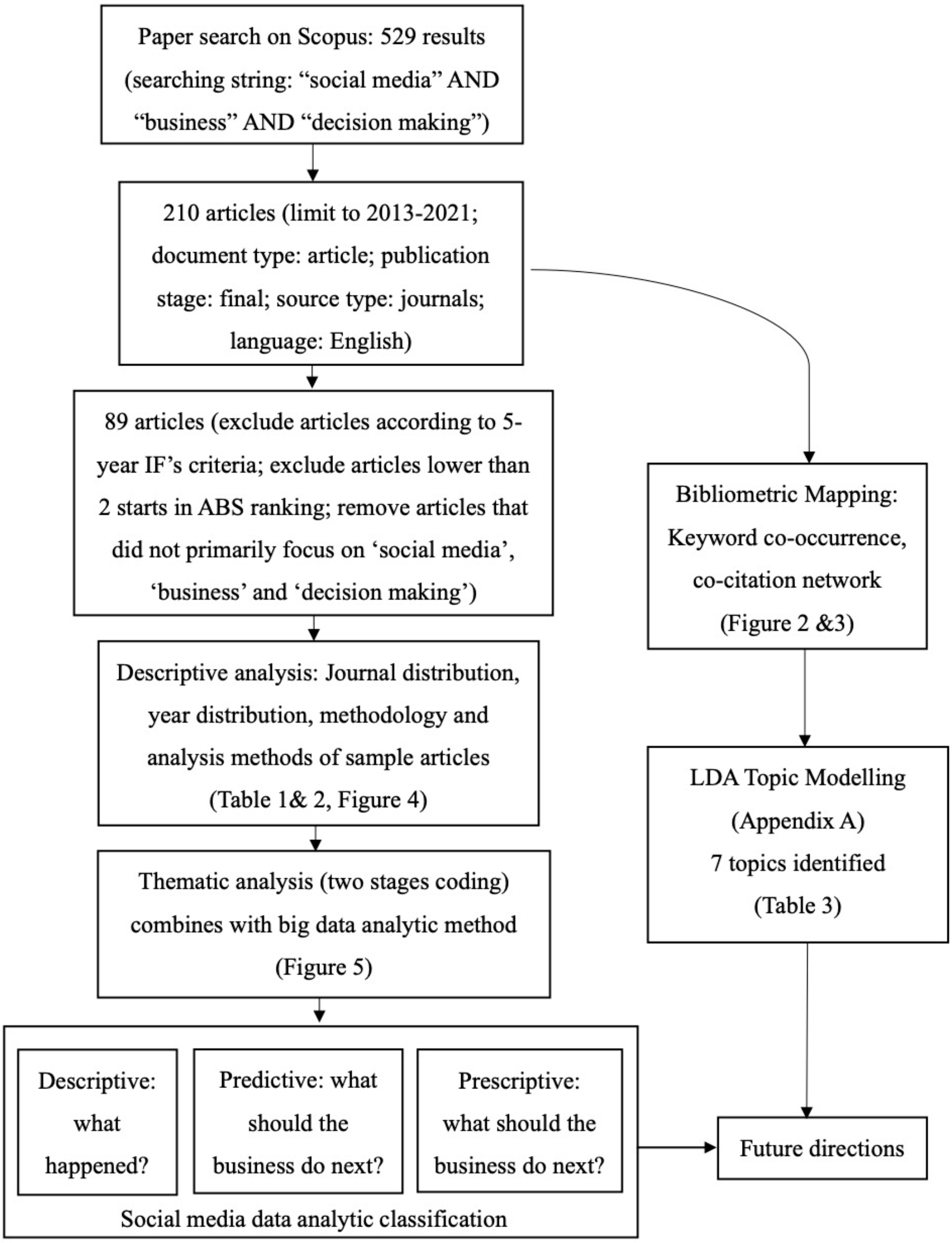


Figure 1. The research protocol.

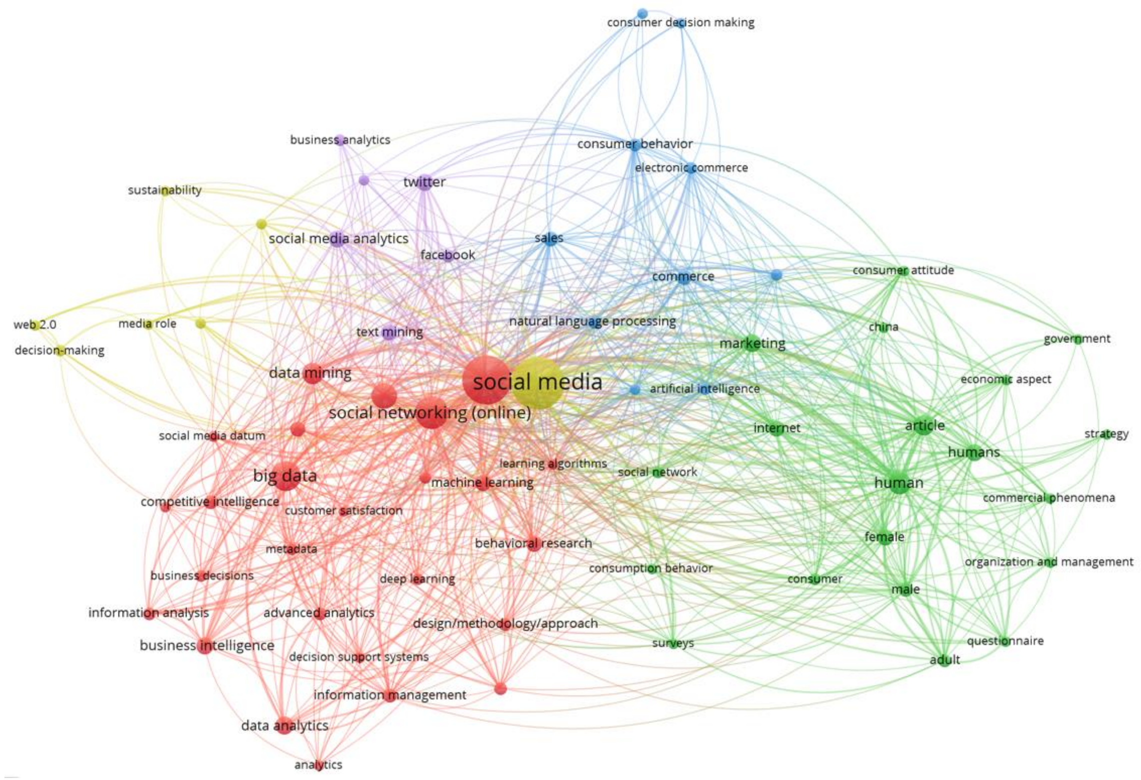


Figure 2. Keyword co-occurrence mapping (Note: the minimum number of a keyword is 4, 69 keywords meet the threshold).

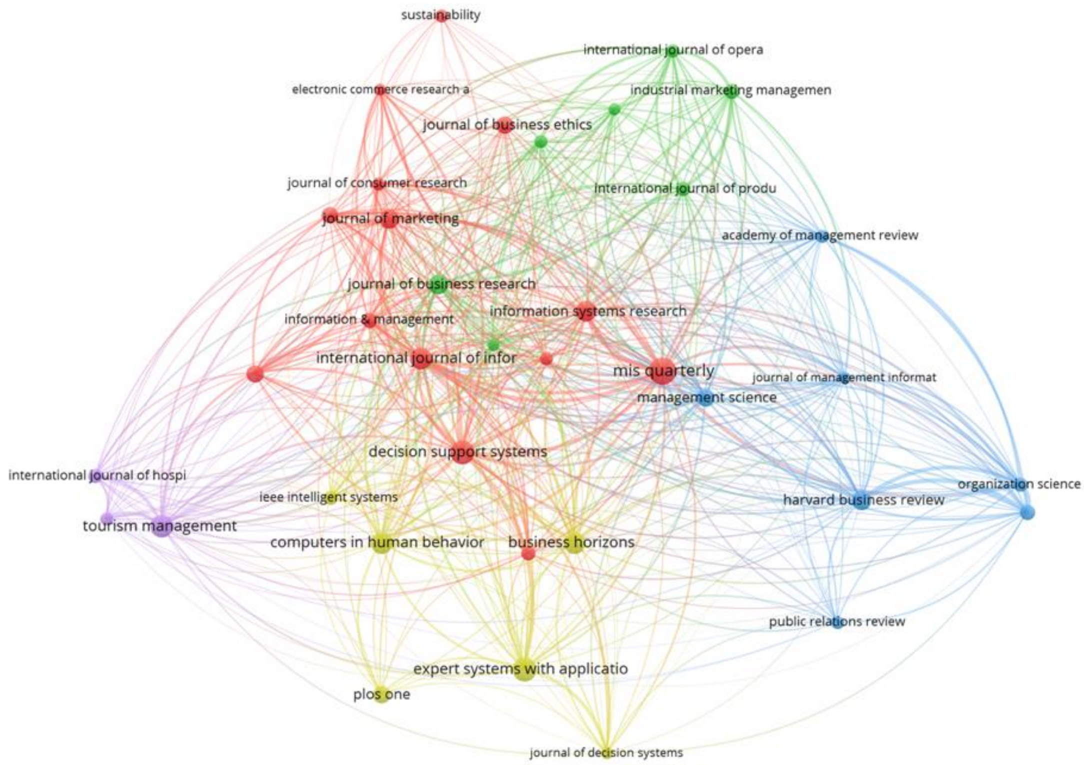


Figure 3. Journal co-citation network (Note: the minimum number of citations of a source is 20, 40 journals meet the threshold).

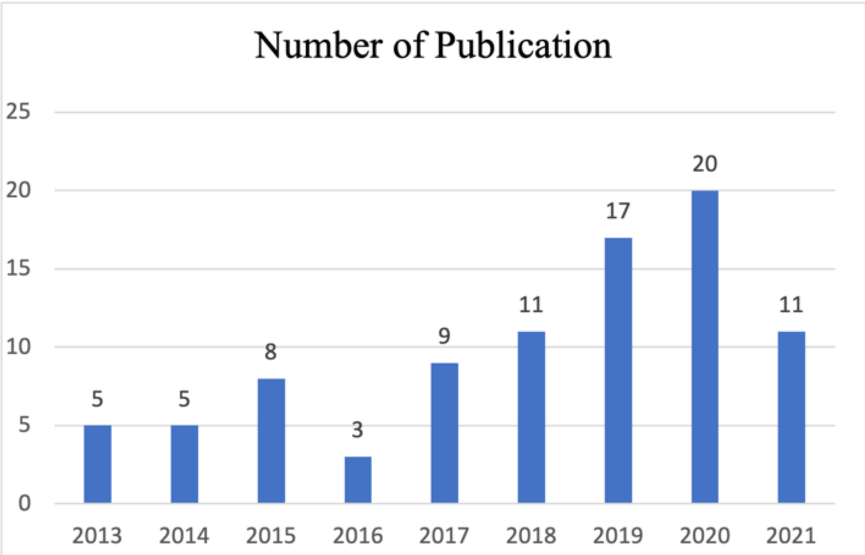


Figure 4. Time distribution of publications.

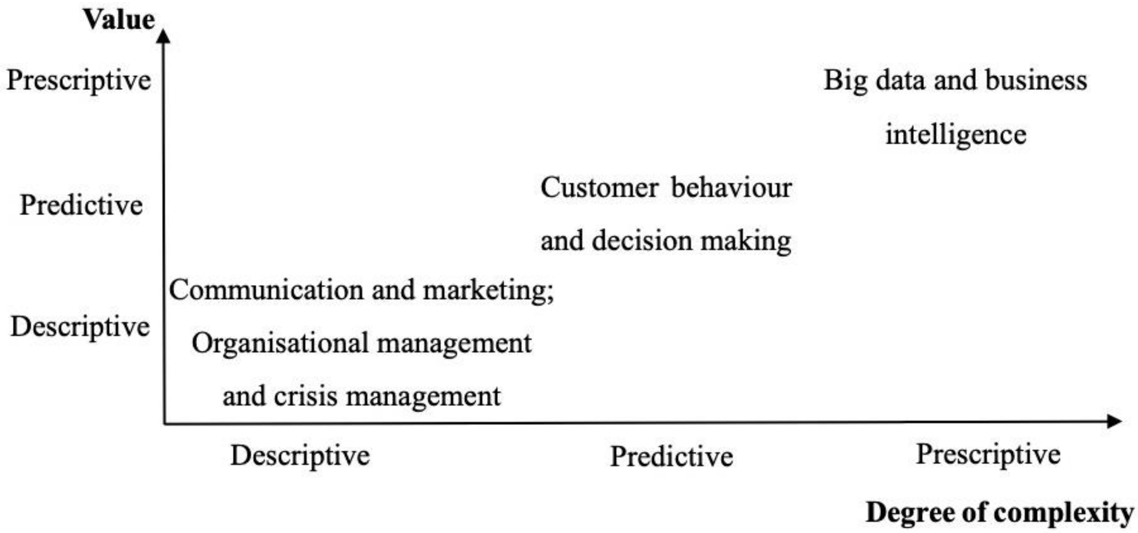


Figure 5. Key themes classification.