Social Big Data: Techniques and Recent Applications

Tina Tian *tina.tian@manhattan.edu*
Department of Computer Science
Manhattan College
New York, USA

Abstract

In the big data era, large volumes of social media data are generated at a high velocity, which we refer to as social big data. It is beyond the ability of traditional methods and algorithms to manage the massive amount of data in a tolerable elapsed time. In this paper, we present a comprehensive overview of the established big data techniques and new achievements on social big data management. The study also highlights a list of state-of-the-art applications based on data gathered from social networking platforms. At the end, we identify the key issues and challenges related to social big data analytics.

Keywords: Social Big Data, Big Data, Social Media Analytics, Social Media.

1. INTRODUCTION

With the rapid development of computing and networking techniques, social media have experienced a fast growth. Massive amount of data has been generated in real-time. Social network sites, such as Twitter and Facebook and other micro-blogging services, have provided a new kind of platform for information sharing. Web users all over the world can directly engage in these networks and share their opinions and perspectives. The content data are presented in the forms of texts, links, images, videos, etc., ranging from daily life stories to the latest local and global news and events [1]. The rich and continuously generated data provide tremendous value for users and organizations. The information gathered in the online communities and shared by their users constructs an important source of big data and provides valuable contribution to decision making.

Social big data is a collection of huge data sets with great diversity extracted from social networks. To better define big data and characterize its technological aspects, [2] proposes the three V model. The three Vs are Volume, Velocity and Variety. Additionally, another three Vs have been added to extend the set of variables. They are Veracity, Variability and Value. The latter three Vs have a stronger focus addressing the big data challenges and managerial perspectives [3] [4]. The 6V model is further explained as follows.

- Volume refers to the large amount of data that consumes excessive storage.
- Velocity represents the speed of data generation and the frequency of data delivery. Big data analysis requires a fast rate to keep up with the speed of data production.
- Variety addresses the importance that big data are generated from a great variety of sources, which may contain both structured and unstructured data.
- Veracity characterizes the quality of data (for example, uncertain and imprecise data) and the level of trust in different data sources.
- Variability stresses the unpredictability of big data.
- Value describes the transformation of information into insights that may create economic value for companies and organizations [5].

As an emerging paradigm, big data refers to any set of data with enormous capacity that traditional methods would require a large amount of time to process [6]. The sheer size of the dataset makes it practically impossible for typical database software tools to capture, store, manage and analyze [3]. Therefore, the management of big data requires advanced and technology-based analytical approaches and sophisticated algorithms and techniques.

In this paper, we provide an overview of the state-of-the-art methods and techniques to store and process social big data. A survey of recent achievements on social big data applications is being conducted, with a clear focus on applications published in the past few years. The paper also highlights the current challenges and security issues in social big data.

The rest of the paper is organized as follows. Section 2 summarizes the techniques and methods used in social big data and investigates the emerging frameworks and libraries. Section 3 describes the recent applications in the field, providing a list of achievements in the past five years. In Section 4, we discuss the open questions and challenges in social big data. Section 5 concludes the paper.

2. SOCIAL BIG DATA TECHNIQUES

Highly scalable and efficient methods are needed to process the increasingly large volume of social media data. High velocity of the big data production requires fast algorithms to avoid bottlenecks during data processing and analysis in order to generate meaningful results to aid decision making in real time. Not-Only Structured Query Language (NoSQL) is a revolutionary database management system that is widely used for big data processing. It stores large volumes of data and retrieves query results based on the consistency model, which is more flexible than the traditional relational databases. The NoSQL databases offer horizontal and vertical scalabilities that provide fast responses for data storage and retrieval [7]. Popular NoSQL database models include MongoDB [8], Cassandra [9], Amazon's DynamoDB [10], Google's BigTable [11], etc.

Hadoop [12] is an open-source framework that was specifically designed to handle big data. The main components of Hadoop are Hadoop Distributed File System (HDFS) and MapReduce. With HDFS, multiple machines that are remotely located can cooperate to achieve a common computational goal. MapReduce is programming model designed to separate operations across different logical units. As indicated by its name, MapReduce contains two key functions: map and reduce. The overall task is first divided into smaller subtasks containing autonomous data chunks. The map function executes each individual subtask in parallel, while the reduce function combines the data tuples produced by the map function and delivers the end result.

With increasing advancement in modeling and analysis of social big data, various methods and techniques are being exploited, such as cloud computing and parallel processing. Cloud computing has become a mainstream solution for big data storage, processing, and distribution [13]. It is designed to use vast computing and storage resources to provide the needed computing capacity for big data applications. The large amount of data requires a new processing approach by applying parallelism on readily available hardware [14]. Parallel computing has been widely adopted by scholars for processing large clusters of social media data in parallel. These system infrastructures facilitate scalable, accessible, and sustainable data streams to expand the processing ability of social big data.

The system infrastructures stated above help to store and load the social big data. To search for valuable information and to extract hidden patterns in the data, analytic methods are applied. Data analysis enables organizations to obtain information that can affect their businesses. Statistical models and machine learning methods are commonly adopted in social big data analysis. Often, they perform a predictive task, which is used to forecast future observations based on historical data. Popular predictive techniques include regression analysis, cluster analysis, decision tree, Bayesian network, Support Vector Machine, neural network, etc.

Regression analysis has been widely applied in many social big data applications. Two common regression models are autoregressive models [15] and moving average models [16]. Clustering is a process to allocate data records into groups. Five types of clustering algorithms are discussed in [17], which are partitioning-based, hierarchy-based, density-based, grid-based, and model-based, respectively. For example, [18] developed an advanced data-driven application based on density-based clustering models for real-time social media news classification.

Neural network is constructed with artificial neurons linked by weighted edges. The commonly used backpropagation algorithm [19] repeatedly adjusts the weight on the edges based on the difference between the predicted result and the actual value. The emerging deep learning models split data into groups and map them to separate layers in the network for processing. Deep learning algorithms have achieved remarkable results in social big data analysis. For example, sentiment analysis is one of the major tasks in social big data analysis which aims at interpreting social network users' sentiment tendency. A Sentiment polarity classifier is typically used to classify the data, which are categorized into positive, negative, and neutral preferences. [20] used a sentiment polarity classifier based on a deep learning model to investigate demands for regional events.

Software and tools have been developed to meet the increasing need for social big data mining and visualization. The expanding development of large-scale social networks is boosting the rise of new data analytic models and tools for big data techniques [13]. Tensor, a tool for big data analysis, has been adopted in various research work because of its prominent advantages in representing and handling complex and high-order data [21]. In [22], Liu et al. present an effective processing framework based on tensor networks to store and analyze social big data. [23] describes a method for managing big data from social network. Specifically, it is presented in distributed settings so as to support big data mining of frequently occurring patterns from social networks. [24] established an information diffusion model based on data from an online microblogging service. In their work, the impact of users' network activities was studied to enhance the accuracy of the model. Paul et al. introduced a high-level system architecture for large-scale data processing services [25]. Sandryhaila et al. discussed a possible paradigm for large-scale data analysis based on the discrete signal processing on graphs [26].

Vinay et al. proposed a new classifier, named Extreme Learning Machine classifier, to perform face tagging for social networks operating on big data [27]. Severyn et al. developed a sentiment polarity classifier for Twitter using deep convolutional neural networks [28]. In [29], a novel framework is described to deliver mobile big data in content-centric mobile social networks. [30] developed a hierarchical framework for feature extraction in social internet of things big data using MapReduce framework and a supervised classifier model. [31] presents an ontology-based approach as a means to extract semantics of textual social data. They semantically analyze tweets at both the entity level and the domain level. Ontologies are used to capture domain knowledge and enhance the semantics of tweets by providing the conceptual representation of entities. Information control and detection is a critical issue in social big data. In order to detect outliers in various complex datasets, [32] presents an outlier detection method by incorporating density-based and clustering-based methods. [33] describes a study case that creates a corporate knowledge base exploiting linked open data and social big data.

3. RECENT APPLICATIONS

In the previous section, we examined the system infrastructure and analytical techniques to store and manage social big data. Big data analysis can provide tremendous values in the decision-making process. It has brought increased attention among research scholars in various disciplines. In this section, we review several major application fields of social big data, focusing on recent achievements published in the past five years.

3.1 Social Big Data for Event Detection

Big data have been employed as a source of information for event detection, such as natural disasters. Social media plays an important role in disaster management by channeling emergency information to communities that are affected by the disaster. In the past, researchers have utilized social media data to better understand the characteristics and develop relief plans [34] [35]. Kim et al. explored patterns generated by interactions of online users on Facebook during the 2016 Louisiana flood [36]. Yoo et al. applied Twitter data during Hurricane Sandy to evaluate information diffusion speed and its determinants [37]. In [38], Social Big Board, a real-time monitoring system of social big data, is introduced for disaster management.

Crowdsourcing is a process of acquisition, integration, and analysis of big and heterogeneous data from diverse sources in urban spaces. Researchers have used social media content to detect urban emergency events, such as fires, storms, traffic jams, etc. [39] proposed a 5W model (What, Where, When, Who, and Why) to detect and describe real-time urban emergencies based on crowdsourcing, using data collected from Weibo, a microblogging service. The spatial and temporal information was extracted from the platform to detect real-time events.

3.2 Building Intelligent Systems

Big data for social transportation has brought unprecedented opportunities for building the next-generation intelligent transportation systems. [40] reviewed a suite of schemes that are demanded for utilizing big data in social transportation systems, including data sources, analytical approaches, and application systems. D'Andrea et al. proposed a real-time monitoring system for traffic event detection from Twitter stream analysis [41]. Internet of things enables ubiquitous information exchange and content sharing among vehicles. In [42], Zhou et al. studied how real-world social big data can play a key factor in building an intelligent transportation industry.

With the continuing growth of social media data, information overload has become a great challenge. Recommendation system is one of the emerging solutions to handle the vast volume of data. It supports users to quickly access their desired content. [43] proposed a multi-domain recommendation system based on multi-source social big data.

Social media provides a valuable data source for tourism information [44]. Social big data from user-generated content, together with their aggregation, integration, real time analysis and synchronization, have become the major drive for the value creation process in smart destinations. In [45], Vecchio et al. demonstrated how vast amount of social big data from tourists can nurture the value creation process for a smart tourism destination. The study explored a set of regional tourist experiences in a Southern European destination. The integration of social big data analysis, including preliminary interviews, clustering, and sentiment analysis, makes it possible to build an intelligent system catered for tourists. The ongoing monitoring of social media can identify critical features and take real-time corrective actions. [46] presents a case study in Barcelona on tourism analytics with massive user-generated content.

The value creation process in big data involves using analytical techniques and tools to deliver information that can identify hidden patterns and provide insights on customer experiences to aid decision making. Raguseo et al. explored the moderating effects of market positioning based on customers' reviews [47]. The study of [48] demonstrates the utility of big data analytics to better understand important hospitality issues, namely the relationship between hotel guest experience and satisfaction. The huge amount of data generated on social media by tourists serves as a valuable source of open innovation. [49] studies the digital records of tourist experiences in a destination in Southern Europe. The findings demonstrate how social big data can leverage the building of smart destination and support sustainable travel experiences.

In order to evaluate nature-based tourism and the features attractive to tourists, [50] applied social big data in investigating a protected area in an ASEAN heritage park. The research effectively illustrates the spatial patterns of visitation using 10 years of geo-tagged Flickr data.

The study offers insights into the applicability of social big data to protected area management and sustainable tourism.

3.3 Social Big Data in Economics

Social textual data have been used to predict economic growth. Yamada et al. introduced a systematic method to estimate the economic indicator by analyzing big data from social blogs [51]. Comparing to the announcement of the economic index from the government, which usually has a delay due to the time required to gather and analyze information, their approach can significantly reduce the time lag, producing results in nearly real-time. [52] studied Bitcoin OTC trust-weighted signed network datasets and monitored the Bitcoin economy via tracking users' financial situation using attached identities.

3.4 Big Data and Social Issues

Social influence analysis explores a quantifiable way to measure the influence of each social network user and it aims to identify the most influential users in the networks. Analyzing social networks enables fresh insights into how web users interact with each other. Studying the influencing mode among users provides a guidance in understand the way information propagates across the networks. [13] provides a comprehensive investigation of social influence analysis and studies the characteristics and architecture of social influence analysis based on data in large-scale social networks. Social influence analysis has significant social impact and application value, which help us understand social behaviors of people and promote communication of political, economic, and cultural activities [13].

Zhang et al. present a systematic study to investigate how online users' reposting activities in a popular online social network affect computer networks and users' offline behaviors [53]. In the study, they present a voting strategy that identifies the most influential users for information dissemination in mobile cellular networks. The traffic load in the backbone network was predicted by observing interactions among users.

In [54], Zhang et al. studied and predicted the trends and patterns linked with youth sexting in South Korea using social big data extracted from domestic online news sites, social networking sites, and forums. Social big data have the advantages to incorporate a large volume of data and to gather information of a diverse range of participants, which lend to more accurate predictions of social issues.

4. CHALLENGES

The rich content on social big data brings unprecedented opportunities and challenges on data acquisition, management, and analysis. In the reminder of this section, we summarize the key issues and research challenges in social big data.

4.1 Privacy and Security of Social Big Data

Privacy and information disclosure are one of the major concerns with the explosive growth of social big data and its emerging development. For example, the risk of users' spatial information disclosure has become a serious issue with the increasing use of the "check in" feature on social networks [55]. It is crucial to prevent any possible identification of personal data. One solution is to apply anonymization algorithms to protect individual privacy. While open data provide a valuable resource for organizations and scholars, their accessibility should be regulated in order to avoid anticompetitive business practices [56]. [57] provides a critical discussion on open research issues in the field of privacy and security of social big data, concentrating on algorithmic paradigms and model-oriented paradigms. [58] evaluates the Internet of Things security criteria, while [59] presents security issues pertaining to Internet of Things cloud. Neyaz et al. analyze the code behind social media apps, such as TikTok, and evaluate their robustness against attacks, privacy, and security [60].

With ever-increasing social big data being delivered on mobile devices, security issues should be taken into consideration. Besides the users' physical location as mentioned earlier, the wireless connection, which is used by mobile users to connect to the social networking platforms, should also be protected to avoid malicious attacks from three parties [61]. A new challenge has arisen in terms of how to allocate the security resources, for example, the computational resource to implement encryption in the wireless connection. Su et al. proposes a security-aware resource allocation scheme to deliver mobile social big data based on joint matching-coalition game [61]. In [62], Liang et al. discuss various security and privacy challenges of mobile social networking, including private information leakage, cheating detection, and sybil attacks.

In personalized recommendation systems, privacy of the users has raised a concern. To deliver the recommendation results, users' sensitive context information, such as their social status, age, and hobbies, may be exposed [63]. It is extremely challenging to keep a balance between the system accuracy and users' privacy preserving level. To tackle the problem of potential privacy leakage, [64] proposed a cloud-assisted private video recommendation system based on distributed online learning. The approach uses differential privacy to produce an efficient and highly accurate recommendation system.

4.2 Challenges of Text Mining In Social Media

Data quality on social networks is another issue. Social big data, for example Twitter streams, may contain a huge number of irrelevant and even polluted messages, including spams, advertisements, viruses, etc. [65] [66]. Social network studies indicate that over 70% of raw data sets may be noise or irrelevant message [67]. These contents significantly affect event detection performance. Weibo, another microblogging site, has become a tool for lawbreakers to diffuse false and illegal information, causing serious consequences [24]. Apart from spams and malicious information, a particular problem for social media is that discussions and comments can rapidly diverge into unrelated topics, resulting in irrelevant data being retrieved [68].

There are many underlying challenges in applying text mining and sentiment analysis techniques to social big data. For example, tweets particularly, do not produce the best results with traditional data mining techniques due to its short messages. The maximum capacity of 140 characters in a tweet does not allow each tweet to provide much contextual information and implicit knowledge. Tweets are also known to exhibit frequent occurrences of informal and less grammatical language, and a large number of misspelled words [69]. Tweets often contain emojis and abbreviations, which add further complications to the language variation. Irony and sarcasm are common in tweets, which are particularly difficult for machines to detect [68].

Moreover, data acquired from social media often contain biases. In [70], Hargittai discusses the methodological challenges of using big data retrieved from specific sites. Results reveal that "age, gender, race/ethnicity, socioeconomic status, online experiences, and Internet skills all influence the social network sites people use and thus where traces of their behavior show up" [70]. Crawford points out that hidden biases in both the big data collection and analysis stages present significant risks [71]. In her work, the use of tweets for studying Hurricane Sandy was brought up as an example. The data retrieved from social networks exhibit considerable gaps with little or no input from certain communities affected by the disaster. Harford addresses that social network users are not representative of the population as a whole [72]. Tufekci presents a series of methodological and conceptual challenges in social big data in [70]. Example issues include the over-emphasis of a single platform, sampling biases from selection by hashtags, and vague and unrepresentative sampling problem [73].

5. CONCLUSIONS

With the exponential growth of big data and rapid development of social media, social big data is gaining increasing attention in the research community. Recent technology revolutions have enabled social media data to be generated at an unprecedented speed. In this paper, we review the background and the state-of-the-art of social big data. The massive amount of data is characterized by the multi-V model, which are Volume, Velocity, Variety, Veracity, Variability and Value. When data are measured in terabytes or petabytes, traditional data computing models can no longer meet the needs. New forms of techniques are demanded for storing, querying, processing, and analyzing social big data to improve the decision-making process. Modern data warehouses such as NoSQL databases provide a solution to store large quantities of data. Distributed systems use parallel computing and the MapReduce framework to support parallel processing of different data at the same time. Due to the complexity of the data that need to be analyzed and the scalability of the underlying algorithms that support such processes, social big data processing is becoming a challenging task [74] [75]. To handle social big data efficiently, statistical models and machine learning algorithms are applied for data analysis. Besides the established data processing methods, our study also investigated newly developed data analytic techniques.

The survey paper aims to provide a comprehensive overview of this emerging field. Numerous representative applications of social big data were reviewed in Section 3, including using social network data for event detection, building intelligent systems, and applications that are linked to economics and social issues. The big data era creates opportunities in data applications and management advancement, but it also brings challenges to data processing and real-time analytics. The study discussed open questions and key challenges in social big data, including privacy issues and challenges in terms of data quality and underlying problems in text mining of the social big data.

6. REFERENCES

[1] F. Atefeh and W. Khreich. “A Survey of techniques for event detection in Twitter.” Computer intelligence, vol. 31, no. 1, 2015.

[2] D. Laney. “3D data management: controlling data volume, velocity, and variety.” META group research note, vol. 6, no. 70, 2001.

[3] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. Hung Byers. “Big data: The next frontier for innovation, competition, and productivity.” McKinsey Global Institute, 2011.

[4] A. McAfee and E. Brynjolfsson. “Big data. The management revolution.” Harvard business review, vol. 90, no. 10, pp. 60-68, 2012.

[5] A. De Mauro, M. Greco, and M. Grimaldi. "A formal definition of big data based on its essential features." Library Review, 2016.

[6] A. Gandomi and M. Haider, M. “Beyond the hype: big data concepts, methods, and analytics.” International Journal of Information Management, vol. 35, no. 2, pp. 137–144, 2015.

[7] I, Ha, B. Back and B. Ahn, “MapReduce functions to analyze sentiment information from social big data.” International Journal of Distributed Sensor Networks, vol. 11, no. 6, pp. 417502, 2015.

[8] MongoDB, https://www.mongodb.com/, retrieved on Oct. 31, 2020.

[9] Cassandra, https://cassandra.apache.org/, retrieved on Oct. 31, 2020.

[10] DynamoDB, https://aws.amazon.com/dynamodb/, retrieved on Oct. 31, 2020.

[11] BigTable, https://cloud.google.com/bigtable, retrieved on Oct. 31, 2020.

[12] Hadoop, http://hadoop.apache.org/, retrieved on Oct. 31, 2020.

[13] S. Peng, G. Wang and D. Xie. "Social influence analysis in social networking big data: Opportunities and challenges." IEEE network, vol. 31, no. 1, pp. 11-17, 2016.

[14] D. Lian and Y. Xiong. "Big data analytics and business analytics." Journal of Management Analytics, vol. 2, no. 1, pp. 1-21, 2015.

[15] H. Akaike. “Fitting autoregressive models for prediction.” Annals of the institute of Statistical Mathematics, vol. 21, no. 1, pp. 243-247, 1969.

[16] S.E. Said, and D.A. Dickey. “Testing for unit roots in autoregressive-moving average models of unknown order.: Biometrika, vol. 71, no. 3, pp. 599-607, 1984.

[17] A. Fahad, N. Alshatri, Z. Tari, A. Alamri, I. Khalil, A.Y. Zomaya, S. Foufou, and A. Bouras. “A survey of clustering algorithms for big data: Taxonomy and empirical analysis.” IEEE transactions on emerging topics in computing, vol. 2 no. 3, pp.267-279, 2014.

[18] C.H. Lee and T.F. Chien. “Leveraging microblogging big data with a modified density-based clustering approach for event awareness and topic ranking.” Journal of Information Science, vol. 39, no. 4, pp. 523–543, 2013.

[19] A.E. Bryson. “Applied optimal control: optimization, estimation and control.” CRC Press, 1975.

[20] T. Ohbe, O. Tadachika and S. Toramatsu. "A sentiment polarity classifier for regional event reputation analysis." In proceedings of the international conference on web intelligence, 2017.

[21] L. Kuang, L.T. Yang and Y. Liao, “An integration framework on cloud for cyber physical social systems big data,” IEEE Transactions on Cloud Computing, 2015.

[22] H. Liu, L.T. Yang, Y. Guo, X. Xie, J. Ma. “An incremental tensor-train decomposition for cyber-physical-social big data.” IEEE Transactions on Big Data, 2018.

[23] C.K. Leung, H. Zhang. “Management of distributed big data for social networks.” In 2016 16th IEEE/ACM international symposium on cluster, cloud and grid computing, pp. 639-648, 2016.

[24] D. Jin, X. Ma, Y. Zhang, H. Abbas and H. Yu. “Information diffusion model based on social big data.” Mobile networks and applications, vol. 23, no. 4, pp. 717-722, 2018.

[25] A. Paul, A. Awais, R. Mazhar and J. Sohail. "Smartbuddy: defining human behaviors using big data analytics in social internet of things." IEEE Wireless communications vol. 23, no. 5, pp. 68-74, 2016.

[26] A. Sandryhaila and J. Moura. "Big data analysis with signal processing on graphs: representation and processing of massive data sets with irregular structure." IEEE signal processing magazine, vol. 31, no. 5, pp. 80-90, 2014.

[27] A. Vinay, S. Vinay, J. Rituparna, A. Tushar, K. Balasubramanya and S. Murthya. "Cloud based big data analytics framework for face recognition in social networks using machine learning." Procedia Computer Science, vol. 50, pp. 623-630, 2015.

[28] A. Severyn and A. Moschitti. “Twitter sentiment analysis with deep convolutional neural networks.” In proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval, pp. 959–962, 2015.

[29] Z. Su, Q. Xu and Q. Qi. "Big data in mobile social networks: A QoE-oriented framework." IEEE network, vol. 30, no. 1, pp. 52-57, 2016.

[30] S.K. Lakshmanaprabu, K. Shankar, A. Khanna, D. Gupta, J. Rodrigues, P.R. Pinheiro, and V. Albuquerque. “Effective features to classify big data using social internet of things.” IEEE access, vol. 6, pp. 24196-24204, 2018.

[31] P. Wongthongtham and B.A. Salih. "Ontology-based approach for identifying the credibility domain in social Big Data." Journal of organizational computing and electronic commerce, vol. 28, no.4, pp. 354-377, 2018.

[32] J. He and N. Xiong. "An effective information detection method for social big data." Multimedia tools and applications, vol. 77, no. 9, pp. 11277-11305, 2018.

[33] A.I. Torre-Bastida, E. Villar-Rodriguez, J. Del Ser, and S. Gil-Lopez. “Semantic information fusion of linked open data and social big data for the creation of an extended corporate CRM database.” Intelligent distributed computing VIII, pp. 211-221, 2015.

[34] J. Kim and M. Hastak. “Social Network Analysis: the role of social media after a disaster.” 10th anniversary homeland defense/security education summit, 2017.

[35] J. Yin, S. Karimi, A. Lampert, M. Cameron, B. Robinson and R. Power. “Using social media to enhance emergency situation awareness.” International joint conference on artificial intelligence, pp. 4234–4239, 2015.

[36] J. Kim and M. Hastak. "Social network analysis: Characteristics of online social networks after a disaster." International journal of information management, vol. 38, no. 1, pp. 86-96, 2018.

[37] E. Yoo, W. Rand, M. Eftekhar, and E. Rabinovich. “Evaluating information diffusion speed and its determinants in social media networks during humanitarian crises.” Journal of operations management, vol. 45, pp. 123–133, 2016.

[38] S. Choi and B. Bae. "The real-time monitoring system of social big data for disaster management." Computer science and its applications, pp. 809-815, 2015.

[39] Z. Xu, Y. Liu, N. Yen, L. Mei, X. Luo, X. Wei and C. Hu. “Crowdsourcing based description of urban emergency events using social media big data.” IEEE Transactions on Cloud Computing, 2016.

[40] X. Zheng, W. Chen, P. Wang, D. Shen, S. Chen, X. Wang, Q. Zhang and L. Yang. "Big data for social transportation." IEEE transactions on intelligent transportation systems, vol. 17, no. 3, pp. 620-630, 2015.

[41] E. D'Andrea, P. Ducange, B. Lazzerini, and F. Marcelloni. “Real-time detection of traffic from twitter stream analysis.” IEEE transactions on intelligent transportation systems, vol. 16, no. 4, pp. 2269-2283, 2015.

[42] Z. Zhou, C. Gao, C. Xu, Y. Zhang, S. Mumtaz, and J. Rodriguez. “Social big-data-based content dissemination in internet of vehicles.” IEEE transactions on industrial informatics, vol. 14, no. 2, pp. 768-777, 2017.

[43] Y. Zhang, X. Ma, S. Wan, H. Abbas, and M. Guizani. “CrossRec: Cross-domain recommendations based on social big data and cognitive computing.” Mobile networks and applications, vol. 23, no. 6, pp. 1610-1623, 2018.

[44] S. J. Miah, H.Q. Vu, J. Gammack and M. McGrath. “A big data analytics method for tourist behavior analysis.” Information and management, vol. 54, no. 6, pp. 771–785, 2016.

[45] P. Del Vecchio, G. Mele, V. Ndou, and G. Secundo. “Creating value from social big data: Implications for smart tourism destinations.” Information processing & management, vol. 54, no. 5, pp. 847-860, 2018.

[46] E. Marine-Roig and S.A. Clavé. “Tourism analytics with massive user-generated content: A case study of Barcelona.” Journal of destination marketing & management, vol. 4, no. 3, pp. 162–172, 2015.

[47] P. Neirotti, E. Raguseo, and E. Paolucci. “Are customers’ reviews creating value in the hospitality industry? Exploring the moderating effects of market positioning.” International journal of information management, vol. 36, no. 6, pp. 1133–1143, 2016.

[48] Z. Xiang, Z. Schwartz, J.H. Gerdes and M. Uysal. “What can Big Data and text analytics tell us about hotel guest experience and satisfaction?” International journal of hospitality management, vo. 44, pp. 120–130, 2015.

[49] P. Del Vecchio, G. Mele, V. Ndou, and G. Secundo. “Open innovation and social big data for sustainability: evidence from the tourism industry.” Sustainability, vol. 10, no. 9, pp. 3215, 2018.

[50] Y. Kim, C.K. Kim, D.K. Lee, H.W. Lee, and R.I.T. Andrada. “Quantifying nature-based tourism in protected areas in developing countries by using social big data.” Tourism management, vol. 72, pp. 249-256, 2019.

[51] K. Yamada, H. Takayasu and M. Takayasu. "Estimation of economic indicator announced by government from social big data." Entropy, vol. 20, no. 11, pp. 852, 2018.

[52] M. Spagnuolo, F. Maggi, F and S. Zanero, S. “Bitiodine: Extracting intelligence from the bitcoin network.” International conference on financial cryptography and data security, 2014.

[53] Y. Zhang, Z. Li, C. Gao, K. Bian, L. Song, S. Dong, and X. Li. “Mobile social big data: Wechat moments dataset, network applications, and opportunities.” IEEE network, vol. 32, no. 3, pp. 146-153, 2018.

[54] J. Song, T. Song, and J. Lee. "Stay alert: forecasting the risks of sexting in Korea using social big data." Computers in human behavior vol. 81, pp. 294-302, 2018.

[55] M. Tsou and M. Leitner. “Visualization of social media: seeing a mirage or a message?” Cartography and geographic information science, vol. 40, no. 2, pp. 55–60, 2013.

[56] L. Manovich. “Trending: the promises and the challenges of big social data.” Debates in the digital humanities, pp. 460–475, 2012.

[57] A. Cuzzocrea. "Privacy and security of big data: current challenges and future research perspectives." In proceedings of the first international workshop on privacy and security of big data, 2014.

[58] C. Uppin and A. Sudhir. "A comprehensive review for security analysis of IoT platforms." International journal of computer science and security, vol. 14, no. 4, 2020.

[59] N. Almolhis, M. Haney, F. Alqhtani, and K. Makdi. "Discovering and understanding the security issues in IoT cloud." International journal of computer science and security, vol. 13, no. 6, 2019.

[60] A. Neyaz, A. Kumar, S. Krishnan, J. Placker and Q. Liu. "Security, privacy and steganographic analysis of FaceApp and TikTok." International journal of computer science and security, vol 14, no. 2, 2020.

[61] Z. Su and Q. Xu. "Security-aware resource allocation for mobile social big data: A matching-coalitional game solution." IEEE transactions on big data, 2017.

[62] X. Liang, K. Zhang, X. Shen and X. Lin, "Security and privacy in mobile social networks challenges and solutions." IEEE wireless communications, vol. 21, no. 1, pp: 33-41, 2014.

[63] A. Jeckmans, M. Beye, Z. Erkin, P. Hartel, R. Lagendijk and Q. Tang "Privacy in recommender systems," Social media retrieval, pp. 263-281, 2013.

[64] P. Zhou, Y. Zhou, D. Wu, and H. Jin. "Differentially private online learning for cloud-based video recommendation with multimedia big data in social networks." IEEE transactions on multimedia, vol. 18, no. 6, pp. 1217-1229, 2016.

[65] J. Hurlock and M. Wilson." Searching Twitter: separating the tweet from the chaff." In international AAAI conference on weblogs and social media, 2011.

[66] K.B. Lee and J. Caverlee. "Seven months with the devils: A long-term study of content polluters on Twitter." In international AAAI conference on weblogs and social media, 2011.

[67] M.H. Tsou. "Research challenges and opportunities in mapping social media and big data." Cartography and geographic information science, vol. 42, pp. 70-74, 2015.

[68] D. Maynard, Diana, K. Bontcheva, and D. Rout. "Challenges in developing opinion mining tools for social media." In proceedings of the @ NLP can u tag# usergeneratedcontent, pp. 15-22, 2012.

[69] D. Metzler, C. Cai and E.H. Hovy. "Structured event retrieval over microblog archives." In proceedings of the 2012 conference of the north American chapter of the association for computational linguistics: human language technologies, pp. 646–655, 2012.

[70] E. Hargittai. "Is bigger always better? Potential biases of big data derived from social network sites." The annals of the American academy of political and social science, vol. 659, no.1, pp. 63-76, 2015.

[71] K. Crawford. "The hidden biases of big data." Harvard business review, 2013.

[72] T. Harford. "Big data: Are we making a big mistake?" FT magazine, 2014.

[73] Z. Tufekci. "Big questions for social media big data: representativeness, validity and other methodological pitfalls." arXiv preprint arXiv:1403.7400, 2014.

[74] A. Labrinidis and H. Jagadish, "Challenges and opportunities with big data," In proceedings of the VLDB endowment, vol. 5, no.12, pp. 2032–2033, 2012.

[75] N. Khan, I. Yaqoob, I. Hashem, Z. Inayat, W.K. Mahmoud Ali, M. Alam, M. Shiraz, and A. Gani. "Big data: survey, technologies, opportunities, and challenges." The scientific world journal, 2014.