International Journal of Data and Network Science 3 (2019) 145–164

Contents lists available at GrowingScience

International Journal of Data and Network Science

homepage: www.GrowingScience.com/ijds

Big data and social media: A scientometrics analysis

Hossein Jelvehgaran Esfahani[a], Keyvan Tavasoli[a] and Armin Jabbarzadeh[a*]

[a]*Business School, McMaster University, Ontario, Canada*

C H R O N I C L E

Article history:
Received: October 29, 2018
Received in revised format: January 21, 2019
Accepted: February 8, 2019
Available online:
February 9, 2019

Keywords:
Social media
Social networking
Big data
Big data analytics
Scientometrics
Bibliometric
Bibliometrix R-package

A B S T R A C T

The purpose of this research is to investigate the status and the evolution of the scientific studies for the effect of social networks on big data and usage of big data for modeling the social networks users' behavior. This paper presents a comprehensive review of the studies associated with big data in social media. The study uses Scopus database as a primary search engine and covers 2000 of highly cited articles over the period 2012-2019. The records are statistically analyzed and categorized in terms of different criteria. The findings show that researches have grown exponentially since 2014 and the trend has continued at relatively stable rates. Based on the survey, decision support systems is the key-word which has carried the highest densities followed by heuristics methods. Among the most cited articles, papers published by re-searchers in United States have received the highest citations (7548), followed by United Kingdom (588) and China with 543 citations. Thematic analysis shows that the subject nearly maintained an important and well-developed research field and for better results we can merge our research with "big data analytics" and "twitter" that are important topics in this field but not developed well.

1. Introduction

The era of Big Data is underway, computer scientists, physicists, economists, mathematicians, political scientists, bio-informaticists, sociologists, and other scholars are clamoring for access to the massive quantities of information produced by and about people, things, and their interactions (Boyd et al., 2012). Parliamentary office of science and technology in its journal Houses of parliament, number 460 March 2014 write an article and brought some truths about social media and big data: 57% of over-16s in the UK use social media, generating vast amounts of accessible data. Analyzing social media data can help organizations understand behaviors and target products and services more effectively. Key applications include profiling voters and complementing traditional polling, targeting adverts at consumers, credit scoring and informing policing decisions. There is a debate about how to analyze social media data, including which methods to use and how to control for biases. Personal data can be shared or sold with

* Corresponding author.
E-mail address: Jabbarza@mcmaster.ca (A. Jabbarzadeh)

doi: 10.5267/j.ijdns.2019.2.007

users' consent as long as they are anonymized. There are concerns that users are not fully aware of how their data are being used and that it is often possible to identify individuals from linking anonymized datasets. Analyzing large quantities of readily available data from social media has created new opportunities to understand and influence how people think and act. The rate of unstructured data production on social media makes it difficult to analyze using traditional methods that rely on human analysts. Social media analytics is a new field of study that is developing automated or semi-automated methods for analyzing data. Some advocates of big data argue that the sheer size of the datasets reduces, or even eliminates, the need for established statistical methods such as random sampling, because all the data can be analyzed. However, in the case of social media data, it only contains data about people that use social media. In the UK, around 49% of the population use Facebook and 24% use Twitter and not all users create content. There are concerns that social media data may not represent vulnerable groups in society, such as the elderly or those from lower income backgrounds. This means that there are significant gaps in the data, and there are not yet accepted methods for controlling for biases.

This paper presents an overview on studies associated with big data in social media. The study uses Scopus database as a primary search engine and analyzes the data over the period 2012-2019.
In this article we use science mapping technic with Bibliometrix R-package that performing bibliometric analysis and building data matrices for co-citation, coupling, scientific collaboration analysis and co-word analysis on topic of use of big data in social media.

Table 1
The main information and summary

Description	Results
Documents	2000
Sources (Journals, Books, etc.)	1077
Keywords Plus (ID)	7500
Author's Keywords (DE)	4496
Period	2012 - 2019
Average citations per documents	8.467
Authors	4979
Author Appearances	6362
Authors of single-authored documents	241
Authors of multi-authored documents	4738
Single-authored documents	296
Documents per Author	0.402
Authors per Document	2.49
Co-Authors per Documents	3.18
Collaboration Index	2.78
Document types	
ARTICLE	754
ARTICLE IN PRESS	70
BOOK	34
BOOK CHAPTER	77
CONFERENCE PAPER	900
CONFERENCE REVIEW	37
EDITORIAL	20
ERRATUM	1
LETTER	3
NOTE	19
REVIEW	80
SHORT SURVEY	5

2. About Bibliometrix R-package

Science mapping is complex and confusing because it is multi-step and frequently requires numerous and diverse software tools. Bibliometrix R-package is a tool for quantitative research in scientometrics and bibliometrics. Bibliometrix package provides various routines for importing bibliographic data from Scopus, Clarivate Analytics' Web of Science, PubMed and Cochrane databases, performing bibliometric analysis and building data matrices for co-citation, coupling, scientific collaboration analysis and co-word analysis (Aria et al., 2017).

3. Most cited countries

Our survey demonstrates that United States maintained the most contribution in the field of big data in social media, followed by United Kingdom and China. Table 2 shows details of our survey.

Table 2
The summary of the contributions of different countries:

Country	Total Citations	Average Article Citations
USA	7548	19.454
UNITED KINGDOM	588	8.4
CHINA	543	5.902
AUSTRALIA	398	7.96
KOREA	352	6.769
GERMANY	327	10.548
INDIA	282	2.35
ITALY	236	4.291
SPAIN	174	6.96
HONG KONG	151	6.04
MALAYSIA	139	6.043
CANADA	130	5.417
POLAND	129	25.8
NETHERLANDS	113	6.647
GREECE	107	5.35
DENMARK	104	5.778
TAIWAN	92	3.286
NEW ZEALAND	75	15
SINGAPORE	71	6.455
FRANCE	58	4.143
JAPAN	51	2.217
SWEDEN	48	12
AUSTRIA	43	8.6
NORWAY	36	12
INDONESIA	30	2.143
IRELAND	29	7.25
ISRAEL	28	4.667
CZECH REPUBLIC	25	8.333
IRAN	20	6.667
MOROCCO	19	1.9
URUGUAY	19	19
ROMANIA	18	9
ALGERIA	17	17
FINLAND	16	2
PAKISTAN	15	1.875
SAUDI ARABIA	15	2.143
CROATIA	14	7
TURKEY	14	1.167
BRAZIL	10	0.909
MEXICO	9	4.5
SWITZERLAND	9	2.25
SRI LANKA	8	4
TUNISIA	8	2.667
CHILE	6	6
CYPRUS	6	1.5
NIGERIA	6	6
BELGIUM	5	1.667
OMAN	5	5
QATAR	5	2.5
SOUTH AFRICA	5	0.625

According to Table 2, researchers from USA have published 7548 papers followed by United Kingdom with 588 papers and China with 543 papers. In terms of the average citation, papers published by researchers in Poland and USA have maintained the highest citations. Fig. 1 shows the results of the collaborations among various countries.

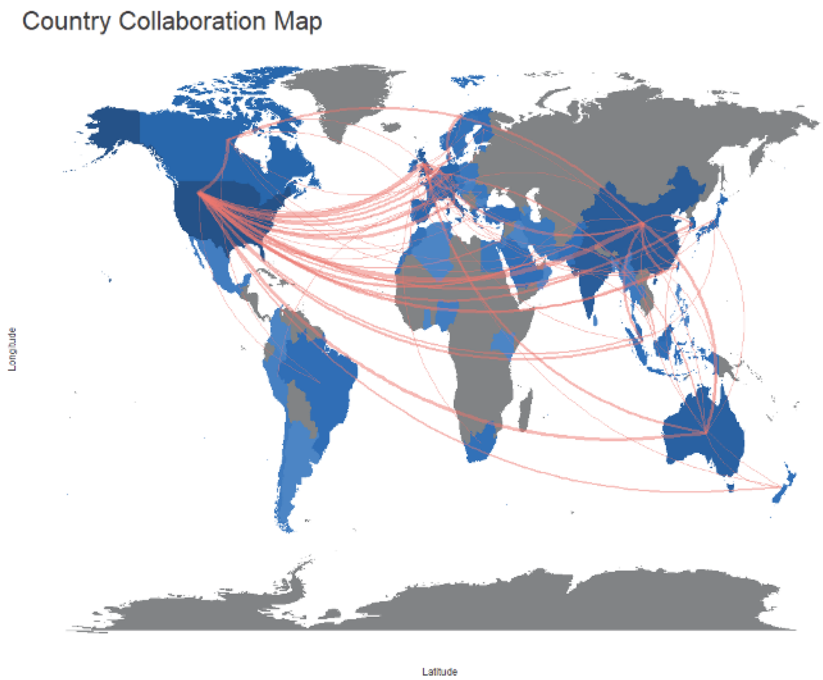

Fig. 1. Word Map collaboration (Social Structure)

As we can observe from the results of Fig. 1, there were strong collaboration from the researchers in United States from one side and other countries as shown in below:

Table 3
Country collaboration Table

From	To	Frequency
USA	UNITED KINGDOM	39
	TAIWAN	7
	SINGAPORE	8
	SAUDI ARABIA	5
	PAKISTAN	6
	NEW ZEALAND	5
	NETHERLANDS	10
	KOREA	8
	ITALY	15
	INDIA	8
	HONG KONG	8
	GERMANY	13
	FRANCE	9
	DENMARK	5
	CHINA	43
	CANADA	18
	AUSTRALIA	21
UNITED KINGDOM	SWITZERLAND	5
	NETHERLANDS	7
	GERMANY	6
	CHINA	9
	AUSTRALIA	9
SPAIN	UNITED KINGDOM	5
NORWAY	DENMARK	14
ITALY	UNITED KINGDOM	7
CHINA	TAIWAN	5
	SINGAPORE	8
	HONG KONG	12
	CANADA	8
AUSTRALIA	GERMANY	7
	CHINA	14

4. Country Scientific Production

One of the interesting areas of the interest is to learn more about the contribution of different countries in big data in social media. As we can observe from the results of Fig. 2, researchers from USA (1289 papers), China (383 papers), India (305 papers), UK (254 papers) and Australia (175 papers) have contributed the most on big data in social media.

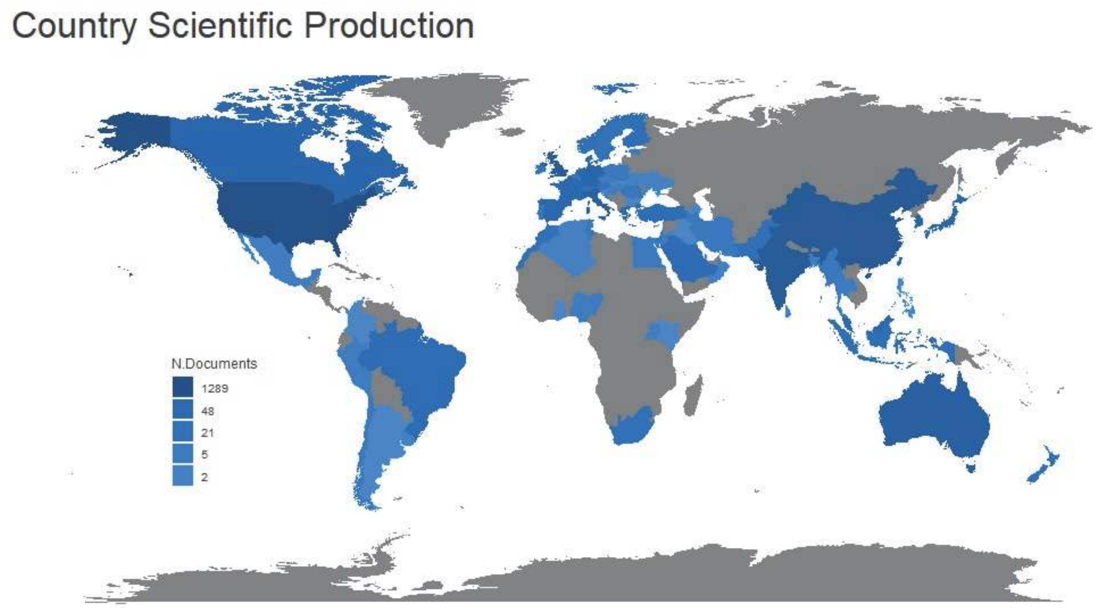

Fig. 2. The frequency of the keywords used in different big data in social media studies

5. Highly cited papers (Most Global Cited Documents)

Table 4 shows a summary of the most cited articles. As we can observe from the results of Table 4, the study by Boyd et al. (2012) has received the highest citations. The second highly cited work is associated with Lazer et al. (2014) where they investigated a trap in big data. The third highly cited work belongs to Kramer et al. (2014) where they proposed an important and emerging area of social science research that needs to be approached with sensitivity and with vigilance regarding personal privacy issues. According to Stephens et al. (2015), Genomics is a Big Data science and will become much bigger as time passes on, but we still do not know whether the requirements of genomics will surpass other Big Data domains. Morone and Makse (2015) stated that big data analyses are associated with the set of optimal influencers is much smaller than the one forecasted by previous heuristic centralities. According to Bello-Orgaz et al. (2016) big data plays an essential role for a large number of research areas such as data mining, machine learning, computational intelligence, information fusion, the semantic Web, and social networks. The rise of various big data structures such as Apache Hadoop and, more recently, Spark, for huge data processing has provided an opportunity for an efficient utilization of data mining techniques and machine learning methods in various domains. Bello-Orgaz et al. (2016) provided a revision of the new techniques designed to help for active data mining and information fusion from social media and of the new applications and frameworks which are presently are available under the “umbrella” of the social networks, social media and big data paradigms. Mohr et al. (2013) concentrated on the barriers and the costs associated with big data storage and specified that any improvements in the collection, storage, analysis and visualization of big data could help practitioner better target sales.

Table 4
The summary of the most cited articles

Paper , Year , Source	Total Citations	TC per Year
BOYD D, 2012, INF COMMUN SOC	1439	205.571
LAZER D, 2014, SCIENCE	739	147.8
KRAMER ADI, 2014, PROC NATL ACAD SCI U S A	731	146.2
STEPHENS ZD, 2015, PLOS BIOL	295	73.75
MORONE F, 2015, NATURE	272	68
BELLO-ORGAZ G, 2016, INF FUSION	212	70.667
MOHR DC, 2013, GEN HOSP PSYCHIATRY	190	31.667
YOUYOU W, 2015, PROC NATL ACAD SCI U S A	176	44
VAN DIJCK J, 2014, SURVEILL SOC	171	34.2
TUFEKCI Z, 2014, PROC INT CONF WEBLOGS SOC MEDIA, ICWSM	152	30.4
WOOD SA, 2013, SCI REP	151	25.167
CRAMPTON JW, 2013, CARTOGR GEOGR INF SCI	150	25
XIANG Z, 2015, INT J HOSP MANAGE	133	33.25
RUSSELL NEUMAN W, 2014, J COMMUN	130	26
MOCANU D, 2013, PLOS ONE	126	21
KHOURY MJ, 2014, SCIENCE	124	24.8
EICHSTAEDT JC, 2015, PSYCHOL SCI	122	30.5
KRAWCZYK B, 2016, PROG ARTIF INTELL	109	36.333
CHAE B, 2015, INT J PROD ECON	109	27.25
BRUNS A, 2013, AM BEHAV SCI	103	17.167
PARK G, 2015, J PERS SOC PSYCHOL	99	24.75
HERLAND M, 2014, J BIG DATA	95	19
LEEFLANG PSH, 2014, EUR MANAGE J	94	18.8
ZHENG Y, 2014, UBICOMP - PROC ACM INT JT CONF PERVASIVE UBIQUITOUS COMPUT	91	18.2
PROCTER R, 2013, INT J SOC RES METHODOL	90	15
BIAN J, 2012, INT CONF INF KNOWLEDGE MANAGE	88	12.571
HAY SI, 2013, PLOS MED	87	14.5
LIU C, 2014, IEEE TRANS PARALLEL DISTRIB SYST	82	16.4
GOLDER SA, 2014, ANNU REV SOCIOL	78	15.6
SHELTON T, 2015, LANDSC URBAN PLANN	73	18.25
TUFEKCI Z, 2014, FIRST MONDAY	73	14.6
SHELTON T, 2014, GEOFORUM	72	14.4
LAM W, 2012, PROC VLDB ENDOW	72	10.286
WATSON HJ, 2014, COMMUN ASSOC INFO SYST	71	14.2
BRAVO-MARQUEZ F, 2014, KNOWL BASED SYST	71	14.2
HASAN S, 2014, TRANSP RES PART C EMERG TECHNOL	67	13.4
BAIL CA, 2014, THEORY SOC	67	13.4
BURNAP P, 2015, POLICY INTERNET	65	16.25
DRISCOLL K, 2014, INT J COMMUN	64	12.8
O'DEA B, 2015, INTERNET INTERV	63	15.75
KENNEY M, 2016, ISSUES SCI TECHNOL	62	20.667

MARINE-ROIG E, 2015, J DESTIN MARK MANAGE	62	15.5
SINGH S, 2012, PROC - INT CONF COMMUN , INF COMPUT TECHNOL , ICCICT	62	8.857
CHIANG RHL, 2012, ACM TRANS MANAGE INF SYST	62	8.857
YOUNG SD, 2014, PREV MED	61	12.2
STIEGLITZ S, 2014, BUSIN INFO SYS ENG	61	12.2
HE W, 2015, INF MANAGE	60	15
WHITTINGTON R, 2014, J STRATEGIC INFORM SYST	60	12
VATSAVAI RR, 2012, PROC ACM SIGSPATIAL INT WORKSHOP ANAL BIG GEOSPATIAL DATA, BIGSPATIAL	58	8.286
COMPTON R, 2015, PROC - IEEE INT CONF BIG DATA, IEEE BIG DATA	57	14.25
YANG M, 2015, J BIOMED INFORMATICS	57	14.25
BLISS CA, 2012, J COMPUT SCI	56	8
RAM S, 2015, IEEE J BIOMEDICAL HEALTH INFORMAT	55	13.75
SMITH M, 2012, IEEE INT CONF DIGIT ECOSYST TECHNOL	55	7.857
BAKER TB, 2014, J MED INTERNET RES	54	10.8
LIU X, 2013, LECT NOTES COMPUT SCI	54	9
YAQOOB I, 2016, INT J INF MANAGE	53	17.667
ISHWARAPPA I, 2015, PROCEDIA COMPUT SCI	51	12.75
MARIANI MM, 2016, TOUR MANAGE	50	16.667
BUHALIS D, 2015, J DESTIN MARK MANAGE	49	12.25
ARAGÕN P, 2013, POLICY INTERNET	49	8.167
PROCTER R, 2013, POLICING SOC	48	8
HAUSTEIN S, 2016, SCIENTOMETRICS	46	15.333
XIE H, 2014, NEURAL NETW	46	9.2
MORONE F, 2016, SCI REP	44	14.667
PAPACHARISSI Z, 2016, INF COMMUN SOC	43	14.333
HANSEN MM, 2014, YEARB MED INFORM	43	8.6
WHITE M, 2012, BUS INF REV	43	6.143
BAYM NK, 2013, FIRST MONDAY	42	7
ZHONG E, 2012, PROC ACM SIGKDD INT CONF KNOWL DISCOV DATA MIN	41	5.857
BENTLEY RA, 2014, BEHAV BRAIN SCI	40	8
OU M, 2013, PROC ACM SIGKDD INT CONF KNOWL DISCOV DATA MIN	40	6.667
WU KJ, 2017, J CLEAN PROD	39	19.5
YANG W, 2015, PROC NATL ACAD SCI U S A	39	9.75
COUPER MP, 2013, SURV RES METHODS	39	6.5
LOHRMANN B, 2015, PROC INT CONF DISTRIB COMPUT SYST	38	9.5
ARTIKIS A, 2012, PROC ACM INT CONF DISTRIB EVENT-BASED SYST , DEBS	38	5.429
BAIL C, 2014, TERRIFIED: HOW ANTI-MUSLIM FRINGE ORGAN BE-CAME MAINSTREAM	37	7.4
CAO G, 2015, COMPUT ENVIRON URBAN SYST	36	9
HU H, 2015, IEEE NETWORK	35	8.75
BURNS R, 2015, GEOJOURNAL	35	8.75
JIANG B, 2015, PROF GEOGR	35	8.75

DE FRANCISCI MORALES G, 2013, WWW COMPANION - PROC INT CONF WORLD WIDE WEB	35	5.833
JIANG W, 2015, PLOS ONE	34	8.5
FERNÁNDEZ-LUQUE L, 2015, HEALTHC INFORMATICS RES	34	8.5
ALI A, 2015, INT J ADV SOFT COMPUT APPL	34	8.5
FLEURENCE RL, 2014, HEALTH AFF	33	6.6
HU H, 2014, IEEE MULTIMEDIA	33	6.6
MARTIN-SANCHEZ F, 2014, YEARB MED INFORM	32	6.4
BRUNS A, 2013, FIRST MONDAY	32	5.333
CIULLA F, 2012, EPJ DATA SCI	32	4.571
HUDA M, 2018, INT J EMERG TECHNOL LEARN	31	31
LIU SQ, 2017, INT J HOSP MANAGE	31	15.5
WILLIAMS ML, 2016, BR J CRIMINOL	31	10.333
TSOU MH, 2015, CARTOGR GEOGR INF SCI	31	7.75
BEAM AL, 2018, JAMA	30	30
JIANG B, 2015, CITIES	30	7.5
PALDINO S, 2015, EPJ DATA SCI	30	7.5
ROSS MK, 2014, YEARB MED INFORM	30	6
KEPNER J, 2013, IEEE HIGH PERFORM EXTREME COMPUT CONF , HPEC	30	5
OBOLER A, 2012, FIRST MONDAY	30	4.286
MIAH SJ, 2017, INF MANAGE	29	14.5
CONWAY M, 2017, STUD CONFL TERRORISM	29	14.5
EDITORIAL DEPARTMENT OF CHINA JOURNAL OF HIGHWAY EDCJH, 2016, ZONGGUO GONGLU XUEBAO	29	9.667
DE MAIO C, 2016, INF FUSION	29	9.667
LIMA ACES, 2015, APPL MATH COMPUT	28	7
GITTELMAN S, 2015, J MED INTERNET RES	28	7
JIANG K, 2013, LECT NOTES COMPUT SCI	28	4.667
MILLER HJ, 2013, J TRANSP GEOGR	28	4.667
STIEGLITZ S, 2018, INT J INF MANAGE	27	27
ORDENES FV, 2017, J CONSUM RES	27	13.5
LEVIN N, 2015, ECOL APPL	27	6.75
FRIED D, 2015, PROC - IEEE INT CONF BIG DATA, IEEE BIG DATA	27	6.75
SHARMA S, 2014, DATA SCI J	27	5.4
HUSSAIN A, 2014, LECT NOTES COMPUT SCI	27	5.4
DEDE E, 2013, IEEE INT CONF CLOUD COMPUT , CLOUD	27	4.5
WILLIAMS ML, 2017, BR J CRIMINOL	26	13
KHARE R, 2016, BRIEF BIOINFORM	26	8.667
LEV-ON A, 2015, GOV INF Q	26	6.5
PEEK N, 2014, YEARB MED INFORM	26	5.2
KIM HS, 2015, J COMMUN	25	6.25
FULGONI G, 2014, J ADVERT RES	25	5
HUANG Y, 2016, COMPUT ENVIRON URBAN SYST	24	8
CULOTTA A, 2016, MARK SCI	24	8
DEHGHANI M, 2016, J EXP PSYCHOL GEN	24	8

STEPHANSEN HC, 2014, INF COMMUN SOC	24	4.8
SIKOS LF, 2015, MASTERING STRUCTURED DATA ON THE SEMANTIC WEB: FROM HTML5 MICRODATA TO LINKED OPEN DATA	23	5.75
JIANG B, 2015, GEOJOURNAL	23	5.75
YOUNG SD, 2015, PREV MED	23	5.75
KAFEZA E, 2014, PROC - IEEE INT CONGR BIG DATA, BIGDATA CONGR	23	4.6
SANG ETK, 2013, COMPUT LINGUIST NETHERLANDS J	23	3.833
SINGH VK, 2012, MM - PROC ACM INT CONF MULTIMEDIA	23	3.286
PARK SB, 2016, J TRAVEL TOUR MARK	22	7.333
WILSON MW, 2015, CULT GEOGR	22	5.5
KEPNER J, 2014, IEEE HIGH PERFORM EXTREM COMPUT CONF , HPEC	22	4.4
RIBARSKY W, 2014, COMPUT GRAPHICS (PERGAMON)	22	4.4
CAI Y, 2014, NEURAL NETW	22	4.4
MCKELVEY K, 2014, INF COMMUN SOC	22	4.4
CAI J, 2017, REMOTE SENS ENVIRON	21	10.5
CARLEY KM, 2016, SAF SCI	21	7
ULDAM J, 2016, NEW MEDIA AND SOCIETY	21	7
ZHU W, 2015, IEEE MULTIMEDIA	21	5.25
IMMONEN A, 2015, IEEE ACCESS	21	5.25
WOOD D, 2014, FRONT NEUROINFORMATICS	21	4.2
BAKILLAH M, 2014, BIG DATA: TECHNIQUES AND TECHNOLOGIES IN GEOINFORMATICS	21	4.2
SLAVAKIS K, 2014, IEEE SIGNAL PROCESS MAG	21	4.2
KERN ML, 2014, DEV PSYCHOL	21	4.2
BANSAL S, 2016, J INFECT DIS	20	6.667
SHARMA S, 2016, FUTURE GENER COMPUT SYST	20	6.667
KWOK L, 2016, INT J CONTEMP HOSP MANAGE	20	6.667
BAGHERI H, 2015, INT J ELECTR COMPUT ENG	20	5

6. The most common keywords

Table 5 demonstrates some of the mostly cited references associated with big data in social media. As we can observe from the results of Table 5, big data, social media and social networking (online) are three well recognized keywords used in the literature. Fig. 3 shows the most important words used over times.

Table 5
The most popular keywords used in studies associated with big data in social media

Words	Occurrences	Words2	Occurrences3
big data	1139	data privacy	43
social media	836	marketing	43
social networking (online)	811	big data analytics	42
data mining	445	social media analysis	41
human	180	data analytics	40
internet	157	disasters	40
sentiment analysis	152	information retrieval	40
data handling	145	male	40
artificial intelligence	142	female	39
learning systems	133	procedures	39
twitter	132	data analysis	38

humans	121	facebook	38
social media datum	115	sales	38
decision making	114	social sciences computing	38
natural language processing systems	106	data set	36
digital storage	91	internet of things	36
semantics	91	location	36
information management	89	population statistics	36
article	80	text mining	36
classification (of information)	80	clustering algorithms	34
behavioral research	79	surveys	34
cloud computing	71	information systems	33
forecasting	69	on-line social networks	32
priority journal	65	health	31
commerce	60	risk assessment	31
united states	57	database systems	30
hadoop	55	decision support systems	30
data visualization	54	machine learning	30
visualization	54	unstructured data	30
social media platforms	52	websites	30
big datum	50	world wide web	30
distributed computer systems	49	computational linguistics	29
natural language processing	49	medical informatics	29
social network	49	neural networks	29
map-reduce	48	online social medias	29
public health	47	search engines	29
social media analytics	47	china	28
learning algorithms	46	linguistics	28
algorithms	45	privacy	28
information processing	45	data processing	27
text processing	45	deep learning	27
health care	44	online systems	27
information analysis	44	statistics	27
information dissemination	44	geographic information systems	26

Fig. 3. The frequency of the keywords used in different big data in social media

7. Word Dynamics

Word dynamic graph prepared on keywords helps us learn more about the keyword dynamics over time. Their growing or declining trend can help us choose a better topic in any survey. There are two types of keywords: Author keywords and Keywords plus. Author keywords are the ones that authors state in their articles and keyword plus are the results of the Thomson Reuters editorial expertise in science. What they do is to review the titles of all references and highlight additional relevant but overlooked keywords that were not listed by the authors or publishers. With keywords plus, it is possible to uncover more papers that may not have appeared in a search due to changes in scientific keywords over time.

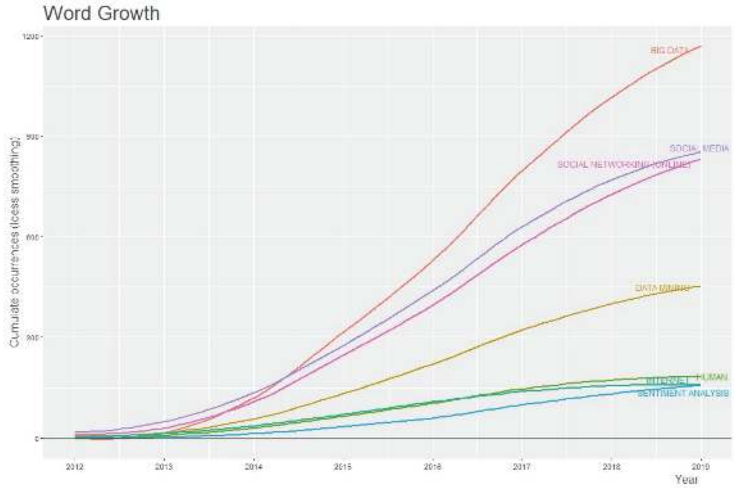

Fig. 4. Keywords plus dynamic view over time

As we can observe from the results of Fig. 4, big data, social media, social network (online) and data mining, show good growth in the chart unlike sentiment analysis and internet.

8. Conceptual structure, Co-occurrence network

A keywords co-occurrence network (KCN) focuses on understanding the knowledge components and knowledge structure of a scientific/technical field by examining the links between keywords in the literature. Fig. 5 focuses on the analysis methods based on KCNs, which have been used in theoretical and empirical studies to explore research topics and their relationships in selecting scientific fields. If keywords are grouped into the same cluster, they are more likely to reflect identical topics. Each cluster has different number of subject keyword.

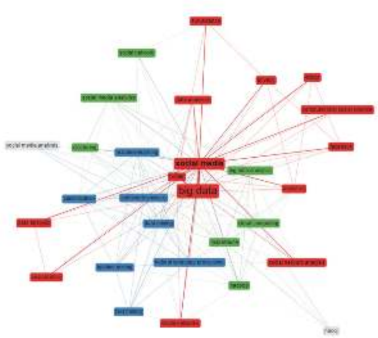

Fig. 5. Co-occurrence network (2012-2019)

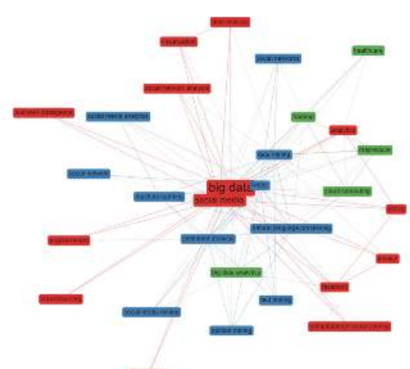

Fig. 6. Co-occurrence network (2012-2016)

To see the growth and the evolution of this network more tangibly, Fig. 6 shows the same graph over the period 2012-2016 (beginning of the survey until the first significant growth of articles production).

9. Thematic Map (Well developed or not? Important or not?)

When co-word analysis is used for mapping science, clusters of keywords and their interconnections are obtained. These clusters are considered as themes. Each research theme obtained in this process is characterized by two parameters; namely “density” and “centrality”. Both median and mean values for density and centrality can be used in classifying themes in to our groups. In a theme, the keywords and their interconnections draw a network graph, called a "thematic network" that "centrality" is horizontal axis and "density" is vertical axis in it. In a network, if the node has a large amount of relations with others, it has a higher centrality and lies in an essential position in the network. Centrality is therefore used to measure the correction degree among different topics. Similarly, a higher density means higher cohesiveness or equals the higher internal correlation degree among nodes. The density of a research field represents its capability to maintain and develop itself. Thematic map is a very intuitive plot and we can analyze themes according to the quadrant in which they are placed. Upper-right quadrant is motor-themes, lower-right quadrant is basic themes, lower-left quadrant is emerging or disappearing themes, upper-left quadrant is very specialized/niche themes. Themes in the upper-right quadrant are both well developed and important for the structuring of a research field such as "big data" and "big data analytics". Themes in the upper-left quadrant have well developed internal ties but unimportant external ties and so are of only marginal importance for the field such as "social network". Themes in the lower-left quadrant are both "weakly developed and marginal", mainly representing either emerging or disappearing themes such as "social media" and "Hadoop". Themes in the lower-right quadrant are "important for a research field but are not developed", so this quadrant groups transversal and general, basic themes such as "twitter". Thematic analysis shows that for better results we can merge our research focus with "big data analytics" and "twitter" that are important topics in this field but not developed well.

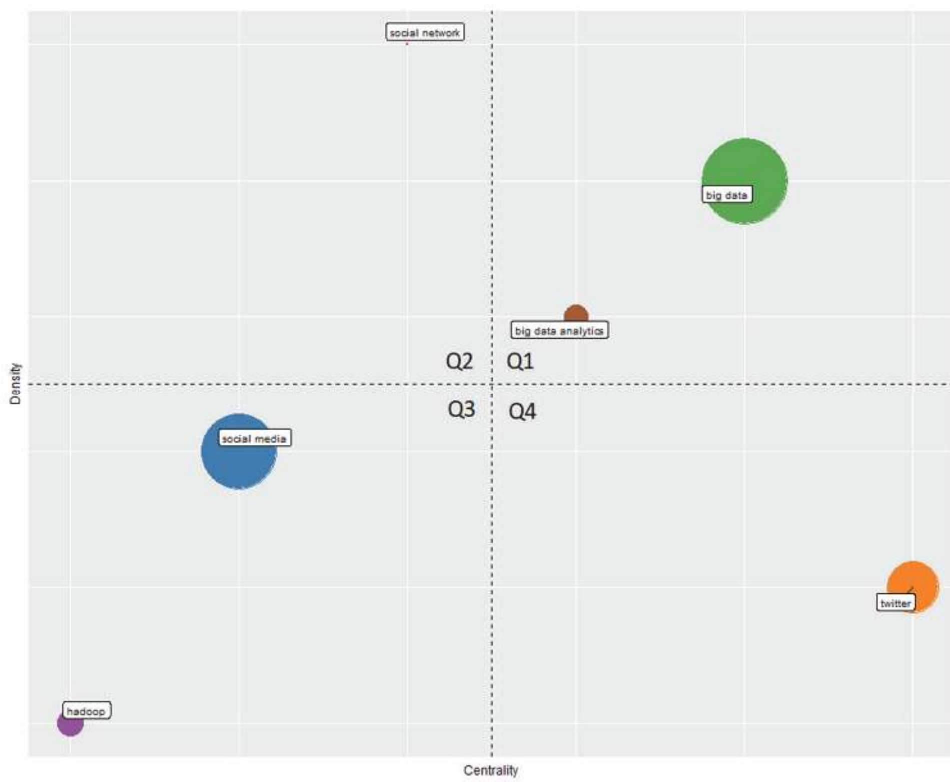

Fig. 7. Thematic Map

10. Intellectual Structure, Historiograph

The historiographic map is a graph proposed by Garfield to represent a chronological network map of the most relevant direct citations resulting from a bibliographic collection. The citation network technique provides the scholar with a new modus operandi which may significantly affect future historiography.

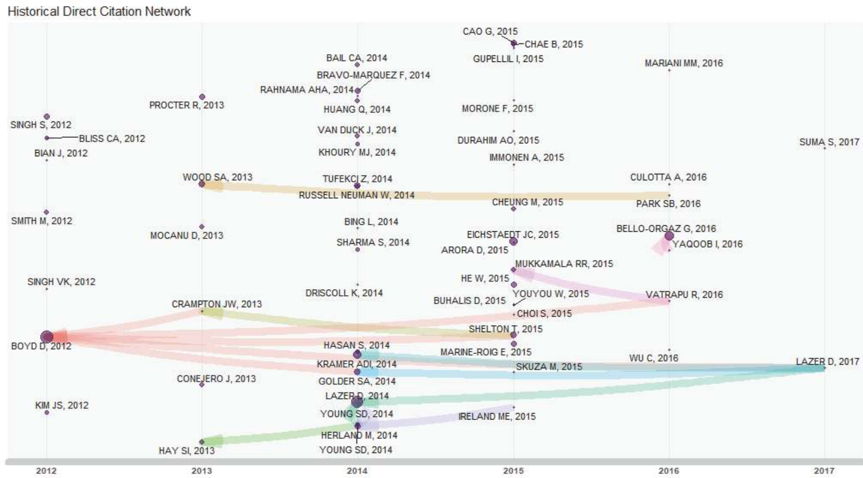

Fig. 8. Historiograph

Fig. 8 shows Boyd (2012), Wood (2013), Hay (2013) and Crampton (2013) were the beginner of new trends at their own time. The direction of the arrows in Fig. 8 explains the chronicle change of research trends from the past. Research accomplished by Boyd (2012) was about the effects of big data on knowledge. Crampton (2013), Kramer (2014), Hassan (2014), Shelton (2015) and Vatrapu (2016) provided more development on big data. Wood (2013) tried to understand which elements of nature influence more on people to locations around the globe, and whether changes in ecosystems could alter visitation rates. Hay (2013), in his research used big data approaches to routinely map all of vast majority of infectious diseases of clinical significance. It would be of public health benefit to map about half of conditions. Research of Crampton (2013) presented an overview and initial results of a geoweb analysis designed to provide the foundation for a continued discussion of the potential impacts of 'big data' for the practice of critical human geography. They believed while Haklay's (2012) observation that social media content is generated by a small number of 'outliers' is correct. They could explore alternative methods and conceptual frameworks that might allow for one to overcome the limitations of previous analyses of user-generated geographic information.

11. Conclusion

This study has tried to provide a comprehensive review of the studies published in the literature associated with big data in social media. The study has indicated that this field has been popular mostly among researchers in USA, China, India, UK and Australia. The study has also indicated that while researchers from USA and UK published a relatively high number of papers, they were also successful to publish highly cited papers. Many big data in social media studies have dealt with combinatorial optimization techniques and our survey has concluded that meta-heuristics methods have been popular among researchers to locate the near-optimal solutions. We hope this study could guide other researchers find important research gaps.

References

Ali, A., Shamsuddin, S. M., & Ralescu, A. L. (2015). Classification with class imbalance problem: a review. *International Journal of Advances in Soft Computing and its Applications*, *7*(3), 176-204.

Aragón, P., Kappler, K. E., Kaltenbrunner, A., Laniado, D., & Volkovich, Y. (2013). Communication dynamics in twitter during political campaigns: The case of the 2011 Spanish national election. *Policy & Internet*, *5*(2), 183-206.

Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, *11*, 959-975.

Artikis, A., Etzion, O., Feldman, Z., & Fournier, F. (2012, July). Event processing under uncertainty. In *Proceedings of the 6th ACM International Conference on Distributed Event-Based Systems* (pp. 32-43). ACM.

Bagheri, H., & Shaltooki, A. A. (2015). Big Data: challenges, opportunities and Cloud based solutions. International Journal of Electrical and Computer Engineering (IJECE), 5(2), 340-343.

Bail, C. A. (2014). The cultural environment: Measuring culture with big data. *Theory and Society*, *43*(3-4), 465-482.

Bail, C. A. (2014). Terrified: How anti-Muslim fringe organizations became mainstream. *Princeton University Press*.

Baker, T. B., Gustafson, D. H., & Shah, D. (2014). How can research keep up with eHealth? Ten strategies for increasing the timeliness and usefulness of eHealth research. *Journal of Medical Internet Research*, *16*(2).

Bakillah, M., Lauer, J., Liang, S. H., Zipf, A., Jokar Arsanjani, J., Mobasheri, A., & Loos, L. (2014). Exploiting big VGI to improve routing and navigation services. *Big data techniques and technologies in geoinformatics*, 177-192.

Bansal, S., Chowell, G., Simonsen, L., Vespignani, A., & Viboud, C. (2016). Big data for infectious disease surveillance and modeling. *The Journal of Infectious Diseases*, *214*(suppl_4), S375-S379.

Baym, N. K. (2013). Data not seen: The uses and shortcomings of social media metrics. *First Monday*, *18*(10).

Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. *Journal of the American Medical Association*, *319*(13), 1317-1318.

Bello-Orgaz, G., Jung, J., & Camacho, D. (2016). Social big data: Recent achievements and new challenges. *Information Fusion*, *28*, 45--59.

Bentley, R. A., O'Brien, M. J., & Brock, W. A. (2014). Mapping collective behavior in the big-data era. *Behavioral and Brain Sciences*, *37*(1), 63.

Bian, J., Topaloglu, U., & Yu, F. (2012, October). Towards large-scale twitter mining for drug-related adverse events. In *Proceedings of the 2012 international workshop on Smart health and wellbeing* (pp. 25-32). ACM.

Bliss, C. A., Kloumann, I. M., Harris, K. D., Danforth, C. M., & Dodds, P. S. (2012). Twitter reciprocal reply networks exhibit assortativity with respect to happiness. *Journal of Computational Science*, 3(5), 388-397.

Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society*, *15*(5), 662-679.

Bravo-Marquez, F., Mendoza, M., & Poblete, B. (2014). Meta-level sentiment models for big social data analysis. *Knowledge-Based Systems*, *69*, 86-99.

Bruns, A. (2013). Faster than the speed of print: Reconciling 'big data' social media analysis and academic scholarship. *First Monday*, *18*(10).

Bruns, A., Highfield, T., & Burgess, J. (2013). The Arab Spring and social media audiences: English and Arabic Twitter users and their networks. *American Behavioral Scientist*, *57*(7), 871-898.

Buhalis, D., & Foerste, M. (2015). SoCoMo marketing for travel and tourism: Empowering co-creation of value. *Journal of Destination Marketing & Management*, *4*(3), 151-161.

Burnap, P., & Williams, M. L. (2015). Cyber hate speech on twitter: An application of machine classification and statistical modeling for policy and decision making. *Policy & Internet*, *7*(2), 223-242.

Burns, R. (2015). Rethinking big data in digital humanitarianism: Practices, epistemologies, and social relations. *GeoJournal*, *80*(4), 477-490.

Cai, J., Huang, B., & Song, Y. (2017). Using multi-source geospatial big data to identify the structure of polycentric cities. *Remote Sensing of Environment*, *202*, 210-221.

Cai, Y., Li, Q., Xie, H., & Min, H. (2014). Exploring personalized searches using tag-based user profiles and resource profiles in folksonomy. *Neural Networks*, *58*, 98-110.

Carley, K. M., Malik, M., Landwehr, P. M., Pfeffer, J., & Kowalchuck, M. (2016). Crowd sourcing disaster management: The complex nature of Twitter usage in Padang Indonesia. *Safety Science*, *90*, 48-61.

Cao, G., Wang, S., Hwang, M., Padmanabhan, A., Zhang, Z., & Soltani, K. (2015). A scalable framework for spatiotemporal analysis of location-based social media data. *Computers, Environment and Urban Systems*, *51*, 70-82.

Chae, B. K. (2015). Insights from hashtag# supplychain and Twitter Analytics: Considering Twitter and Twitter data for supply chain practice and research. *International Journal of Production Economics*, *165*, 247-259.

Chiang, R. H., Goes, P., & Stohr, E. A. (2012). Business intelligence and analytics education, and program development: A unique opportunity for the information systems discipline. *ACM Transactions on Management Information Systems* (TMIS), *3*(3), 12.

Ciulla, F., Mocanu, D., Baronchelli, A., Gonçalves, B., Perra, N., & Vespignani, A. (2012). Beating the news using social media: the case study of American Idol. *EPJ Data Science*, *1*(1), 8.

Compton, R., Jurgens, D., & Allen, D. (2014, October). Geotagging one hundred million twitter accounts with total variation minimization. In *Big Data (Big Data), 2014 IEEE International Conference* on (pp. 393-401). IEEE.

Conway, M. (2017). Determining the role of the internet in violent extremism and terrorism: Six suggestions for progressing research. *Studies in Conflict & Terrorism*, I(1), 77-98.

Couper, M. P. (2013, December). Is the sky falling? New technology, changing media, and the future of surveys. In *Survey Research Methods, 7*(3), 145-156.

Crampton, J., Graham, M., Poorthuis, A., Shelton, T., Stephens, M., Wilson, M., & Zook, M. (2013). Beyond the geotag: situating 'big data'and leveraging the potential of the geoweb. *Cartography and Geographic Information Science*, *40*(2), 130--139.

Culotta, A., & Cutler, J. (2016). Mining brand perceptions from twitter social networks. *Marketing Science*, *35*(3), 343-362.

Dede, E., Sendir, B., Kuzlu, P., Hartog, J., & Govindaraju, M. (2013, June). An evaluation of cassandra for hadoop. In *Cloud Computing (CLOUD), 2013 IEEE Sixth International* Conference on (pp. 494-501). IEEE.

De Francisci Morales, G. (2013, May). SAMOA: A platform for mining big data streams. In *Proceedings of the 22nd International Conference on World Wide Web* (pp. 777-778). ACM.

Dehghani, M., Johnson, K., Hoover, J., Sagi, E., Garten, J., Parmar, N. J. & Graham, J. (2016). Purity homophily in social networks. *Journal of Experimental Psychology: General*, *145*(3), 366.

De Maio, C., Fenza, G., Loia, V., & Parente, M. (2016). Time aware knowledge extraction for microblog summarization on twitter. *Information Fusion*, *28*, 60-74.

Driscoll, K., & Walker, S. (2014). Big data, big questions| working within a black box: Transparency in the collection and production of big twitter data. *International Journal of Communication*, *8*, 20.

Eichstaedt, J. C., Schwartz, H. A., Kern, M. L., Park, G., Labarthe, D. R., Merchant, R. M. & Weeg, C. (2015). Psychological language on Twitter predicts county-level heart disease mortality. *Psychological science*, 26(2), 159-169.

Fernández-Luque, L., & Bau, T. (2015). Health and social media: perfect storm of information. *Healthcare Informatics Research*, *21*(2), 67-73.

Fleurence, R. L., Beal, A. C., Sheridan, S. E., Johnson, L. B., & Selby, J. V. (2014). *Patient-powered research networks aim to improve patient care and health research. Health Affairs*, *33*(7), 1212-1219.

Fried, D., Surdeanu, M., Kobourov, S., Hingle, M., & Bell, D. (2014, October). Analyzing the language of food on social media. In *Big Data (Big Data), 2014 IEEE International* Conference on (pp. 778-783). IEEE.

Fulgoni, G., & Lipsman, A. (2014). Digital game changers: how social media will help usher in the era of mobile and multi-platform campaign-effectiveness measurement. *Journal of Advertising Research*, *54*(1), 11-16.

Gittelman, S., Lange, V., Crawford, C. A. G., Okoro, C. A., Lieb, E., Dhingra, S. S., & Trimarchi, E. (2015). A new source of data for public health surveillance: Facebook likes. *Journal of medical Internet research*, *17*(4).

Golder, S. A., & Macy, M. W. (2014). Digital footprints: Opportunities and challenges for online social research. *Annual Review of Sociology*, *40*, 129-152.

Hansen, M. M., Miron-Shatz, T., Lau, A. Y. S., & Paton, C. (2014). Big data in science and healthcare: a review of recent literature and perspectives. *Yearbook of Medical Informatics*, *23*(01), 21-26.

Hasan, S., & Ukkusuri, S. V. (2014). Urban activity pattern classification using topic models from online geo-location data. *Transportation Research Part C: Emerging Technologies*, *44*, 363-381.

Haustein, S. (2016). Grand challenges in altmetrics: heterogeneity, data quality and dependencies. *Scientometrics*, *108*(1), 413-423.

Hay, S. I., George, D. B., Moyes, C. L., & Brownstein, J. S. (2013). Big data opportunities for global infectious disease surveillance. *PLoS Medicine*, *10*(4), e1001413.

He, W., Wu, H., Yan, G., Akula, V., & Shen, J. (2015). A novel social media competitive analytics framework with sentiment benchmarks. *Information & Management*, *52*(7), 801-812.

Herland, M., Khoshgoftaar, T. M., & Wald, R. (2014). A review of data mining using big data in health informatics. *Journal of Big data*, 1(1), 2.

Huang, Y., Guo, D., Kasakoff, A., & Grieve, J. (2016). Understanding US regional linguistic variation with Twitter data analysis. *Computers, Environment and Urban Systems*, *59*, 244-255.

Hu, H., Wen, Y., Luan, H., Chua, T. S., & Li, X. (2014). Toward multiscreen social TV with geolocation-aware social sense. *IEEE MultiMedia*, *21*(3), 10-19.

Hu, H., Wen, Y., Gao, Y., Chua, T. S., & Li, X. (2015). *Towards SDN-Enabled Big Data Platform for Social TV Analytics.*

Huda, M., Maseleno, A., Atmotiyoso, P., Siregar, M., Ahmad, R., Jasmi, K., & Muhamad, N. (2018). Big data emerging technology: insights into innovative environment for online learning resources. *International Journal of Emerging Technologies in Learning* (iJET), *13*(1), 23-36.

Hussain, A., & Vatrapu, R. (2014, May). Social data analytics tool (sodato). In *International Conference on Design Science Research in Information Systems* (pp. 368-372). Springer, Cham.

Immonen, A., Pääkkönen, P., & Ovaska, E. (2015). Evaluating the quality of social media data in big data architecture. *IEEE Access*, *3*, 2028-2043.

Jiang, B. (2015). Geospatial analysis requires a different way of thinking: The problem of spatial heterogeneity. *GeoJournal*, *80*(1), 1-13.

Jiang, B. (2016). Head/tail breaks for visualization of city structure and dynamics. *European Handbook of Crowdsourced Geographic Information*, 169.

Jiang, B., & Miao, Y. (2015). The evolution of natural cities from the perspective of location-based social media. *The Professional Geographer*, *67*(2), 295-306.

Jiang, K., & Zheng, Y. (2013, December). Mining twitter data for potential drug effects. In *International Conference on Advanced Data Mining and Applications* (pp. 434-443). Springer, Berlin, Heidelberg.

Jiang, W., Wang, Y., Tsou, M. H., & Fu, X. (2015). Using social media to detect outdoor air pollution and monitor air quality index (AQI): a geo-targeted spatiotemporal analysis framework with Sina Weibo (Chinese Twitter). *PloS one*, *10*(10), e0141185.

Kafeza, E., Kanavos, A., Makris, C., & Vikatos, P. (2014, June). T-PICE: Twitter personality based influential communities' extraction system. In *Big Data (BigData Congress), 2014 IEEE International Congress* on (pp. 212-219). IEEE.

Kenney, M., & Zysman, J. (2016). The rise of the platform economy. *Issues in Science and Technology*, *32*(3), 61.

Kepner, J., Anderson, C., Arcand, W., Bestor, D., Bergeron, B., Byun, C. & Prout, A. (2013, September). D4M 2.0 schema: A general purpose high performance schema for the Accumulo database. In *2013 IEEE High Performance Extreme Computing Conference (HPEC*) (pp. 1-6). IEEE.

Kepner, J., Gadepally, V., Michaleas, P., Schear, N., Varia, M., Yerukhimovich, A., & Cunningham, R. K. (2014, September). Computing on masked data: a high performance method for improving big data veracity. In *High Performance Extreme Computing Conference (HPEC)*, 2014 IEEE (pp. 1-6). IEEE.

Kern, M. L., Eichstaedt, J. C., Schwartz, H. A., Park, G., Ungar, L. H., Stillwell, D. J. & Seligman, M. E. (2014). From "Sooo excited!!!" to "So proud": Using language to study development. *Developmental psychology*, *50*(1), 178.

Khare, R., Good, B. M., Leaman, R., Su, A. I., & Lu, Z. (2015). Crowdsourcing in biomedicine: challenges and opportunities. *Briefings in Bioinformatics*, *17*(1), 23-32.

Khoury, M., & Ioannidis, J. (2014). Big data meets public health. *Science*, 1054--1055.

Kim, H. S. (2015). Attracting views and going viral: How message features and news-sharing channels affect health news diffusion. *Journal of Communication*, *65*(3), 512-534.

Kramer, A., Guillory, J., & Hancock, J. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 201320040.

Krawczyk, B. (2016). Learning from imbalanced data: open challenges and future directions. *Progress in Artificial Intelligence*, *5*(4), 221-232.

Kwok, L., & Xie, K. L. (2016). Factors contributing to the helpfulness of online hotel reviews: Does manager response play a role?. *International Journal of Contemporary Hospitality Management*, *28*(10), 2156-2177.

Lam, W., Liu, L., Prasad, S. T. S., Rajaraman, A., Vacheri, Z., & Doan, A. (2012). Muppet: MapReduce-style processing of fast data. *Proceedings of the VLDB Endowment*, *5*(12), 1814-1825.

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: traps in big data analysis. *Science*, *343*(6176), 1203--1205.

Leeflang, P. S., Verhoef, P. C., Dahlström, P., & Freundt, T. (2014). Challenges and solutions for marketing in a digital era. *European Management Journal*, *32*(1), 1-12.

Levin, N., Kark, S., & Crandall, D. (2015). Where have all the people gone? Enhancing global conservation using night lights and social media. *Ecological Applications*, *25*(8), 2153-2167.

Lev-On, A., & Steinfeld, N. (2015). Local engagement online: Municipal Facebook pages as hubs of interaction. *Government Information Quarterly*, *32*(3), 299-307.

Lima, A. C. E., de Castro, L. N., & Corchado, J. M. (2015). A polarity analysis framework for Twitter messages. *Applied Mathematics and Computation*, *270*, 756-767.

Liu, C., Chen, J., Yang, L. T., Zhang, X., Yang, C., Ranjan, R., & Kotagiri, R. (2014). Authorized public auditing of dynamic big data storage on cloud with efficient verifiable fine-grained updates. *IEEE Transactions on Parallel and Distributed Systems*, *25*(9), 2234-2244.

Liu, S. Q., & Mattila, A. S. (2017). Airbnb: Online targeted advertising, sense of power, and consumer decisions. *International Journal of Hospitality Management*, *60*, 33-41.

Liu, X., & Chen, H. (2013, August). AZDrugMiner: an information extraction system for mining patient-reported adverse drug events in online patient forums. In *International conference on smart health* (pp. 134-150). Springer, Berlin, Heidelberg.

Lohrmann, B., Janacik, P., & Kao, O. (2015, June). Elastic stream processing with latency guarantees. In *Distributed Computing Systems (ICDCS), 2015 IEEE 35th International Conference* on (pp. 399-410). IEEE.

Mariani, M. M., Di Felice, M., & Mura, M. (2016). Facebook as a destination marketing tool: Evidence from Italian regional Destination Management Organizations. *Tourism Management*, *54*, 321-343.

Marine-Roig, E., & Clavé, S. A. (2015). Tourism analytics with massive user-generated content: A case study of Barcelona. *Journal of Destination Marketing & Management*, 4(3), 162-172.

Martin-Sanchez, F., & Verspoor, K. (2014). Big data in medicine is driving big changes. *Yearbook of medical informatics*, *9*(1), 14.

McKelvey, K., DiGrazia, J., & Rojas, F. (2014). Twitter publics: How online political communities signaled electoral outcomes in the 2010 US house election. *Information, Communication & Society*, *17*(4), 436-450.

Miah, S. J., Vu, H. Q., Gammack, J., & McGrath, M. (2017). A big data analytics method for tourist behaviour analysis. *Information & Management*, *54*(6), 771-785.

Miller, H. J. (2013). Beyond sharing: cultivating cooperative transportation systems through geographic information science. *Journal of Transport Geography*, *31*, 296-308.

Mocanu, D., Baronchelli, A., Perra, N., Gon alves, B., Zhang, Q., & Vespignani, A. (2013). The twitter of babel: Mapping world languages through microblogging platforms. *PloS one*, *8*(4), e61981.

Mohr, D., Burns, M., Schueller, S., Clarke, G., & Klinkman, M. (2013). Behavioral intervention technologies: evidence review and recommendations for future research in mental health. *General hospital psychiatry*, *35*(4), 332--338.

Morone, F., & Makse, H. (2015). Influence maximization in complex networks through optimal percolation. *Nature*, *524*(7563), 65.

Morone, F., Min, B., Bo, L., Mari, R., & Makse, H. A. (2016). Collective influence algorithm to find influencers via optimal percolation in massively large social media. *Scientific Reports*, *6*, 30062.

Oboler, A., Welsh, K., & Cruz, L. (2012). The danger of big data: Social media as computational social science. *First Monday*, *17*(7).

O'Dea, B., Wan, S., Batterham, P. J., Calear, A. L., Paris, C., & Christensen, H. (2015). Detecting suicidality on Twitter. *Internet Interventions*, *2*(2), 183-188.

Ordenes, FV. , Ludwig, S., De Ruyter, K., Grewal, D., & Wetzels, M. (2017). Unveiling what is written in the stars: Analyzing explicit, implicit, and discourse patterns of sentiment in social media. *Journal of Consumer Research*, *43*(6), 875-894.

Ou, M., Cui, P., Wang, F., Wang, J., Zhu, W., & Yang, S. (2013, August). Comparing apples to oranges: a scalable solution with heterogeneous hashing. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 230-238). ACM.

Paldino, S., Bojic, I., Sobolevsky, S., Ratti, C., & González, M. C. (2015). Urban magnetism through the lens of geo-tagged photography. *EPJ Data Science*, *4*(1), 5.

Papacharissi, Z. (2016). Affective publics and structures of storytelling: Sentiment, events and mediality. *Information, Communication & Society*, *19*(3), 307-324.

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J. & Seligman, M. E. (2015). Automatic personality assessment through social media language. *Journal of personality and social psychology*, *108*(6), 934.

Park, S. B., Ok, C. M., & Chae, B. K. (2016). Using twitter data for cruise tourism marketing and research. *Journal of Travel & Tourism Marketing*, *33*(6), 885-898.

Peek, N., Holmes, J. H., & Sun, J. (2014). Technical challenges for big data in biomedicine and health: data sources, infrastructure, and analytics. *Yearbook of Medical Informatics*, *9*(1), 42.

Procter, R., Crump, J., Karstedt, S., Voss, A., & Cantijoch, M. (2017). Reading the riots: What were the police doing on Twitter?. *In Policing Cybercrime* (pp. 5-28). Routledge.

Procter, R., Vis, F., & Voss, A. (2013). Reading the riots on Twitter: methodological innovation for the analysis of big data. *International Journal of Social Research Methodology*, *16*(3), 197-214.

Ram, S., Zhang, W., Williams, M., & Pengetnze, Y. (2015). Predicting asthma-related emergency department visits using big data. IEEE J. *Biomedical and Health Informatics*, *19*(4), 1216-1223.

Ribarsky, W., Wang, D. X., & Dou, W. (2014). Social media analytics for competitive advantage. *Computers & Graphics*, *38*, 328-331.

Ross, M. K., Wei, W., & Ohno-Machado, L. (2014). “Big data” and the electronic health record. *YearBook of Medical Informatics*, *23*(01), 97-104.

Russell Neuman, W., Guggenheim, L., Mo Jang, S., & Bae, S. (2014). The dynamics of public attention: Agenda-setting theory meets big data. *Journal of Communication*, *64*(2), 193--214.

Sang, E. T. K., & van den Bosch, A. (2013). Dealing with big data: The case of twitter. *Computational Linguistics in the Netherlands Journal*, *3*(121-134), 2013.

Sharma, S. (2016). Expanded cloud plumes hiding Big Data ecosystem. *Future Generation Computer Systems*, *59*, 63-92.

Sharma, S., Tim, U. S., Wong, J., Gadia, S., & Sharma, S. (2014). A brief review on leading big data models. *Data Science Journal*, *13*, 138-157.

Shelton, T., Poorthuis, A., Graham, M., & Zook, M. (2014). Mapping the data shadows of Hurricane Sandy: Uncovering the sociospatial dimensions of ‘big data’. *Geoforum*, *52*, 167-179.

Shelton, T., Poorthuis, A., & Zook, M. (2015). Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information. *Landscape and Urban Planning*, *142*, 198-211.

Sikos, L. (2015). Mastering structured data on the Semantic Web: From HTML5 microdata to linked open data. *Apress*.

Singh, V. K., Gao, M., & Jain, R. (2012, October). Situation recognition: an evolving problem for heterogeneous dynamic big multimedia data. *In Proceedings of the 20th ACM international conference on Multimedia* (pp. 1209-1218). ACM.

Slavakis, K., Kim, S. J., Mateos, G., & Giannakis, G. B. (2014). Stochastic approximation vis-a-vis online learning for big data analytics [lecture notes]. *IEEE Signal Processing Magazine, 31*(6), 124-129.

Smith, M., Szongott, C., Henne, B., & Von Voigt, G. (2012, June). Big data privacy issues in public social media. *In Digital Ecosystems Technologies (DEST), 2012 6th IEEE International Conference* on (pp. 1-6). IEEE.

Stephansen, H. C., & Couldry, N. (2014). Understanding micro-processes of community building and mutual learning on Twitter: a 'small data'approach. *Information, Communication & Society*, *17*(10), 1212-1227.

Stephens, Z., and Lee, S., Faghri, F., Campbell, R., Zhai, C., Efron, M., Robinson, G. (2015). Big data: astronomical or genomical? *PLoS biology*, *13*(7), e1002195.

Stieglitz, S., Dang-Xuan, L., Bruns, A., & Neuberger, C. (2014). Social media analytics. *Wirtschaftsinformatik*, *56*(2), 101-109.

Stieglitz, S., Mirbabaie, M., Ross, B., & Neuberger, C. (2018). Social media analytics–Challenges in topic discovery, data collection, and data preparation. *International Journal of Information Management*, *39*, 156-168.

Tsou, M. H. (2015). Research challenges and opportunities in mapping social media and Big Data. *Cartography and Geographic Information Science*, *42*(sup1), 70-74.

Tufekci, Z. (2014). Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls. *Proceedings of the 8th International Conference on Weblogs and Social Media (ICWSM 2014)*, *14*, 505--514.

Tufekci, Z. (2014). Engineering the public: Big data, surveillance and computational politics. *First Monday*, 19(7).

Uldam, J. (2016). Corporate management of visibility and the fantasy of the post-political: Social media and surveillance. *New Media & Society*, *18*(2), 201-219.

Vatsavai, R. R., Ganguly, A., Chandola, V., Stefanidis, A., Klasky, S., & Shekhar, S. (2012, November). Spatiotemporal data mining in the era of big spatial data: algorithms and applications. In *Proceedings of the 1st ACM SIGSPATIAL international workshop on analytics for big geospatial data* (pp. 1-10). ACM.

Van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance \& Society*, *12*(2), 197--208.

Watson, H. J. (2014). Tutorial: Big data analytics: Concepts, technologies, and applications. *Communications of the Association for Information Systems*, *34*, 65.

White, M. (2012). Digital workplaces: Vision and reality. *Business information review*, *29*(4), 205-214.

Whittington, R. (2014). Information systems strategy and strategy-as-practice: a joint agenda. *The Journal of Strategic Information Systems*, *23*(1), 87-91.

Williams, M. L., & Burnap, P. (2015). Cyberhate on social media in the aftermath of Woolwich: A case study in computational criminology and big data. *British Journal of Criminology*, *56*(2), 211-238.

Williams, M. L., Burnap, P., & Sloan, L. (2017). Crime sensing with big data: The affordances and limitations of using open-source communications to estimate crime patterns. *The British Journal of Criminology*, *57*(2), 320-340.

Wilson, M. W. (2015). Morgan Freeman is dead and other big data stories. *Cultural geographies*, *22*(2), 345-349.

Wood, D., King, M., Landis, D., Courtney, W., Wang, R., Kelly, R. & Calhoun, V. D. (2014). Harnessing modern web application technology to create intuitive and efficient data visualization and sharing tools. *Frontiers in Neuroinformatics*, *8,* 71.

Wood, S., Guerry, A., Silver, J., & Lacayo, M. (2013). Using social media to quantify nature-based tourism and recreation. *Scientific Reports*, *3*, 2976.

Wu, K. J., Liao, C. J., Tseng, M. L., Lim, M. K., Hu, J., & Tan, K. (2017). Toward sustainability: using big data to explore the decisive attributes of supply chain risks and uncertainties. *Journal of Cleaner Production*, *142*, 663-676.

Xie, H., Li, Q., Mao, X., Li, X., Cai, Y., & Rao, Y. (2014). Community-aware user profile enrichment in folksonomy. *Neural Networks*, *58*, 111-121.

Xiang, Z., Schwartz, Z., Gerdes Jr, J., & Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management*, *44*, 120--130.

Yang, M., Kiang, M., & Shang, W. (2015). Filtering big data from social media–Building an early warning system for adverse drug reactions. *Journal of Biomedical Informatics*, *54*, 230-240.

Yaqoob, I., Hashem, I. A. T., Gani, A., Mokhtar, S., Ahmed, E., Anuar, N. B., & Vasilakos, A. V. (2016). Big data: From beginning to future. *International Journal of Information Management*, *36*(6), 1231-1247.

Yang, W., Lipsitch, M., & Shaman, J. (2015). Inference of seasonal and pandemic influenza transmission dynamics. *Proceedings of the National Academy of Sciences*, *112*(9), 2723-2728.

Young, S. D., Rivers, C., & Lewis, B. (2014). Methods of using real-time social media technologies for detection and remote monitoring of HIV outcomes. *Preventive medicine*, *63*, 112-115.

Young, S. D. (2015). A "big data" approach to HIV epidemiology and prevention. *Preventive medicine*, *70*, 17-18.

Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, *112*(4), 1036--1040.

Zheng, Y., Liu, T., Wang, Y., Zhu, Y., Liu, Y., & Chang, E. (2014, September). Diagnosing New York City's noises with ubiquitous data. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 715-725). ACM.

Zhong, E., Fan, W., Wang, J., Xiao, L., & Li, Y. (2012, August). Comsoc: adaptive transfer of user behaviors over composite social network. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 696-704). ACM.

Zhu, W., Cui, P., Wang, Z., & Hua, G. (2015). Multimedia big data computing. *IEEE multimedia*, *3*, 96-c3.