IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received February 12, 2020, accepted March 6, 2020, date of publication March 26, 2020, date of current version April 22, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2983656

Twitter and Research: A Systematic Literature Review Through Text Mining

AMIR KARAMI [1], MORGAN LUNDY [2], FRANK WEBB [3], AND YOGESH K. DWIVEDI [4]
[1]College of Information and Communications, University of South Carolina, Columbia, SC 29208, USA
[2]College of Arts and Sciences, University of South Carolina, Columbia, SC 29208, USA
[3]South Carolina Honors College, University of South Carolina, Columbia, SC 29208, USA
[4]School of Management, Swansea University, Swansea SA2 8PP, U.K.

Corresponding author: Amir Karami (karami@sc.edu)

This work was supported in part by the Social Science Research Grant Program, and in part by the Open Access Fund at the University of South Carolina.

ABSTRACT Researchers have collected Twitter data to study a wide range of topics. This growing body of literature, however, has not yet been reviewed systematically to synthesize Twitter-related papers. The existing literature review papers have been limited by constraints of traditional methods to manually select and analyze samples of topically related papers. The goals of this retrospective study are to identify dominant topics of Twitter-based research, summarize the temporal trend of topics, and interpret the evolution of topics withing the last ten years. This study systematically mines a large number of Twitter-based studies to characterize the relevant literature by an efficient and effective approach. This study collected relevant papers from three databases and applied text mining and trend analysis to detect semantic patterns and explore the yearly development of research themes across a decade. We found 38 topics in more than 18,000 manuscripts published between 2006 and 2019. By quantifying temporal trends, this study found that while 23.7% of topics did not show a significant trend ($P => 0.05$), 21% of topics had increasing trends and 55.3% of topics had decreasing trends that these hot and cold topics represent three categories: application, methodology, and technology. The contributions of this paper can be utilized in the growing field of Twitter-based research and are beneficial to researchers, educators, and publishers.

INDEX TERMS Literature review, social media, survey, text mining, topic modeling, Twitter.

I. INTRODUCTION

Twitter is a social media platform for computer-mediated online communication, which shapes an emerging social structure. This communication platform has 1.3 billion accounts and 336 million active users posting 500 million tweets per day [1].Twitter users can post comments known as "tweets," each restricted to 140 characters prior to October 2018 and currently, 280 characters. Unless tweets are made private, they are publicly available and Twitter users can show their reaction to and engagement with a tweet by sharing it on their profile (retweet), clicking the like button, tagging someone's user name, or responding to the author of the tweet [2].

Twitter has also provided Application Programming Interfaces (APIs) to facilitate data collection. To access the API, a user can apply for a developer account.[1] Following the application approval, the user has access to four keys: consumer key, consumer secret, access token, and access secret [20]. These keys authenticate the user to access Twitter data such as tweets and profile information. Twitter's own API is the most potent available tool for collecting data generated through the interaction of Twitter users. Representing different demographic categories, Twitter data is a diverse and salient data source for researchers [21], [22] and policymakers [23].

This global data source has earned the focus of several studies to address a wide range of research questions in different applications such as health and politics [24], [25]. While most studies used Twitter APIs for data collection such as [26], other studies manually collected data like [27], acquired Twitter data from commercial companies such as

The associate editor coordinating the review of this manuscript and approving it for publication was Xiao Liu.

[1]https://developer.twitter.com/en/apply

TABLE 1. Twitter literature review studies.

Publication Type [Ref]	Publication Year	Number of Analyzed Papers		Trend Analysis	Time Frame	#Years	Topic(s)
		Qualitatively	*Computationally*				
Journal [3]	2012	NA (#Ref=25)	NA	NA	NA	NA	Recommender Systems
Journal [4]	2013	NA (#Ref=57)	NA	NA	NA	NA	Election Prediction
Journal [5]	2014	21	NA	NA	2009-2014	5	Spam Detection
Book Chapter [6]	2015	NA (#Ref=110)	NA	NA	2006-2010	4	Politics
Book Chapter [7]	2015	NA (#Ref=30)	NA	NA	NA	NA	Socioeconomics
Book Chapter [8]	2015	NA (#Ref=31)	NA	NA	NA	NA	Psychology
Book Chapter [9]	2015	NA (#Ref=53)	NA	NA	NA	NA	Health
Book Chapter [10]	2015	NA (#Ref=34)	NA	NA	NA	NA	Disaster Analysis
Journal [11]	2015	31	NA	NA	2012-2014	2	Health
Journal [12]	2015	37	NA	NA	NA	NA	Disaster Analysis
Journal [13]	2016	127	NA	NA	2008-2014	6	Election
Conference [14]	2016	40	NA	NA	2006-2015	9	Politics
Journal [15]	2016	NA (#Ref=24)	NA	NA	NA	NA	Sentiment Analysis
Journal [16]	2017	137	NA	NA	2010-2015	6	Health
Journal [17]	2018	NA (#Ref=97)	NA	NA	NA	NA	Spam Detection
Journal [18]	2018	158	NA	NA	2009-2018	9	Disaster Analysis
Journal [19]	2020	60	NA	NA	NA	NA	Sentiment Analysis

[28], or utilized previously collected data from other studies like [29].

The diverse interests of interdisciplinary fields, like Twitter-based research, have constantly been evolving. Twitter-related studies have experienced an explosion of research development during the last decade. To reflect this development, some papers have reviewed relevant scientific publications systematically through traditional literature reviews. Table 1 shows the related studies found in the Google Scholar and Web of Science databases using two queries: *"Twitter AND Survey"* and *"Twitter AND Review*."

We have extracted some relevant features of these studies such as data collection method, number of analyzed papers, topic(s), and whether each included trend analysis. For the studies which did not mention the exact number of analyzed Twitter-based papers, we assumed the maximum number of reviewed papers was equal to the total number of references (#Ref). In the case of articles without research time frame information, we assumed the maximum range of time frame was the difference between the publication year and the launching year of Twitter that is 2006. For example, if a paper was published in 2012, the maximum number of years (#Years) is 6 (2012-2006).

While the relevant studies in Table 1 provide valuable insights, these studies have some limitations. First, a traditional literature review process starts with a large number of manuscripts. However, this manual process is not feasible. Therefore, researchers limit the number of papers to review – meaning that only a sample of relevant papers was selected, not all relevant papers [30]. Second, the first limitation indicates that the traditional approach could be prone to various biases such as focusing on journal articles and highly cited authors or studies [31]. Third, the traditional literature review process is not efficient, resulting in a time-consuming and labor-intensive process with a small sample of papers. For example, the maximum number of Twitter-related papers analyzed in a study was less than 160. Fourth, the existing literature does not provide a temporal trend perspective. Fifth, the current studies are too specific to a few topics, often only one. Sixth, they did not show a macro-level perspective that synthesized major disciplines and applications from the literature. Seventh, it is a difficult task to replicate and compare the results.

This study combats these limitations using a systematic approach to collect and analyze all relevant abstracts containing a condensed representation of the breadth of interdisciplinary Twitter-related studies. This endeavor can provide a macro-level intellectual viewpoint to track dynamic semantic patterns of what has been accomplished and what could be in future Twitter-based research.

Given the large number of Twitter-related studies, a systematic review can provide a valuable perspective to better define the literature landscape. The goals of this retrospective study are to identify dominant topics of Twitter-based research, summarize the temporal trend of topics, and interpret the evolution of topics withing the last ten years. To achieve these goals, utilizing computational methods offers a promising, more time efficient route [32]. Therefore, we applied text mining and trend analysis to reveal the main research themes and trends in the abstracts of more than 18,000 papers published between 2006 and 2019. Text mining is an effective and efficient approach that has been applied in a wide range of research interests. With the purpose of organizing and understanding documents, text mining disclosed hidden semantic patterns in a corpus. Then trend analysis, or the process of measuring the variation of topic distributions over several years, made it possible to track the temporal changes of research activities across the years represented in the corpus.

The contributions of this paper are four-fold. First, this is the first research that investigates thousands of papers to disclose the main themes of Twitter-related studies. Secondly, the proposed framework offers a fully functional approach to review a large number of research papers of any discipline and track their trends across several years. So, our approach can be utilized to understand the landscape of additional fields of research. Thirdly, the publicly available data of this research can be used to pursue further studies or replicate results of this study. Fourth, the trend analysis aids both supplies

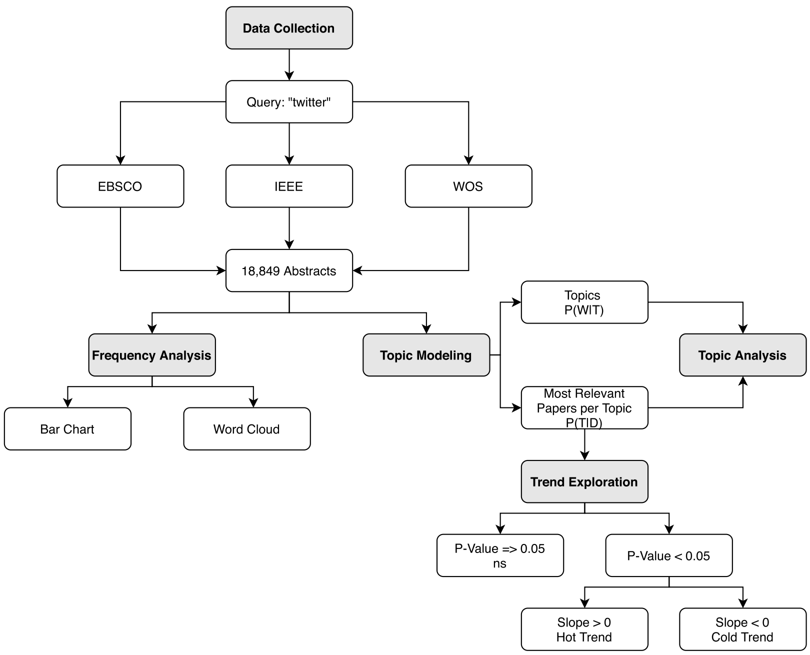

FIGURE 1. **Research framework.**

an overview of past research as well as insights for future research.

II. METHODS

This section introduces our research framework in four phases: data collection, frequency analysis, topic modeling, topic analysis, and trend exploration (Figure 1).

A. DATA COLLECTION

To better understand the current status of Twitter research, studying the content of journal and conference publications can help us to detect methodological and practical concepts [33], [34]. The abstract of an article contains concise information which discloses the larger picture of the article. We retrieved relevant abstracts containing "twitter" in their title or abstract from three major databases: Web of Science (WOS),[2] EBSCO,[3] and IEEE[4] in March 2019. We focused on the journal and conference abstracts published between 2006 and 2019. The collected data is available at https://github.com/amir-karami/Twitter-Research-Papers (Figure 1).

B. FREQUENCY ANALYSIS

Published in various domains, the abstracts are unstructured textual data and needed to be deciphered. Text mining techniques can provide exploratory analysis to detect semantic patterns [33]. To have an overall perspective, we analyzed the frequency of top-10 and top-50 words using the bar chart and word cloud, respectively. A word cloud represents the frequency of words in a corpus using word size – where larger size denotes higher frequency [36] (Figure 1).

C. TOPIC MODELING

As one of the most popular text mining methods, topic modeling is an efficient and systematic approach to analyze thousands of documents in a few minutes [37]. Among topic models, Latent Dirichlet Allocation (LDA) [38] is a valid and widely used model based on statistical distributions [39]–[43]. LDA assumes that there is an exchange between words and documents in a corpus representing by bag-of-words. LDA identifies semantically related words, which occur together in multiple documents of a corpus. These word lists or "topics" are then interpreted by human intuition as meaningful "themes" [44]. For example, LDA assigns "gene," "dna," and "genetic" to a theme interpreting as "genetic" (Figure 2).

LDA has been applied on both long-length (e.g., abstracts) and short-length (e.g., tweets) corpora for different applications such as health [26], [45]–[47], e-petitions [48], politics [29], [49], analysis of sexual harassment experiences [50], [51], opinion mining [52], investigation of social media strategy [53], [54], SMS spam detection [55], transportation literature [56], and mobile work [57], and literature review surveys relevant to depressive disorder [58], wearable technology [59], biomedical [36], [60], and medical case reports [37].

We considered abstracts as our documents, hereafter using abstract and document as interchangeable terms. LDA assigns a degree of probability for each set of words with respect to

[2] www.webofknowledge.com
[3] https://www.ebsco.com/
[4] https://ieeexplore.ieee.org/Xplore/home.jsp

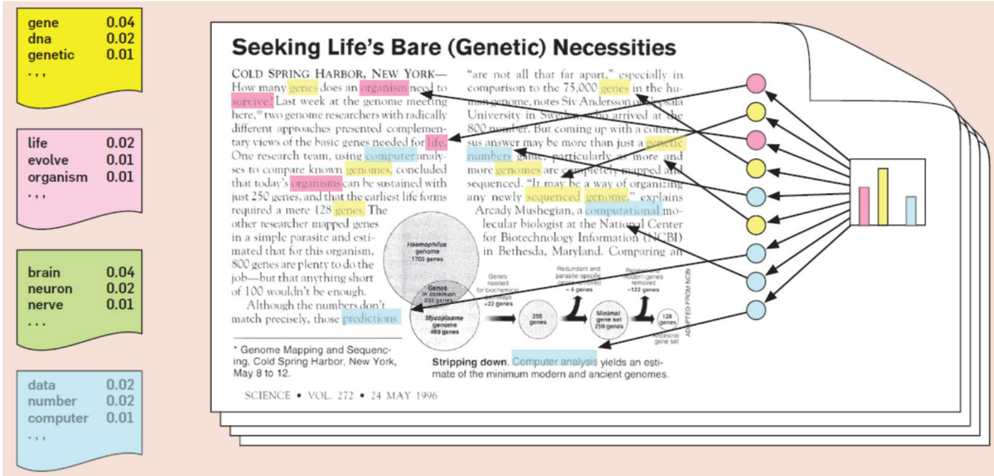

FIGURE 2. An example of LDA [35].

$$LDA \rightarrow \begin{array}{cc} & Topics \\ Words & \begin{bmatrix} P(W_1|T_1) & \dots & P(W_1|T_t) \\ \vdots & \ddots & \vdots \\ P(W_m|T_1) & \dots & P(W_m|T_t) \end{bmatrix} \end{array} \& \begin{array}{cc} & Documents \\ Topics & \begin{bmatrix} P(T_1|D_1) & \dots & P(T_1|D_n) \\ \vdots & \ddots & \vdots \\ P(T_t|D_1) & \dots & P(T_t|D_n) \end{bmatrix} \end{array}$$

FIGURE 3. Matrix interpretation of LDA.

each of the topics and a degree of probability for each of the topics with respect to each of the documents. In summary, LDA identifies the relationship between the topics and documents, $P(T|D)$, and words and topics, $P(W|T)$ (Figure 1). For n documents (abstracts), m words, and t topics, the outputs of LDA were: the probability of each of the words given a topic or $P(W_i|T_k)$ and the probability of each of the topics given a document or $P(T_k|D_j)$ (Figure 3). The top words per each topic, based on the descending order of $P(W_i|T_k)$, were used to represent the topics. We also used $P(T_k|D_j)$ to find the weight of topics per year. For example, if the first 100 abstracts (D_1, D_2, $\dots$, D_{100}) were published in 2009, the weight of T_1 in 2009 would be $\sum_{j=1}^{100} P(T_1|D_j)$.

D. TOPIC ANALYSIS

The inferred words of topics do not have inherent meaning without additional qualitative analysis. Two of the authors coded the topics independently by investigating the top-10 words based on descending order of $P(W|T)$ and top-5 abstracts based on descending order of $P(T|D)$ to decode the content of topics (Figure 1). For the disagreements, the two coders employed another round of annotation. Finally, a third annotator resolved any remaining disagreements. For example, the coders assigned "*Politics*" label to T_2 based on exploring the top-10 words (*"political, twitter, election, campaign, candidates, politicians, parties, party, elections, communication"*) in Table 2 and investigating top-5 documents ([61]–[65]) inferred by LDA.

E. TREND EXPLORATION

Statistical trend analysis of topics can aid in detecting hidden temporal patterns to move beyond surface-level observations about research trends. To explore the trends, we used a linear trend model to track the weight of topics within each year using $P(T|D)$ [58]. We used the R *lm* function to measure $P - Value$ and slope for each of topics. The $P - Value$ determines whether a trend is significant or meaningful ($P < 0.05$) and the slope shows whether a trend is increasing (hot) or decreasing (cold) (Figure 1).

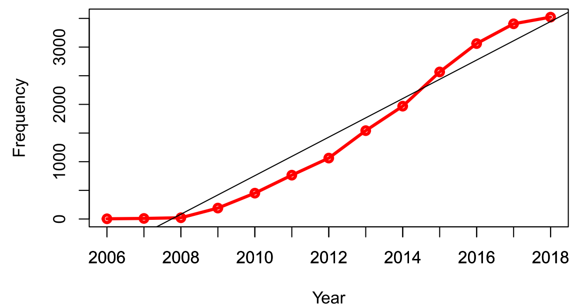

FIGURE 4. Frequency of twitter-related studies from 2006 to 2018.

III. RESULTS

After removing duplicate records, we found 18,849 unique abstracts written in English, published between 2006 and 2019. Figure 4 shows the total number of published papers per year over more than one decade along with the trend line, which illustrates a significant change ($P < 0.05$) with a positive slope indicating an increasing trend. Due to incompleteness, the 2019 papers published were excluded from the figure.

There were 1,706,918 tokens, of which 30,506 words were unique. Word frequency analysis illustrated that more than

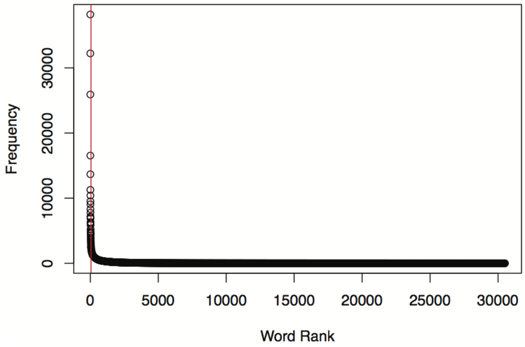

FIGURE 5. Frequency of words. The vertical line shows the cut-off point for the top-50 words in the word cloud.

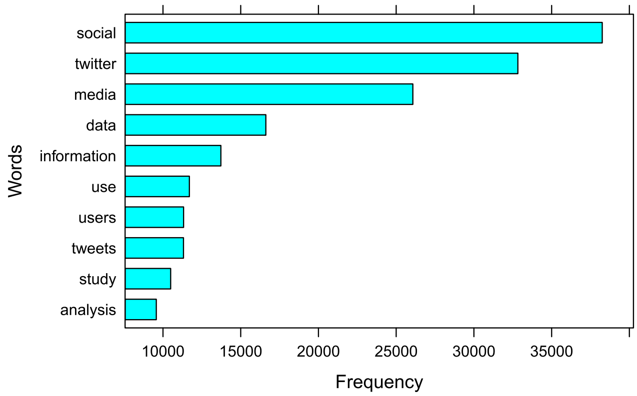

FIGURE 6. Frequency of the top-10 high-frequency words.

FIGURE 7. Word cloud of the top-50 high-frequency words.

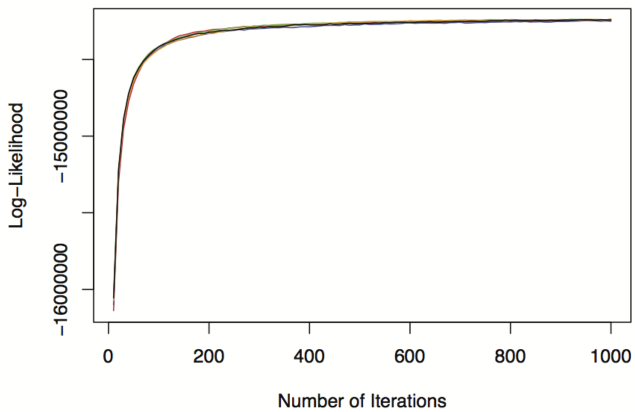

FIGURE 8. Convergence of the log-likelihood for 5 sets of 1000 integrations.

90% of the unique words had less than 100-frequency. The word frequency ranged between 2 and 38,219 occurrences with a median 5 and average 55.95. Figure 5 shows the alignment between the frequency of words and Zipf's law stating the inverse relationship between the frequency of words and their frequency rank [66]. This figure also shows the position of the top-50 words among more than 30,000 words in the word cloud using the vertical line.

Figures 6 and 7 show the top-10 high-frequency words and the top-50 high- frequency words using the bar chart and word cloud, respectively. We saw three categories of words. The first one was expected-social media words such as *"users"* and *"tweets"* which were among the high-frequency words. The second category represented common research paper-related words like *"data," "information," "use," "study,"* and *"analysis."* The third category illustrated research topics such as *"health," "political," "news,"* and *"behavior."*

Beyond preliminary frequency assessments, we used LDA to disclose the hidden semantic structure of papers. LDA has an important pre-processing step, which is defining the optimal number of topics. We used the latent concept modeling [67] to estimate the optimum point. This model maximizes the overall dissimilarity between the word distributions of topics. Using the *ldatuning* R package,[5] we found the optimum number of topics at 40 by applying the latent concept modeling on the number of topics from 2 to 100. After removing stop words (e.g., "the" and "a"), we applied the Mallet implementation of LDA [68] with its default settings on the abstracts to detect the 40 topics.

To evaluate the robustness of LDA, this study investigated the log-likelihood for five sets of 1000 iterations (Figure 8). The comparison of the mean and standard deviation of iterations showed that there was not a significant difference ($P => 0.05$) in log-likelihood convergence over the iterations.

Table 2 and Figure 9 show the detected 40 topics and their weight, respectively. For example, T_2 was named "Politics" based on interpretation of "*political, twitter, election, campaign, candidates, politicians, parties, party, elections, communication*."

[5]https://cran.r-project.org/web/packages/ldatuning/vignettes/topics.html

TABLE 2. The 40 topics of twitter-related studies generated by LDA.

Label	ID	Topic
Nature/Tourism	T_1	twitter tourism human early calls tourists species complex destination sound
Politics	T_2	political twitter election campaign candidates politicians parties party elections communication
Disaster Management	T_3	information crisis social media disaster events emergency twitter response management
Sentiment Analysis	T_4	sentiment analysis classification tweets twitter learning machine accuracy data text
Topic Modeling	T_5	topics topic events tweets event twitter information method detection semantic
Spatial Analysis	T_6	data location spatial patterns urban city traffic human activity geographic
Digital Communication	T_7	mobile internet communication technology information devices applications people technologies access
Medical Education	T_8	twitter medical library professional nursing academic conference education librarians media
Ethics, Law, and Privacy	T_9	privacy law security issues twitter legal speech concerns ethical policies
Information Behavior	T_{10}	twitter social factors influence effect model information perceived theory behavior
Social Movement	T_{11}	news media journalists twitter social political movement traditional protests activists
Abstract Words	T_{12}	twitter topics social include presented discussed offers reports mentions focuses
Social Media Platforms	T_{13}	social media platforms facebook youtube content instagram role usage communication
Content Analysis	T_{14}	tweets twitter users content analysis messages number hashtags posted retweets
Disease Surveillance	T_{15}	health public disease twitter surveillance trends tweets information outbreak monitoring
Social Media Technology	T_{16}	social information facebook sites networking online websites internet twitter blogs
Sports/Entertainment	T_{17}	twitter women television sports media fans game celebrities athletes men
Big Data Mining	T_{18}	data analysis social twitter big mining techniques large collection process
Cloud Technology	T_{19}	system data performance processing applications cloud distributed network real-time service
Stock Market	T_{20}	sentiment stock emotions negative market positive analysis tweets predict financial
Recommender System	T_{21}	users information social twitter weibo recommendation content interests network services
Altmetrics	T_{22}	impact articles twitter scientific metrics academic journals publications altmetrics citations
Survey	T_{23}	survey twitter respondents data age significant differences gender sample participants
Experiment	T_{24}	posts social media quality activity online engagement group participants recruitment
NLP	T_{25}	language text words linguistic corpus natural features written messages processing
Public Relations	T_{26}	media social communication twitter organizations public engagement stakeholders platforms relations
Opinion Mining	T_{27}	public media twitter opinion issues social analysis discussion debate content
Health Discussion	T_{28}	health patients media social professionals medical public support treatment awareness
Citizen-Government Interaction	T_{29}	public government local countries citizens national policy global international agencies
Community Analysis	T_{30}	social community groups networks users support members interactions sharing connections
Marketing	T_{31}	marketing brand companies business consumers customers product service media advertising
Social Media Interactions	T_{32}	social facebook online networking twitter users relationship people life relationships
Network Analysis	T_{33}	network social information model influence users diffusion interactions nodes analysis
Image/Video Analysis	T_{34}	content images visual media videos social shared online photos multimedia
Drug	T_{35}	health drug online tobacco smoking media marijuana drugs posts abuse
Activism Discourse	T_{36}	digital media discourse social practices identity ways space cultural communication
Web Technology	T_{37}	knowledge design work web software development technologies framework systems tools
Graph Mining	T_{38}	model proposed algorithm data method graph prediction performance experiments clustering
Pedagogical Use	T_{39}	students learning twitter education university teaching teachers tools educators classroom
Security	T_{40}	social users twitter accounts detection features spam bots fake malicious

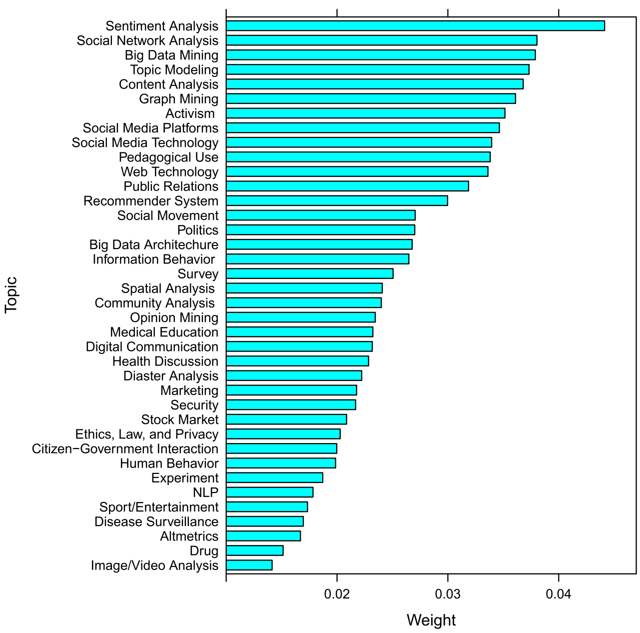

FIGURE 9. Sorted the 38 meaningful and relevant topics from the highest to the lowest weight.

We removed T_1 and T_{12} because T_1 was an unrelated topic representing animal calls and T_{12} contained common terms found in writing abstracts that cannot be mapped to a specific research theme. The weight of topics ranges from 0.044 for sentiment analysis to 0.014 for image/video analysis with median 0.024 and average 0.026. After the coding process and removing the two unrelated topics, we investigated the yearly trend of 38 topics based on aggregating $P(T|D)$ at the year level for ten years from 2009 to 2018 (Table 3). Due to the low number of publications, we did not consider 2006, 2007, and 2008 trends. Out of the 38 topics, nine topics did not have significant trends ($P => 0.05$), but 29 topics had significant trends ($P < 0.05$) including 21 hot and 8 cold topics. Among the 29 topics with a meaningful trend, 17 topics had extremely significant ($P < 0.001$), 4 topics had very significant ($P < 0.01$), and 8 topics had significant ($P < 0.05$) changes (Figure 10).

To provide examples for each of the 38 meaningful topics and better illustrate their relevance, the coders analyzed the five most related papers for each of the topics based on descending order of $P(T|D)$. The following summaries provide more details based on studying 190 articles (5 articles per topic $\times$ 38 topic) offered by LDA. We also provide useful additional resources for methodology-related topics.

Politics: Several studies have utilized Twitter data for election purposes such as the impact of Twitter adoption on the voting behavior of congressmen [61], analyzing populist social media strategies of political actors [62], examining the Twitter engagement between the candidates and followers in the 2010 US midterm elections [63], investigating the variance in partisan rhetoric of the US senators in their tweets [64], and examining the tweets posted by the candidates

TABLE 3. Linear trend of the 38 meaningful and relevant topics (^{ns}P => 0.05; *P < 0.05; **P < 0.01; ***P < 0.001).

Topic	Slope	P-Value	Topic	Slope	P-Value
Politics	0.001824223	***	Altmetrics	0.000572231	*
Disaster Analysis	-	ns	Survey	0.001744732	***
Sentiment Analysis	0.00529812	***	Experiment	0.000990999	***
Topic Modeling	0.00254395	**	NLP	0.000934663	***
Spatial Analysis	0.001899396	***	Public Relations	-	ns
Digital Communication	-0.002895787	***	Opinion Mining	0.001237115	***
Medical Education	-0.002131603	*	Health Discussion	0.00122912	***
Ethics, Law, and Privacy	-0.001903798	***	Citizen-Government Interaction	-	ns
Information Behavior	0.001648176	*	Community Analysis	-0.000824096	*
Social Movement	-	ns	Marketing	-0.001077084	*
Social Media Platforms	0.00120391	*	Human Behavior	-	ns
Content Analysis	0.002016129	***	Social Network Analysis	-	ns
Disease Surveillance	0.000810952	***	Image/Video Analysis	-	ns
Social Media Technology	-0.010435213	***	Drug	0.000763579	**
Sport/Entertainment	0.000510283	*	Activism	0.001650483	**
Big Data Mining	0.002807851	***	Web Technology	-0.001359937	*
Big Data Architecture	-	ns	Graph Mining	0.002862427	***
Stock Market	0.001817865	***	Pedagogical Use	-0.003288796	***
Recommendation System	-	ns	Security	0.001168514	**

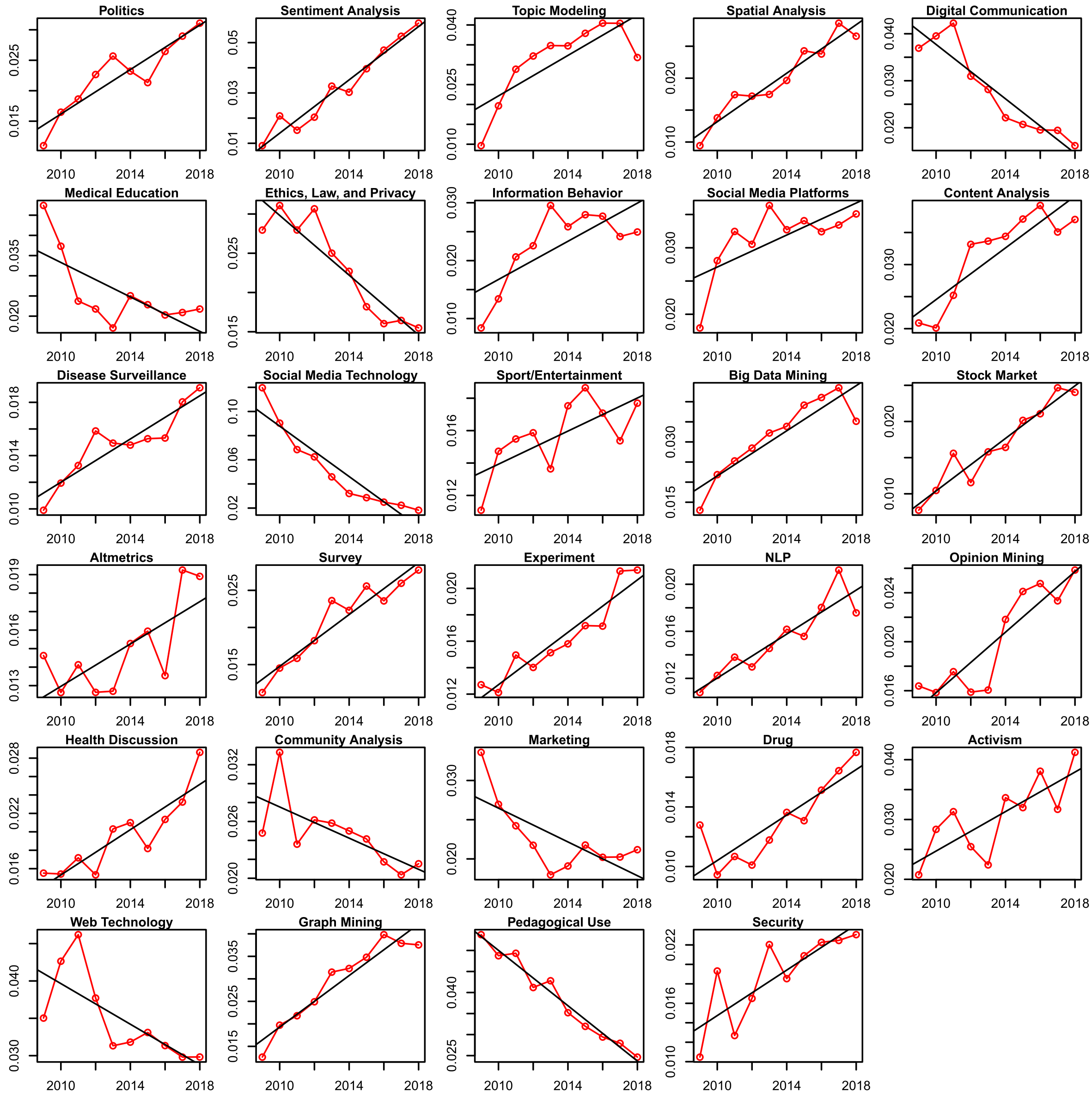

FIGURE 10. Yearly trend of the 29 topics with P < 0.05.

during the Australian federal elections in 2013 and 2016 [65]. Having an extremely significant change (P < 0.05), politics-related twitter studies have a positive slope indicating increasing research activity.

Disaster Analysis: This category of study has used Twitter data for disaster analysis such as investigating the activity of rumor-spreading users and the debunking response behaviors during Hurricane Sandy in 2012 and the Boston Marathon bombings in 2013 [69], understanding how retweets spread information during the Fukushima nuclear radiation disaster [70], investigating the difference between the information shared on public safety organizations' websites and their twitter accounts during the 2015 winter storms in Lexington, KY [71], analyzing the tweets of government organizations during Hurricane Harvey to disclose their disaster response strategies [72], and examining the use of weather warning related hashtags to disclose their effectiveness for information retrieval and processing [73]. The trend analysis does not show a significant change ($P => 0.05$) for the disaster analysis category.

Sentiment Analysis: This topic represents a research methodology that aids researchers in assessing whether the sentiment polarity of tweets is positive, negative, or neutral. The studies reviewed here proposed new methods for sentiment analysis purposes such as using N-gram analysis based on diabetes ontologies for aspect-level sentiment analysis [74], exchanging sentiment labels between words and tweets using feature vectors to reduce the cost of data annotation in supervised methods [75], utilizing a pattern-based approach to classify tweets into seven classes including happiness, sadness, anger, love, hate, sarcasm and neutral [76], proposing a supervised sentiment analysis method using emotion-annotated tweets, unlabeled tweets, and hand-annotated lexicons [77], and utilizing semantic relationships between words and n-grams analysis for measuring public sentiment [78]. Like the politics-related twitter studies, sentiment analysis has a positive slope indicating increasing research activities over time. For extensive details and discussions on sentiment analysis, refer to [79], [80].

Topic Modeling: This method discloses the hidden semantic structure of tweets. The studies in this theme have proposed customized topic models for Twitter data or utilized already developed topic models such as incorporating Twitter-LDA, WordNet, and hashtags to enhance the quality of topic-discovery [81], developing a topic model based on high utility pattern mining to detect emerging topics [82], utilizing an already developed method, called hierarchical Dirichlet processes (HDP), to detect all posts related a given event [83], proposing a model based on Latent Dirichlet Allocation (LDA) to extract key phrases for social media events [84], and integrating the recurrent Chinese restaurant process and word co-occurrence analysis to propose a nonparametric topic model for short text documents such as tweets [85]. It is worth mentioning that the studies that utilized pre-existing topic models used a qualitative approach for coding topics. In addition, the current topic models are based on five main approaches: linear algebra [86], probability [87], statistical distributions [38], neural networks [88], and fuzzy clustering [44], [89]–[91]. The topic modeling theme shows an extremely significant change with a positive slope. For more details and discussions on topic modeling, refer to [35], [88], [92]–[94].

Spatial Analysis: This technique examines Twitter geo-located data such as the assessment of spatial distribution of people's exposure to burglaries and robberies [95], exploring different types of human activities such as shopping in Boston and Chicago [96], creating population maps using geo-located tweets in Indonesia [97], mapping mobility patterns in one of Chile's medium-sized cities [98], and utilizing Twitter data for spatial analysis of crashes in Los Angeles [99]. The trend of spatial analysis illustrates an extremely significant change with an increasing slope. For more details on spatial analysis, refer to [100]–[103].

Digital Communication: This theme represents the papers which studied digital communication issues of Twitter such as investigating the impact of student interactions with social media on their daily lives [104], utilizing Twitter to promote library services [105], assessing fitness, diabetes, and meditation mobile applications based on communications in tweets [106], understanding how Twitter has changed the communication between surgeons and colleagues and patients [107], and investigating the Twitter application for medical communication [108]. Having an extremely significant change, the digital communication theme shows a negative slope indicating decreasing research activities.

Medical Education. This theme shows the research on using Twitter for medical education purposes such as online discussion on program evaluation [109] and research papers [110], professional development [111], and utilizing Twitter in emergency medicine residency programs [112] and medical conferences [113]. Having a significant change, the medical education theme have a decreasing trend.

Ethics, Law, and Privacy: This theme represents the studies discussing ethical and privacy issues such as proposing ethical frameworks for social media platforms to better protect free speech and prevent harms [114], investigating the challenges of current laws for social media legal cases [115], exploring Twitter regulatory mechanisms to protect the users against criminal offenses [116], evaluating the freedom of expression on Twitter based on the interpretation of First Amendment [117], and understanding the legal position of Twitter's services in the US Federal criminal courts [118]. Having an extremely significant change, the theme of ethics, law, and privacy illustrates decreasing research activities.

Information Behavior. This theme includes papers that studied information behavior on Twitter such as examining the impact of content and context factors on retweeting [119], disclosing the motivations behind the continuous use of Twitter services [120], analyzing the factors impacting Twitter's perpetuation [121], investigating the brand-following behavior of Twitter users based on the theory of planned behavior [122], and understanding whether tweets can increase the news knowledge of users [123]. Having a significant change, the theme of information behavior analysis has a decreasing trend.

Social Movement: This topic analyzed the tweets of social movements such as the 2011 revolution in Egypt [124], 2011 revolution in Tunisia [125], 2011 occupy Wall Street protest [126], 2009 Iranian presidential election protests [127], and 2014 protests for the kidnapped and massacred students in Mexico in 2014 [128]. The theme of social movement does not show a significant change over time.

Social Media Platforms. This topic is illustrated by papers which explored and compared Twitter and other social media platforms such as Reddit, Foursquare, Tumblr, Instagram, Facebook, Wikipedia, and YouTube to address different research purposes such as exploring various definitions of social media [129], analyzing uses of social media platforms [130], investigating social media platforms for communication in academic libraries [131], understanding how radiologists use different social media platforms [132], and exploring the impact of social media on the competitiveness, structure, and processes of an organization [133]. The theme of social media platforms has a significant change with an increasing trend.

Content Analysis: This method discloses concepts, detects their relationships, and draws semantic inferences by interpreting and coding tweets. Examples of content analysis studies include characterizing the content of tweets relating to indoor tanning [134], investigating the tweets containing bullying-related words [135], exploring tweets related to Planned Parenthood [136], analyzing the tweets related to kidney stones [137], and comparing the marijuana-related tweets posted before and after the 2012 US election [138]. While the topic modeling related studies analyzed the content of all collected tweets, the content analysis related papers investigated a small sample of tweets. For example, the researchers analyzing the bullying-related tweets selected a sample of 10,000 tweets from millions of potentially related tweets for content analysis. Like topic modeling, content analysis shows an extremely meaningful trend with a positive slope. For more details about content analysis, refer to [139]–[144].

Disease Surveillance: This topic represents the papers utilizing Twitter data for monitoring diseases such as the 2010 Haitian Cholera outbreak [145], 2015 Middle East respiratory syndrome outbreak [146], Chikungunya virus [147], Zika virus [148], and infectious eye diseases [149]. The theme of disease surveillance shows an extremely significant change with a positive slope indicating increasing research activities.

Social Media Technology: This theme covers the papers studying different aspects of social media technology such as information sources used by students [150], financial professionals [151], and library services [152] and exchanging health information [153], [154]. Having an extremely significant change, the theme of social media technology faces decreasing research activities.

Sport/Entertainment: This topic illustrates the manuscripts studying applications of social media for sport and entertainment such as campaigning again racism at the 2016 Oscars [155], TV [156], social media [157], and women's soccer [158], and the #metoo movement against sexual harassment and sexual assault [159]. The trend analysis shows that sport/entertainment related studies have a significant change with a positive slope indicating increasing research activities.

Big Data Mining: Having a foundation in statistics, artificial intelligence, and machine learning, this method detects patterns and correlations within big Twitter data for different applications such as sentiment analysis [160], opinion mining [161], analyzing structured and unstructured data [162], exploring the trend change of languages [163], and spatial analysis [164]. The trend analysis shows that big data mining is one of the attractive research topics with an extremely significant increase and change over time. For more details about big data mining, refer to [165]–[169].

Big Data Architecture: This topic represents the studies focusing on the architecture of big data such as proposing new platforms for analyzing real-time social media data [170], cloud systems in different locations [171], data storage and management [172], data streaming [173], and distributed storage systems [174]. The trend of big data architecture does not show a significant change.

Stock Market: This theme investigates the stock market applications of Twitter data such as investigating the relationship between relevant Twitter trends and trends of stock options pricing [175], studying Twitter as a useful information resource for financial market activity [176], exploring the relationship between the Twitter daily happiness trend and the stock market trends [177], and examining the impact of positive, negative, and neutral tweets on price returns [178] and renewable energy stocks [179]. The stock market theme shows a positive extremely significant change indicating high research activities.

Recommendation System: This topic represents the Twitter-related studies focusing on recommendation systems for different purposes such as recommending new followers [180], detecting interests of users [181], and developing a personalized recommender system based on relationships of users [182], a personalized news recommender system utilizing news popularity on Twitter [183], and an emotion-based music recommender system [184]. The trend analysis does not show a significant change for the recommendation system topic.

Altmetrics: This topic considers non-traditional scholarly impact measurements based on web activities such as tweeting. This theme represents the studies analyzing research-related discussions on Twitter such as understanding the impact of online non-social media discussions on social media activities such as liking and tweeting [185], evaluating mentions of papers as an alternative method for research assessment [186] in different domains such as humanities [187] and dental research [188], and comparing alternative metrics such as Twitter mentions and traditional metrics like citations in medical education [189]. The altmetrics topic has a positive significant trend indicating high research activities. More discussions on altmetrics can be found in [190]–[195].

Survey. Using statistical techniques, this research method investigates a data sample collected from a population by traditional data collection techniques like developing a questionnaire. This topic represents the manuscripts in two directions. The first one is using surveys for studying Twitter-related issues such as investigating social media users' preference for oral health information searching [196], analyzing the relationship between gender, personality, and Twitter addiction [197], sleep disturbance and social media use [198], and eating concerns and social media use [199]. The second direction is posting a survey on Twitter to hire participants for research such as evaluating a home drinking assessment scale based on the initial psychometric properties [200]. The studies based on the survey topic have an extremely significant change with a positive trend. For more details on survey methods, refer to [201]–[204].

Experiment. This method investigates the impact of changing the independent variable (the cause) on the dependent variable (the effect). Experiment-related studies developed trials for different research purposes such as investigating the impact of social media activities on web visits of a journal [205], posting tweets on promoting knowledge products [206], developing engagement strategies on social media webpage visits of a state health-system pharmacy organization [207], and lifestyle related tweets on weight loss [208], and examining the relationship between Twitter use, physical activity, and body composition [209]. The experiment theme has high research activity with an extremely significant change. More discussions and details on developing experiments can be found in [210], [211].

Natural Language Processing (NLP): This method utilized artificial intelligence techniques to analyze natural language text or speech. NLP has been utilized for different purposes on Twitter such as analyzing the use of diminutive interjections [212], the meanings of different combinations of hashtags starting with #jesuis [213], and the geographic patterns of African American Vernacular English posts [214], proposing a normalization method to convert Malay Tweet language to standard Malay [215], and decoding different languages on tweets [216]. Having an extremely significant change, the NLP theme has an increasing trend. For more discussion on NLP, refer to [217]–[221].

Public Relations: This theme is seen in papers which utilize Twitter in public relations for different firms such as nonprofit organizations [222], media organizations [223], state health departments [224], global organizations [225], and Fortune 1000 companies [226]. The trend of public relations does not have a significant change over time.

Opinion Mining: While sentiment analysis explores public feeling about a given topic, opinion mining investigates the reasons or driving forces behind the public feeling. Within this topic, studies investigated public opinion with respect to different issues such as the 2015 Ireland same-sex marriage referendum [227], climate change [228], the Dakota access pipeline [229], the U.S. nuclear energy policy [230], and the 2011 Norwegian election [231]. Having an extremely significant change, the opinion mining theme has an increasing trend. For more discussions and details on opinion mining, refer to [79], [232].

Health Discussion: This theme shows Twitter-related studies focused on health topics such as orthodontic retention [233], lung cancer [234], depression and schizophrenia [235], diabetes [236], and mental health [237]. Having an extremely significant change, the health discussion theme has a positive slope.

Citizen-Government Interaction: This topic appears in studies which investigated the interaction between governments and citizens with respect to different issues such as foreign policy in Canada [238], public policy of the 75 largest U.S. cities [239], food policy in South Korea [240], the transparency of public agencies in Thailand [241], and presentational strategies of the Canadian Toronto Police Service [242]. The trend of the citizen-government interaction theme does not have a meaningful change.

Community Analysis: This topic represents the research which analyzed Twitter communities developed for different purposes such as peer-to-peer file sharing [243], developing personal communities [244], learning [245], and support for weight loss [246] and physical activity [247]. The community analysis has a significant negative slope indicating decreasing research activities.

Marketing: This theme shows papers which analyzed Twitter data for marketing functions such as customer knowledge management [248], competitive advantage [249], consumer behavior analysis [250], consumer opinion analysis [251], and brand-building and customer acquisition [252]. Like community analysis, marketing has a significant decreasing trend.

Human Behavior: This topic shows studies which investigate the intersection of social networking and human behavior such as interpersonal relationships [253], bridging and bonding social capital [254], organizational processes and employee performance [255], face-to-face pro-social behaviors [256], and internal and external motivations for the use of social media [257]. The trend of human behavior does not have a meaningful change.

Social Network Analysis: This method utilizes graph theory to characterize the structure of social networks. This topic represents the studies which analyzed Twitter networks for different purposes such as consensus formation processes [258], classifying complex networks [259], and information diffusion methods like epidemic [260], null [261], and evolutionary game theory [262] models. The trend of social network analysis does not have a significant change. For extensive details and discussion on social network analysis, refer to [263]–[266].

Image/Video Analysis: This method uses qualitative approaches to code image/video data of online posts and categorize them. This topic is illustrated by studies that studied images and videos with respect to different issues such as fitness and thinness [267], the Boston marathon bombing [268], eating disorders [269], life experience [270], the online activists of Islamic State in Iraq and Syria (ISIS) [271]. The trend of image/video analysis does not have a meaningful change. For more discussion on image/video analysis, refer to [272], [273].

Drug: This topic shows Twitter-related studies investigated different drugs such as E-cigarettes [274], blunts [275], marijuana and alcohol [276], hookah [277] and opioids [278]. The Drug topic has a very significant positive slope indicating high research activities.

Activism: This topic illustrates the studies which investigated Twitter activism such as feminism [279], African American activism [280], resistance to political movements [281], indigenous activism [282], and anti-racism [283]. The activism topic has a very significant positive slope indicating high research activities.

Web Technology: This theme represents the papers focused on web technology issues such as comparing Application Programming Interfaces (APIs) of multiple companies like Twitter and Facebook [284], exploring different web development platforms [285], analyzing different aspects of open APIs [286], archiving social media (Web 2.0) content [287], and developing new open source applications based on the Twitter API to archive data for research purposes [288]. The web technology topic has a significant negative slope indicating decreasing research activities.

Graph Mining: This method explores the characteristics of graphs to recognize and predict patterns. This topic includes studies that utilized Twitter data for the evaluation of graph mining models developed for different purposes such as clustering [257], triangle counting [289], anomaly detection [290], understanding dynamic graphs [291], and graph-constrained coalition formation [292]. Having an extremely significant change with a positive slope, the graph mining theme has high research activities. For more discussion on graph mining, refer to [293]–[295].

Pedagogical Use: This theme illustrates the studies that used Twitter for educational purposes such as enhancing learning [296], engaging students [297], designing an open online course [298], professional purposes [299], and developing a professional learning network [300]. Having an extremely significant change, the pedagogical use theme has a negative slope.

Security: This topic shows studies which proposed detection methods for security issues such as spams [301], social bots [302], malicious accounts [303], fake identities [304], and suspicious URLs [305]. The security topic has a very significant change with an increasing trend.

IV. DISCUSSION

To better understand the growing field of Twitter-related studies, this study provides a bird's eye view to explore the overall and temporal patterns of major topics within the past years of Twitter related papers. This research has some methodological advantages over traditional literature review studies. First, traditional methods were limited to a small sample of related papers, while this study investigates all relevant papers in three well known and popular databases. Second, traditional methods develop a codebook based on studying a data sample to detect themes within a set of papers. It is humanly impossible to recognize all themes within scholarly publications, but this research utilizes an unsupervised machine learning technique that is an efficient approach that does not need the codebook. Third, previous research defined the total number of themes in traditional methods, while this paper applies an estimation method to find the optimal number of topics. Fourth, while previous studies selected a sample of papers for the traditional qualitative coding, we use a computational approach to find the most related papers for each of the topics systematically. Fifth, the topic discovery process in this paper has been implemented in an efficient process; in comparison, traditional methods face a time-consuming and labor-intensive process.

Among the meaningful and relevant topics, while 23.7% of topics did not show a significant trend over time, 55.3% had increasing (hot) and 21% had decreasing (cold) trends. These topics can be discussed in three categories: (1) application, (2) methodology, and (3) technology (Table 4). The topics in the first category are made up of studies which utilized Twitter data for different applications including business and management such as marketing, education like medical pedagogy, health such as disease surveillance, media like social media platforms, politics such as elections, and psychology and society like information behavior. The second category indicates the methodology related topics in six sub-categories:

- computational (analytical) techniques (e.g., sentiment analysis)
- qualitative techniques (e.g., traditional content analysis)
- mixed methods using a computational techniques (e.g., text mining) for detecting topics in a corpus and a qualitative approach (e.g., coding) for disclosing the theme of topics
- quantitative techniques (e.g., survey)
- research facilitation to find and hire participants by posting surveys on Twitter
- data resources for evaluating new methods

The third category of topics investigated different technological aspects of social media such as APIs. While the technology related topics have a negative slope, all the methodology-related and most of the application-related topics had high research activities between 2009 and 2018.

Among the application-related topics, politics, stock market, and information behavior were the top hot topics, and marketing, ethics, law, and privacy, and medical education were the top cold topics. Considering the methodology related topics, while survey and experiment are traditional qualitative and quantitative methods, the rest of methodologies are computational methods. Therefore, it seems that the large scale of Twitter data was the reason that the research activities using the computational research methods were more prevalent than the ones applying the qualitative and

TABLE 4. Categories of topics. The topics with $P < 0.05$ ranked from the highest to the lowest slope values.

Category	Trend		
	$P < 0.05$		$P => 0.05$
	$Slope > 0$	$Slope < 0$	
Application	Politics Stock Market Activism Information Behavior Health Discussion Social Media Platforms Security Disease Surveillance Sport/Entertainment Drug	Marketing Ethics, Law, and Privacy Medical Education Pedagogical Use Community Analysis	Disaster Analysis Social Movement Public Relations Citizen-Government Interaction Human Behavior
Methodology	Sentiment Analysis Graph Mining Big Data Mining Topic Modeling Content Analysis Spatial Analysis Survey Opinion Mining Altmetrics Experiment NLP	NA	Social Network Analysis Image/Video Analysis
Technology	NA	Social Media Technology Web Technology Digital Communication	Big Data Architecture Recommendation System

quantitative methods. Table 4 shows sentiment analysis, big data mining, and topic modeling were not only the hot topics but also the high-ranking topics among all the detected topics. In the technology category, the meaningful trends were decreasing, indicating low research activities. Increasing or decreasing trends can also correspond to the research needs and opportunities [306].

In summary, evidenced by the increasing number of publications and trend of most topics having a significant change, Twitter-based research will continue to evolve in formal sciences (e.g., computer science), natural science (e.g., health), and social science (e.g., political science). Due to the positive slope in Figure 4, we also expect to see more Twitter-based research activities in the following years, overall.

Our findings show that researchers utilized different data sizes including a few thousand (small scale), several hundred thousand (medium scale), and millions (large scale) of data records on Twitter. These studies used structured data (e.g., the number of followers) or unstructured data (e.g., image/video and tweets), which unstructured data has been investigated more than structured data on Twitter.

Researchers applied qualitative, quantitative, computational (analytics), and mixed methods to address their research goals and questions. Due to the massive number of tweets, the most popular research approaches were computational methods, including supervised methods such as classification techniques and unsupervised methods like clustering techniques.

The recognized static and dynamic patterns disclose a macro-level perspective into some aspects as follows. First, the frequency analysis provides an overall picture of the Twitter-related studies. Second, the detected topics illustrate major Twitter research themes. Third, the weight of topics shows the importance or popularity of topics. Fourth, the temporal topic analysis demonstrates the changes in research interests during the time frame. Fifth, the detected trends help to provide an overview of past studies and offer insight for future studies. Sixth, the top words of topics can be used as keywords to assist researchers in finding relevant studies with respect to a topic for more in-depth analysis of that topic.

This study is beneficial to researchers for understanding the larger picture of Twitter-related studies and their trends, to educators for defining the scope of social media related courses, to journal editorial boards and publishers for categorizing research topics in social media, to publishers for investing more on hot topics, and to science policymakers and funding agencies for developing strategic plans.

V. CONCLUSION

This research proposes a systematic framework to have a better understanding of Twitter-related studies and their hot and cold topics. Our findings show the potential of this research to understand large-scale research corpora and the usefulness of text mining and trend analysis to investigate research themes and their trends in an efficient time-frame. Some key conclusions of this research are as follows:

- The number of Twitter publications has been increased significantly since 2006 and is expected to grow in the following years.
- Sentiment analysis, social network analysis, big data mining, topic modeling, and content analysis were the most discussed topics.

- There were more research activities in application and methodology related topics than the ones in technology related topics.
- Most of the topics with meaningful trend ($P < 0.05$) had an increasing trend ($Slope > 0$).
- While different research approaches were used, supervised and unsupervised computational methods were discussed more than traditional research methods.
- Different data types (structured and unstructured) and data scales (small, medium, and large size) have been studied in the literature.
- Twitter was used as not only a data source but also a facilitator for hiring research participants.
- Twitter has been studied by researchers in formal, natural, and social sciences.
- Twitter-based research is a growing field recognized for population-level data.
- The collaboration of formal, natural, and social sciences on investigating Twitter data shows that Twitter-based research is an interdisciplinary field.
- Compared to traditional literature review methods, the methods of this paper are systematic, fast, and efficient.

While this survey paints a picture to illustrate where Twitter-related studies have been during the past years and might go in the following years, it has some restrictions. First, our data collection was limited to three databases. Second, this study did not consider other social media platforms (e.g., Facebook) to compare the research of multiple platforms. Third, this research focused on manuscripts published in English. Fourth, while this research is a high-level analysis providing a good overview of major topics and trends, it does not capture a full meaning of our data and sub-categories of the detected topics. Considering these limitations, future directions may consider other databases such as Scopus (https://www.scopus.com), multiple social media platforms, non-English-language publications, and investigating each of the detected topics to find sub-topics.

REFERENCES

[1] M. Ahlgren. (2020). *40+ Twitter Statistics & Facts For 2020*. Accessed: Mar. 22, 2020. [Online]. Available: https://www.websitehostingrating.com/twitter-statistics/

[2] D. Arigo, S. Pagoto, L. Carter-Harris, S. E. Lillie, and C. Nebeker, "Using social media for health research: Methodological and ethical considerations for recruitment and intervention delivery," *Digit. Health*, vol. 4, Jan. 2018, Art. no. 205520761877175.

[3] S. M. Kywe, E.-P. Lim, and F. Zhu, "A survey of recommender systems in Twitter," in *Proc. Int. Conf. Social Informat.* Berlin, Germany: Springer, 2012, pp. 420–433.

[4] D. Gayo-Avello, "A meta-analysis of state-of-the-art electoral prediction from Twitter data," *Social Sci. Comput. Rev.*, vol. 31, no. 6, pp. 649–679, Dec. 2013.

[5] M. Verma, D. Divya, and S. Sofat, "Techniques to detect spammers in Twitter—A survey," *Int. J. Comput. Appl.*, vol. 85, no. 10, pp. 27–32, 2014.

[6] D. Gayo-Avello, "Political opinion," in *Twitter: A Digital Socioscope*, vol. 52. New York, NY, USA: Cambridge Univ. Press, 2015.

[7] H. Mao, "Socioeconomic indicators," in *Twitter: A Digital Socioscope*. New York, NY, USA: Cambridge Univ. Press, 2015, p. 75.

[8] D. Quercia, "Hyperlocal happiness from tweets," in *Twitter: A Digital Socioscope*. New York, NY, USA: Cambridge Univ. Press, 2015, p. 96.

[9] P. Kostkova, "Public health," in *Twitter: A Digital Socioscope*. New York, NY, USA: Cambridge Univ. Press, 2015, p. 96.

[10] B. Robinson, R. Power, and M. Cameron, "Disaster monitoring," in *Twitter: A Digital Socioscope*. New York, NY, USA: Cambridge Univ. Press, 2015, p. 131.

[11] D. Finfgeld-Connett, "Twitter and health science research," *Western J. Nursing Res.*, vol. 37, no. 10, pp. 1269–1283, Oct. 2015.

[12] D. A. Gruber, R. E. Smerek, M. C. Thomas-Hunt, and E. H. James, "The real-time power of Twitter: Crisis management and leadership in an age of social media," *Bus. Horizons*, vol. 58, no. 2, pp. 163–172, Mar. 2015.

[13] A. Jungherr, "Twitter use in election campaigns: A systematic literature review," *J. Inf. Technol. Politics*, vol. 13, no. 1, pp. 72–91, Jan. 2016.

[14] R. Buettner and K. Buettner, "A systematic literature review of Twitter research from a socio-political revolution perspective," in *Proc. 49th Hawaii Int. Conf. Syst. Sci. (HICSS)*, Jan. 2016, pp. 2206–2215.

[15] V. A. and S. S. Sonawane, "Sentiment analysis of Twitter data: A survey of techniques," *Int. J. Comput. Appl.*, vol. 139, no. 11, pp. 5–15, 2016.

[16] L. Sinnenberg, A. M. Buttenheim, K. Padrez, C. Mancheno, L. Ungar, and R. M. Merchant, "Twitter as a tool for health research: A systematic review," *Amer. J. Public Health*, vol. 107, no. 1, pp. 143–143, Jan. 2017.

[17] T. Wu, S. Wen, Y. Xiang, and W. Zhou, "Twitter spam detection: Survey of new approaches and comparative study," *Comput. Secur.*, vol. 76, pp. 265–284, Jul. 2018.

[18] M. Martínez-Rojas, M. D. C. Pardo-Ferreira, and J. C. Rubio-Romero, "Twitter as a tool for the management and analysis of emergency situations: A systematic literature review," *Int. J. Inf. Manage.*, vol. 43, pp. 196–208, Dec. 2018.

[19] A. Kumar and A. Jaiswal, "Systematic literature review of sentiment analysis on Twitter using soft computing techniques," *Concurrency Comput., Pract. Exper.*, vol. 32, no. 1, p. e5107, Jan. 2020.

[20] Y. Roth and R. Johnson. (2018). *New Developer Requirements to Protect Our Platform*. Accessed: Mar. 22, 2020. [Online]. Available: https://blog.twitter.com/developer/en_us/topics/tools/2018/new-developer-requirements-to-protect-our-platform.html

[21] A. Hino and R. A. Fahey, "Representing the Twittersphere: Archiving a representative sample of Twitter data under resource constraints," *Int. J. Inf. Manage.*, vol. 48, pp. 175–184, Oct. 2019.

[22] A. Sarker, A. DeRoos, and J. Perrone, "Mining social media for prescription medication abuse monitoring: A review and proposal for a data-centric framework," *J. Amer. Med. Inform. Assoc.*, vol. 27, no. 2, pp. 315–329, Feb. 2020.

[23] A. M. Aladwani, "Facilitators, characteristics, and impacts of Twitter use: Theoretical analysis and empirical illustration," *Int. J. Inf. Manage.*, vol. 35, no. 1, pp. 15–25, Feb. 2015.

[24] Y. Mejova, I. Weber, and M. W. Macy, *Twitter: A Digital Socioscope*. Cambridge, U.K.: Cambridge Univ. Press, 2015.

[25] P. Grover, A. K. Kar, and G. Davies, "'Technology enabled Health'– insights from Twitter analytics with a socio-technical perspective," *Int. J. Inf. Manage.*, vol. 43, pp. 85–97, Dec. 2018.

[26] A. Karami, A. A. Dahl, G. Turner-McGrievy, H. Kharrazi, and G. Shaw, "Characterizing diabetes, diet, exercise, and obesity comments on Twitter," *Int. J. Inf. Manage.*, vol. 38, no. 1, pp. 1–6, Feb. 2018.

[27] T. D. Nascimento, M. F. DosSantos, T. Danciu, M. DeBoer, H. van Holsbeeck, S. R. Lucas, C. Aiello, L. Khatib, M. A. Bender, J.-K. Zubieta, and A. F. DaSilva, "Real-time sharing and expression of migraine headache suffering on Twitter: A cross-sectional infodemiology study," *J. Med. Internet Res.*, vol. 16, no. 4, p. e96, 2014.

[28] A. Karami, V. Shah, R. Vaezi, and A. Bansal, "Twitter speaks: A case of national disaster situational awareness," *J. Inf. Sci.*, Mar. 2019, Art. no. 016555151982862. [Online]. Available: https://journals.sagepub.com/doi/10.1177/0165551519828620

[29] A. Karami, L. S. Bennett, and X. He, "Mining public opinion about economic issues: Twitter and the U.S. presidential election," *Int. J. Strategic Decis. Sci.*, vol. 9, no. 1, pp. 18–28, Jan. 2018.

[30] C. B. Asmussen and C. Müller, "Smart literature review: A practical topic modelling approach to exploratory literature review," *J. Big Data*, vol. 6, no. 1, p. 93, Dec. 2019.

[31] C. R. Sugimoto, D. Li, T. G. Russell, S. C. Finlay, and Y. Ding, "The shifting sands of disciplinary development: Analyzing north American library and information science dissertations using latent Dirichlet allocation," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 62, no. 1, pp. 185–204, Jan. 2011.

[32] R. C. Boyer, W. T. Scherer, and M. C. Smith, "Trends over two decades of transportation research: A machine learning approach," *Transp. Res. Rec.*, vol. 2614, no. 1, pp. 1–9, Jan. 2017.

[33] S. Das, K. Dixon, X. Sun, A. Dutta, and M. Zupancich, "Trends in transportation research: Exploring content analysis in topics," *Transp. Res. Rec.*, vol. 2614, no. 1, pp. 27–38, Jan. 2017.

[34] K. K. Kapoor, K. Tamilmani, N. P. Rana, P. Patil, Y. K. Dwivedi, and S. Nerur, "Advances in social media research: Past, present and future," *Inf. Syst. Frontiers*, vol. 20, no. 3, pp. 531–558, Jun. 2018.

[35] D. M. Blei, "Probabilistic topic models," *Commun. ACM*, vol. 55, no. 4, pp. 77–84, Apr. 2012.

[36] A. J. van Altena, P. D. Moerland, A. H. Zwinderman, and S. D. Olabarriaga, "Understanding big data themes from scientific biomedical literature through topic modeling," *J. Big Data*, vol. 3, no. 1, p. 23, Dec. 2016.

[37] A. Karami, M. Ghasemi, S. Sen, M. F. Moraes, and V. Shah, "Exploring diseases and syndromes in neurology case reports from 1955 to 2017 with text mining," *Comput. Biol. Med.*, vol. 109, pp. 322–332, Jun. 2019.

[38] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.

[39] A. Karami and A. Gangopadhyay, "FFTM: A fuzzy feature transformation method for medical documents," in *Proc. BioNLP*, 2014.

[40] A. Karami, A. Gangopadhyay, B. Zhou, and H. Kharrazi, "A fuzzy approach model for uncovering hidden latent semantic structure in medical text collections," in *Proc. iConference*, 2015, pp. 1–5.

[41] Y. Lu, Q. Mei, and C. Zhai, "Investigating task performance of probabilistic topic models: An empirical study of PLSA and LDA," *Inf. Retr.*, vol. 14, no. 2, pp. 178–203, Apr. 2011.

[42] L. Hong and B. D. Davison, "Empirical study of topic modeling in Twitter," in *Proc. 1st Workshop Social Media Analytics (SOMA)*, 2010, pp. 80–88.

[43] J. D. Mcauliffe and D. M. Blei, "Supervised topic models," in *Proc. Adv. neural Inf. Process. Syst.*, 2008, pp. 121–128.

[44] A. Karami, A. Gangopadhyay, B. Zhou, and H. Kharrazi, "Fuzzy approach topic discovery in health and medical corpora," *Int. J. Fuzzy Syst.*, vol. 20, no. 4, pp. 1334–1345, Apr. 2018.

[45] F. Webb, A. Karami, and V. L. Kitzie, "Characterizing diseases and disorders in gay Users' tweets," in *Proc. Southern Assoc. Inf. Syst. (SAIS)*, 2018, pp. 1–6.

[46] A. Karami, F. Webb, and V. L. Kitzie, "Characterizing transgender health issues in Twitter," *Proc. Assoc. Inf. Sci. Technol.*, vol. 55, no. 1, pp. 207–215, Jan. 2018.

[47] A. Karami and G. Shaw, "An exploratory study of (#) exercise in the Twittersphere," in *Proc. iConference*, 2019, pp. 1–6.

[48] L. Hagen, "Content analysis of e-petitions with topic modeling: How to train and evaluate LDA models?" *Inf. Process. Manage.*, vol. 54, no. 6, pp. 1292–1307, Nov. 2018.

[49] A. Karami and E. Aida, "Political popularity analysis in social media," in *Proc. iConference*, Washington, DC, USA, 2019, pp. 456–465.

[50] A. Karami, S. C. Swan, C. N. White, and K. Ford, "Hidden in plain sight for too long: Using text mining techniques to shine a light on workplace sexism and sexual harassment.," *Psychol. Violence*, Jun. 2019. [Online]. Available: https://psycnet.apa.org/doiLanding?doi=10.1037%2Fvio0000239

[51] A. Karami, C. N. White, K. Ford, S. Swan, and M. Y. Spinel, "Unwanted advances in higher education: Uncovering sexual harassment experiences in academia with text mining," *Inf. Process. Manage.*, vol. 57, no. 2, Mar. 2020, Art. no. 102167.

[52] A. Karami and N. M. Pendergraft, "Computational analysis of insurance complaints: Geico case study," in *Proc. Int. Conf. Social Comput., Behav.-Cultural Modeling, Predict. Behav. Represent. Modeling. Simul.*, 2018, pp. 1–8.

[53] M. Collins and A. Karami, "Social media analysis for organizations: US northeastern public and state libraries case study," in *Proc. Southern Assoc. Inf. Syst. (SAIS)*. Atlanta, Georgia, 2018, pp. 1–6.

[54] A. Karami and M. Collins, "What do the US west coast public libraries post on Twitter?" *Proc. Assoc. Inf. Sci. Technol.*, vol. 55, no. 1, pp. 216–225, Jan. 2018.

[55] A. Karami and L. Zhou, "Exploiting latent content based features for the detection of static SMS spams," *Proc. Amer. Soc. Inf. Sci. Technol.*, vol. 51, no. 1, pp. 1–4, 2014.

[56] L. Sun and Y. Yin, "Discovering themes and trends in transportation research using topic modeling," *Transp. Res. C, Emerg. Technol.*, vol. 77, pp. 49–66, Apr. 2017.

[57] J. Hemsley, I. Erickson, M. H. Jarrahi, and A. Karami, "Digital nomads, coworking, and other expressions of mobile work on Twitter," *1st Monday*, vol. 25, Feb. 2020. [Online]. Available: https://firstmonday.org/ojs/index.php/fm/article/view/10246

[58] Y. Zhu, M.-H. Kim, S. Banerjee, J. Deferio, G. S. Alexopoulos, and J. Pathak, "Understanding the research landscape of major depressive disorder via literature mining: An entity-level analysis of PubMed data from 1948 to 2017," *JAMIA Open*, vol. 1, no. 1, pp. 115–121, Jul. 2018.

[59] G. Shin, M. H. Jarrahi, Y. Fei, A. Karami, N. Gafinowitz, A. Byun, and X. Lu, "Wearable activity trackers, accuracy, adoption, acceptance and health impact: A systematic literature review," *J. Biomed. Informat.*, vol. 93, May 2019, Art. no. 103153.

[60] Q. Chen, N. Ai, J. Liao, X. Shao, Y. Liu, and X. Fan, "Revealing topics and their evolution in biomedical literature using bio-DTM: A case study of ginseng," *Chin. Med.*, vol. 12, no. 1, p. 27, Dec. 2017.

[61] R. Mousavi and B. Gu, "The impact of Twitter adoption on decision making in politics," in *Proc. 48th Hawaii Int. Conf. Syst. Sci.*, Jan. 2015, pp. 4854–4863.

[62] N. Ernst, S. Engesser, F. Büchel, S. Blassnig, and F. Esser, "Extreme parties and populism: An analysis of facebook and Twitter across six countries," *Inf., Commun. Soc.*, vol. 20, no. 9, pp. 1347–1364, Sep. 2017.

[63] J. Yang and Y. M. Kim, "Equalization or normalization? Voter–candidate engagement on Twitter in the 2010 U.S. midterm elections," *J. Inf. Technol. Politics*, vol. 14, no. 3, pp. 232–247, Jul. 2017.

[64] A. Russell, "U.S. senators on Twitter: Asymmetric party rhetoric in 140 characters," *Amer. Politics Res.*, vol. 46, no. 4, pp. 695–723, Jul. 2018.

[65] A. Bruns and B. Moon, "Social media in Australian federal elections: Comparing the 2013 and 2016 campaigns," *Journalism Mass Commun. Quart.*, vol. 95, no. 2, pp. 425–448, Jun. 2018.

[66] G. K. Zipf, *Human Behavior and the Principle of Least Effort*. Boston, MA, USA: Addison-Wesley, 1949.

[67] R. Deveaud, E. SanJuan, and P. Bellot, "Accurate and effective latent concept modeling for ad hoc information retrieval," *Document numérique*, vol. 17, no. 1, pp. 61–84, Apr. 2014.

[68] A. K. McCallum. (2002). *MALLET: A Machine Learning for Language Toolkit*. [Online]. Available: http://mallet.cs.umass.edu/topics.php

[69] B. Wang and J. Zhuang, "Rumor response, debunking response, and decision makings of misinformed Twitter users during disasters," *Natural Hazards*, vol. 93, no. 3, pp. 1145–1162, Sep. 2018.

[70] J. Li, A. Vishwanath, and H. R. Rao, "Retweeting the Fukushima nuclear radiation disaster," *Commun. ACM*, vol. 57, no. 1, pp. 78–85, Jan. 2014.

[71] R. G. Rice and P. R. Spence, "Thor visits lexington: Exploration of the knowledge-sharing gap and risk management learning in social media during multiple winter storms," *Comput. Hum. Behav.*, vol. 65, pp. 612–618, Dec. 2016.

[72] W. Liu, C.-H. Lai, and W. W. Xu, "Tweeting about emergency: A semantic network analysis of government organizations' social media messaging during hurricane harvey," *Public Relations Rev.*, vol. 44, no. 5, pp. 807–819, Dec. 2018.

[73] V. Grasso and A. Crisci, "Codified hashtags for weather warning on Twitter: An Italian case study," *PLoS Currents*, 2016. [Online]. Available: http://currents.plos.org/disasters/article/codified-hashtags-for-weather-warning-on-twitter-an-italian-case-study/

[74] M. D. P. Salas-Zárate, J. Medina-Moreira, K. Lagos-Ortiz, H. Luna-Aveiga, M. Á. Rodríguez-García, and R. Valencia-García, "Sentiment analysis on tweets about diabetes: An aspect-level approach," *Comput. Math. Methods Med.*, vol. 2017, pp. 1–9, Feb. 2017.

[75] F. Bravo-Marquez, E. Frank, and B. Pfahringer, "From opinion lexicons to sentiment classification of tweets and vice versa: A transfer learning approach," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell. (WI)*, Oct. 2016, pp. 145–152.

[76] M. Bouazizi and T. Ohtsuki, "Sentiment analysis: From binary to multiclass classification: A pattern-based approach for multi-class sentiment analysis in Twitter," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–6.

[77] F. Bravo-Marquez, E. Frank, and B. Pfahringer, "Building a Twitter opinion lexicon from automatically-annotated tweets," *Knowl.-Based Syst.*, vol. 108, pp. 65–78, Sep. 2016.

[78] Z. Jianqiang, "Combing semantic and prior polarity features for boosting Twitter sentiment analysis using ensemble learning," in *Proc. IEEE 1st Int. Conf. Data Sci. Cyberspace (DSC)*, Jun. 2016, pp. 709–714.

[79] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retr.*, vol. 2, nos. 1–2, pp. 1–135, 2008.

[80] A. Giachanou and F. Crestani, "Like it or not: A survey of Twitter sentiment analysis methods," *ACM Comput. Surveys*, vol. 49, no. 2, pp. 1–41, Nov. 2016.
[81] S. A. Alkhodair, B. C. M. Fung, O. Rahman, and P. C. K. Hung, "Improving interpretations of topic modeling in microblogs," *J. Assoc. Inf. Sci. Technol.*, vol. 69, no. 4, pp. 528–540, Apr. 2018.
[82] H.-J. Choi and C. H. Park, "Emerging topic detection in Twitter stream based on high utility pattern mining," *Expert Syst. Appl.*, vol. 115, pp. 27–36, Jan. 2019.
[83] P. K. Srijith, M. Hepple, K. Bontcheva, and D. Preotiuc-Pietro, "Sub-story detection in Twitter with hierarchical Dirichlet processes," *Inf. Process. Manage.*, vol. 53, no. 4, pp. 989–1003, Jul. 2017.
[84] M. Yang, Y. Liang, W. Zhao, W. Xu, J. Zhu, and Q. Qu, "Task-oriented keyphrase extraction from social media," *Multimedia Tools Appl.*, vol. 77, no. 3, pp. 3171–3187, Feb. 2018.
[85] Y. Zhang, W. Mao, and J. Lin, "Modeling topic evolution in social media short texts," in *Proc. IEEE Int. Conf. Big Knowl. (ICBK)*, Aug. 2017, pp. 315–319.
[86] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, "Indexing by latent semantic analysis," *J. Amer. Soc. Inf. Sci.*, vol. 41, no. 6, pp. 391–407, 1990.
[87] T. Hofmann, "Unsupervised learning by probabilistic latent semantic analysis," *Mach. Learn.*, vol. 42, no. 1, pp. 177–196, Jan. 2001.
[88] J. Boyd-Graber, Y. Hu, and D. Mimno, *Applications of Topic Models* (Foundations and Trends in Information Retrieval), vol. 11, nos. 2–3, D. Oard, Ed. Boston, MA, USA: Now, 2017. [Online]. Available: http://www.nowpublishers.com/article/Details/INR-030
[89] A. Karami, A. Gangopadhyay, B. Zhou, and H. Kharrazi, "FLATM: A fuzzy logic approach topic model for medical documents," in *Proc. Annu. Conf. North Amer. Fuzzy Inf. Process. Soc. (NAFIPS), Held Jointly 5th World Conf. Soft Comput. (WConSC)*, Aug. 2015, pp. 1–6.
[90] A. Karami, "Taming wild high dimensional text data with a fuzzy lash," in *Proc. IEEE Int. Conf. Data Mining Workshops (ICDMW)*, Nov. 2017, pp. 518–522.
[91] A. Karami, "Application of fuzzy clustering for text data dimensionality reduction," *Int. J. Knowl. Eng. Data Mining*, vol. 6, no. 1, pp. 289–306, 2019.
[92] M. Steyvers and T. Griffiths, "Probabilistic topic models," in *Handbook of Latent Semantic Analysis*, vol. 427, no. 7. New York, NY, USA: Routledge, 2007, pp. 424–440.
[93] S. P. Crain, K. Zhou, S.-H. Yang, and H. Zha, "Dimensionality reduction and topic modeling: From latent semantic indexing to latent Dirichlet allocation and beyond," in *Mining Text Data*. Boston, MA, USA: Springer, 2012, pp. 129–161.
[94] A. Karami, "Fuzzy topic modeling for medical corpora," Ph.D. dissertation, Univ. Maryland, College Park, MD, USA, 2015.
[95] O. Kounadi, A. Ristea, M. Leitner, and C. Langford, "Population at risk: Using areal interpolation and Twitter messages to create population models for burglaries and robberies," *Cartography Geographic Inf. Sci.*, vol. 45, no. 3, pp. 205–220, May 2018.
[96] X. Zhou and L. Zhang, "Crowdsourcing functions of the living city from Twitter and foursquare data," *Cartography Geographic Inf. Sci.*, vol. 43, no. 5, pp. 393–404, Oct. 2016.
[97] N. N. Patel, F. R. Stevens, Z. Huang, A. E. Gaughan, I. Elyazar, and A. J. Tatem, "Improving large area population mapping using geotweet densities," *Trans. GIS*, vol. 21, no. 2, pp. 317–331, Apr. 2017.
[98] M. H. Salas-Olmedo and C. R. Quezada, "The use of public spaces in a medium-sized city: From Twitter data to mobility patterns," *J. Maps*, vol. 13, no. 1, pp. 40–45, Jan. 2017.
[99] J. Bao, P. Liu, H. Yu, and C. Xu, "Incorporating Twitter-based human activity information in spatial analysis of crashes in urban areas," *Accident Anal. Prevention*, vol. 106, pp. 358–369, Sep. 2017.
[100] C. I. J. Nykiforuk and L. M. Flaman, "Geographic information systems (GIS) for health promotion and public health: A review," *Health Promotion Pract.*, vol. 12, no. 1, pp. 63–73, Jan. 2011.
[101] R. R. Vatsavai, A. Ganguly, V. Chandola, A. Stefanidis, S. Klasky, and S. Shekhar, "Spatiotemporal data mining in the era of big spatial data: Algorithms and applications," in *Proc. 1st ACM SIGSPATIAL Int. Workshop Anal. Big Geospatial Data (BigSpatial)*, 2012, pp. 1–10.
[102] D. Darmofal, *Spatial Analysis for the Social Sciences*. Cambridge, U.K.: Cambridge Univ. Press, 2015.
[103] D. Li, S. Wang, and D. Li, *Spatial Data Mining*. Berlin, Germany: Springer, 2015.
[104] L. Crearie, "Millennial and centennial student interactions with technology," *GSTF J. Comput. (JoC)*, vol. 6, no. 1, pp. 1–10, 2018.
[105] M. Stephens, "Messaging in a 2.0 world: Twitter & SMS," *Library Technol. Rep.*, vol. 43, no. 5, pp. 62–66, 2009.
[106] R. R. Pai and S. Alathur, "Assessing mobile health applications with Twitter analytics," *Int. J. Med. Informat.*, vol. 113, pp. 72–84, May 2018.
[107] M. S. Karpeh and S. Bryczkowski, "Digital communications and social media use in surgery: How to maximize communication in the digital age," *Innov. Surgical Sci.*, vol. 2, no. 3, pp. 153–157, Sep. 2017.
[108] L. Labacher and C. Mitchell, "Talk or text to tell? How young adults in Canada and South Africa prefer to receive STI results, counseling, and treatment updates in a wireless world," *J. Health Commun.*, vol. 18, no. 12, pp. 1465–1476, Dec. 2013.
[109] B. Thoma, M. Gottlieb, M. Boysen-Osborn, A. King, A. Quinn, S. Krzyzaniak, N. Pineda, L. M. Yarris, and T. Chan, "Curated collections for educators: Five key papers about program evaluation," *Cureus*, vol. 9, no. 5, p. e1224, May 2017.
[110] N. S. Trueger, H. Murray, S. Kobner, and M. Lin, "Global emergency medicine journal club: A social media discussion about the outpatient management of patients with spontaneous pneumothorax by using pigtail catheters," *Ann. Emergency Med.*, vol. 66, no. 4, pp. 409–416, Oct. 2015.
[111] M. J. Roberts, M. Perera, N. Lawrentschuk, D. Romanic, N. Papa, and D. Bolton, "Globalization of continuing professional development by journal clubs via microblogging: A systematic review," *J. Med. Internet Res.*, vol. 17, no. 4, p. e103, 2015.
[112] D. Diller and L. M. Yarris, "A descriptive analysis of the use of Twitter by emergency medicine residency programs," *J. Graduate Med. Edu.*, vol. 10, no. 1, pp. 51–55, Feb. 2018.
[113] D. Cohen, T. C. Allen, S. Balci, P. T. Cagle, J. Teruya-Feldstein, S. W. Fine, D. D. Gondim, J. L. Hunt, J. Jacob, and K. Jewett, "#InSituPathologists: How the #USCAP2015 meeting went viral on Twitter and founded the social media movement for the United States and Canadian Academy of Pathology," *Modern Pathol.*, vol. 30, no. 2, pp. 160–168, Feb. 2017.
[114] B. G. Johnson, "Speech, harm, and the duties of digital intermediaries: Conceptualizing platform ethics," *J. Media Ethics*, vol. 32, no. 1, pp. 16–27, Jan. 2017.
[115] A. Mills, "The law applicable to cross-border defamation on social media: Whose law governs free speech in 'Facebookistan?'" *J. Media Law*, vol. 7, no. 1, pp. 1–35, Jan. 2015.
[116] K. Barker, "Virtual spaces and virtual layers–governing the ungovernable?" *Inf. Commun. Technol. Law*, vol. 25, no. 1, pp. 62–70, Jan. 2016.
[117] B. G. Johnson, "Networked communication and the reprise of tolerance theory: Civic education for extreme speech and private governance online," *1st Amendment Stud.*, vol. 50, no. 1, pp. 14–31, Jan. 2016.
[118] J. Dean, "To tweet or not to tweet: Twitter, 'broadcasting,' and federal rule of criminal procedure 53," *Univ. Cincinnati Law Rev.*, vol. 79, no. 2, p. 11, 2011.
[119] A. Rudat and J. Buder, "Making retweeting social: The influence of content and context information on sharing news in Twitter," *Comput. Hum. Behav.*, vol. 46, pp. 75–84, May 2015.
[120] R. Agrifoglio, S. Black, C. Metallo, and M. Ferrara, "Extrinsic versus intrinsic motivation in continued twitter usage," *J. Comput. Inf. Syst.*, vol. 53, no. 1, pp. 33–41, 2012.
[121] W. Shu, "Continual use of microblogs," *Behaviour Inf. Technol.*, vol. 33, no. 7, pp. 666–677, Jul. 2014.
[122] S. Kruikemeier, G. van Noort, R. Vliegenthart, and C. H. de Vreese, "The relationship between online campaigning and political involvement," *Online Inf. Rev.*, vol. 40, no. 5, pp. 673–694, Sep. 2016.
[123] E.-J. Lee and S. Y. Oh, "Seek and you shall find? How need for orientation moderates knowledge gain from Twitter use," *J. Commun.*, vol. 63, no. 4, pp. 745–765, Aug. 2013.
[124] S. Harlow and T. J. Johnson, "The arab springl overthrowing the protest paradigm? how the New York times, global voices and Twitter covered the Egyptian revolution," *Int. J. Commun.*, vol. 5, p. 16, 2011.
[125] S. Lowrance, "Was the revolution tweeted? Social media and the jasmine revolution in tunisia," *Dig. Middle East Stud.*, vol. 25, no. 1, pp. 155–176, Mar. 2016.
[126] M. Tremayne, "Anatomy of protest in the digital era: A network analysis of *Twitter* and occupy wall street," *Social Movement Stud.*, vol. 13, no. 1, pp. 110–126, Jan. 2014.
[127] M. Knight, "Journalism as usual: The use of social media as a news-gathering tool in the coverage of the iranian elections in 2009," *J. Media Pract.*, vol. 13, no. 1, pp. 61–74, Jan. 2012.

[128] S. Harlow, R. Salaverría, D. K. Kilgo, and V. García-Perdomo, "Protest paradigm in multimedia: Social media sharing of coverage about the crime of Ayotzinapa, Mexico," *J. Commun.*, vol. 67, no. 3, pp. 328–349, Jun. 2017.

[129] J. W. Treem, S. L. Dailey, C. S. Pierce, and D. Biffl, "What we are talking about when we talk about social media: A framework for study," *Sociol. Compass*, vol. 10, no. 9, pp. 768–784, Sep. 2016.

[130] K. Weller, "Trying to understand social media users and usage: The forgotten features of social media platforms," *Online Inf. Rev.*, vol. 40, no. 2, pp. 256–264, Apr. 2016.

[131] H. A. Howard, S. Huber, L. V. Carter, and E. A. Moore, "Academic libraries on social media: Finding the students and the information they want," *Inf. Technol. Libraries*, vol. 37, no. 1, pp. 8–18, 2018.

[132] E. R. Ranschaert, P. M. A. Van Ooijen, G. B. McGinty, and P. M. Parizel, "Radiologists' usage of social media: Results of the RANSOM survey," *J. Digit. Imag.*, vol. 29, no. 4, pp. 443–449, Aug. 2016.

[133] S. Kwayu, B. Lal, and M. Abubakre, "Enhancing organisational competitiveness via social media—A strategy as practice perspective," *Inf. Syst. Frontiers*, vol. 20, no. 3, pp. 439–456, Jun. 2018.

[134] M. E. Waring, K. Baker, A. Peluso, C. N. May, and S. L. Pagoto, "Content analysis of Twitter chatter about indoor tanning," *Transl. Behav. Med.*, vol. 9, no. 1, pp. 41–47, Jan. 2019.

[135] A. J. Calvin, A. Bellmore, J.-M. Xu, and X. Zhu, "#bully: Uses of hashtags in posts about bullying on Twitter," *J. School Violence*, vol. 14, no. 1, pp. 133–153, Jan. 2015.

[136] L. Han, M. Rodriguez, and L. Han, "Tweeting PP: An analysis of the 2015–2016 planned parenthood controversy on Twitter," *Contraception*, vol. 94, no. 4, pp. 388–394, Oct. 2016.

[137] J. Salem, H. Borgmann, M. Bultitude, H.-M. Fritsche, A. Haferkamp, A. Heidenreich, A. Miernik, A. Neisius, T. Knoll, C. Thomas, and I. Tsaur, "Online discussion on #KidneyStones: A longitudinal assessment of activity, users and content," *PLoS ONE*, vol. 11, no. 8, 2016, Art. no. e0160863.

[138] L. Thompson, F. P. Rivara, and J. M. Whitehill, "Prevalence of marijuana-related traffic on Twitter, 2012–2013: A content analysis," *Cyberpsychology, Behav., Social Netw.*, vol. 18, no. 6, pp. 311–319, Jun. 2015.

[139] H.-F. Hsieh and S. E. Shannon, "Three approaches to qualitative content analysis," *Qualitative Health Res.*, vol. 15, no. 9, pp. 1277–1288, Nov. 2005.

[140] V. J. Duriau, R. K. Reger, and M. D. Pfarrer, "A content analysis of the content analysis literature in organization studies: Research themes, data sources, and methodological refinements," *Organizational Res. Methods*, vol. 10, no. 1, pp. 5–34, Jan. 2007.

[141] M. Vaismoradi, H. Turunen, and T. Bondas, "Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study," *Nursing Health Sci.*, vol. 15, no. 3, pp. 398–405, Sep. 2013.

[142] S. Lacy, B. R. Watson, D. Riffe, and J. Lovejoy, "Issues and best practices in content analysis," *Journalism Mass Commun. Quart.*, vol. 92, no. 4, pp. 791–811, Dec. 2015.

[143] K. Krippendorff, *Content Analysis: An Introduction to its Methodology*. Newbury Park, CA, USA: Sage, 2018.

[144] L. K. Nelson, D. Burk, M. Knudsen, and L. McCall, "The future of coding: A comparison of hand-coding and three types of computer-assisted text analysis methods," *Sociol. Methods Res.*, May 2018, Art. no. 004912411876911. [Online]. Available: https://journals.sagepub.com/doi/10.1177/0049124118769114

[145] R. Chunara, J. R. Andrews, and J. S. Brownstein, "Social and news media enable estimation of epidemiological patterns early in the 2010 haitian cholera outbreak," *Amer. J. Tropical Med. Hygiene*, vol. 86, no. 1, pp. 39–45, Jan. 2012.

[146] S.-Y. Shin, D.-W. Seo, J. An, H. Kwak, S.-H. Kim, J. Gwack, and M.-W. Jo, "High correlation of middle east respiratory syndrome spread with Google search and Twitter trends in korea," *Sci. Rep.*, vol. 6, no. 1, p. 32920, Dec. 2016.

[147] N. Mahroum, M. Adawi, K. Sharif, R. Waknin, H. Mahagna, B. Bisharat, M. Mahamid, A. Abu-Much, H. Amital, N. L. Bragazzi, and A. Watad, "Public reaction to Chikungunya outbreaks in Italy—Insights from an extensive novel data streams-based structural equation modeling analysis," *PLoS ONE*, vol. 13, no. 5, 2018, Art. no. e0197337.

[148] S. F. McGough, J. S. Brownstein, J. B. Hawkins, and M. Santillana, "Forecasting Zika incidence in the 2016 latin america outbreak combining traditional disease surveillance with search, social media, and news report data," *PLOS Neglected Tropical Diseases*, vol. 11, no. 1, 2017, Art. no. e0005295.

[149] M. S. Deiner, T. M. Lietman, S. D. McLeod, J. Chodosh, and T. C. Porco, "Surveillance tools emerging from search engines and social media data for determining eye disease patterns," *JAMA Ophthalmol.*, vol. 134, no. 9, pp. 1024–1030, Sep. 2016.

[150] K. Aillerie and S. McNicol, "Are social networking sites information sources? Informational purposes of high-school students in using SNSs," *J. Librarianship Inf. Sci.*, vol. 50, no. 1, pp. 103–114, Mar. 2018.

[151] A. S. Chaudhry and H. Al-Ansari, "Information behavior of financial professionals in the arabian gulf region," *Int. Inf. Library Rev.*, vol. 48, no. 4, pp. 235–248, Oct. 2016.

[152] P. G. Tadasad, D. K, and S. Patil, "Use of online social networking services in university libraries: A study of university libraries of karnataka, india," *DESIDOC J. Library Inf. Technol.*, vol. 37, no. 4, pp. 249–258, 2017.

[153] J. Pistolis, S. Zimeras, K. Chardalias, Z. Roupa, G. Fildisis, and A. Diomidous, "Investigation of the impact of extracting and exchanging health information by using Internet and social networks," *Acta Inf. Medica*, vol. 24, no. 3, p. 197, 2016.

[154] S. J. McGee, "To friend or not to friend: Is that the question for healthcare?" *Amer. J. Bioethics*, vol. 11, no. 8, pp. 2–5, Aug. 2011.

[155] A. H. J. Magnan-Park, "Leukocentric hollywood: Whitewashing, alohagate and the dawn of hollywood with chinese characteristics," *Asian Cinema*, vol. 29, no. 1, pp. 133–162, Apr. 2018.

[156] S. Olive, "Defining the BBC shakespeare unlocked season 'in festival terms,'" *J. Adaptation Film Perform.*, vol. 10, no. 2, pp. 127–152, Jul. 2017.

[157] G. Fuller, "Shane Warne versus hoon cyclists: Affect and celebrity in a new media event," *Continuum*, vol. 31, no. 2, pp. 296–306, Mar. 2017.

[158] R. Coche, "Promoting women's soccer through social media: How the US federation used Twitter for the 2011 world cup," *Soccer Soc.*, vol. 17, no. 1, pp. 90–108, Jan. 2016.

[159] K. Mendes, J. Ringrose, and J. Keller, "#MeToo and the promise and pitfalls of challenging rape culture through digital feminist activism," *Eur. J. Women's Stud.*, vol. 25, no. 2, pp. 236–246, May 2018.

[160] D. Sehgal and A. K. Agarwal, "Sentiment analysis of big data applications using Twitter data with the help of Hadoop framework," in *Proc. Int. Conf. Syst. Modeling Advancement Res. Trends (SMART)*, 2016, pp. 251–255.

[161] L. G. S. Selvan and T.-S. Moh, "A framework for fast-feedback opinion mining on Twitter data streams," in *Proc. Int. Conf. Collaboration Technol. Syst. (CTS)*, Jun. 2015, pp. 314–318.

[162] M. B. Kraiem, J. Feki, K. Khrouf, F. Ravat, and O. Teste, "Modeling and OLAPing social media: The case of Twitter," *Social Netw. Anal. Mining*, vol. 5, no. 1, p. 47, Dec. 2015.

[163] I. Moise, E. Gaere, R. Merz, S. Koch, and E. Pournaras, "Tracking language mobility in the Twitter landscape," in *Proc. IEEE 16th Int. Conf. Data Mining Workshops (ICDMW)*, Dec. 2016, pp. 663–670.

[164] A. Poorthuis and M. Zook, "Small stories in big data: Gaining insights from large spatial point pattern datasets," *Cityscape*, vol. 17, no. 1, pp. 151–160, 2015.

[165] G. Barbier and H. Liu, "Data mining in social media," in *Social Network Data Analytics*. Boston, MA, USA: Springer, 2011, pp. 327–352.

[166] J. Han, J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*. Amsterdam, The Netherlands: Elsevier, 2011.

[167] C. M. Biship, *Pattern Recognition and Machine Learning* (Information Science and Statistics). New York, NY, USA: Springer, 2011.

[168] C. C. Aggarwal and C. Zhai, *Mining Text Data*. Boston, MA, USA: Springer, 2012.

[169] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*. San Mateo, CA, USA: Morgan Kaufmann, 2016.

[170] C.-K. Shieh, S.-W. Huang, L.-D. Sun, M.-F. Tsai, and N. Chilamkurti, "A topology-based scaling mechanism for apache storm," *Int. J. Netw. Manage.*, vol. 27, no. 3, p. e1933, May 2017.

[171] L. Jiao, J. Li, T. Xu, W. Du, and X. Fu, "Optimizing cost for online social networks on geo-distributed clouds," *IEEE/ACM Trans. Netw.*, vol. 24, no. 1, pp. 99–112, Feb. 2016.

[172] S.-W. Jun, M. Liu, S. Lee, J. Hicks, J. Ankcorn, M. King, and S. Xu, "Bluedbm: Distributed flash storage for big data analytics," *ACM Trans. Comput. Syst.*, vol. 34, no. 3, p. 7, 2016.

[173] G. P. Tran, J. P. Walters, and S. Crago, ''Reducing tail latencies while improving resiliency to timing errors for stream processing workloads,'' in *Proc. IEEE/ACM 11th Int. Conf. Utility Cloud Comput. (UCC)*, Dec. 2018, pp. 194–203.
[174] J. Huang, X. Ouyang, J. Jose, M. Wasi-ur-Rahman, H. Wang, M. Luo, H. Subramoni, C. Murthy, and D. K. Panda, ''High-performance design of HBase with RDMA over InfiniBand,'' in *Proc. IEEE 26th Int. Parallel Distrib. Process. Symp.*, May 2012, pp. 774–785.
[175] W. Wei, Y. Mao, and B. Wang, ''Twitter volume spikes and stock options pricing,'' *Comput. Commun.*, vol. 73, pp. 271–281, Jan. 2016.
[176] A. Tafti, R. Zotti, and W. Jank, ''Real-time diffusion of information on Twitter and the financial markets,'' *PLoS ONE*, vol. 11, no. 8, 2016, Art. no. e0159226.
[177] W. Zhang, X. Li, D. Shen, and A. Teglio, ''Daily happiness and stock returns: Some international evidence,'' *Phys. A, Stat. Mech. Appl.*, vol. 460, pp. 201–209, Oct. 2016.
[178] G. Ranco, D. Aleksovski, G. Caldarelli, M. Grčar, and I. Mozetič, ''The effects of Twitter sentiment on stock price returns,'' *PLoS ONE*, vol. 10, no. 9, 2015, Art. no. e0138441.
[179] J. C. Reboredo and A. Ugolini, ''The impact of Twitter sentiment on renewable energy stocks,'' *Energy Econ.*, vol. 76, pp. 153–169, Oct. 2018.
[180] H. Wu, V. Sorathia, and V. K. Prasanna, ''Predict whom one will follow: Followee recommendation in microblogs,'' in *Proc. Int. Conf. Social Informat.*, Dec. 2012, pp. 260–264.
[181] F. Zarrinkalam, M. Kahani, and E. Bagheri, ''Mining user interests over active topics on social networks,'' *Inf. Process. Manage.*, vol. 54, no. 2, pp. 339–357, Mar. 2018.
[182] Y.-D. Seo, Y.-G. Kim, E. Lee, and D.-K. Baik, ''Personalized recommender system based on friendship strength in social network services,'' *Expert Syst. Appl.*, vol. 69, pp. 135–148, Mar. 2017.
[183] N. Jonnalagedda, S. Gauch, K. Labille, and S. Alfarhood, ''Incorporating popularity in a personalized news recommender system,'' *PeerJ Comput. Sci.*, vol. 2, p. e63, Jun. 2016.
[184] S. Deng, D. Wang, X. Li, and G. Xu, ''Exploring user emotion in microblogs for music recommendation,'' *Expert Syst. Appl.*, vol. 42, no. 23, pp. 9284–9293, Dec. 2015.
[185] S. K. Papworth, T. P. L. Nghiem, D. Chimalakonda, M. R. C. Posa, L. S. Wijedasa, D. Bickford, and L. R. Carrasco, ''Quantifying the role of online news in linking conservation research to Facebook and Twitter,'' *Conservation Biol.*, vol. 29, no. 3, pp. 825–833, Jun. 2015.
[186] C. Meschede and T. Siebenlist, ''Cross-metric compatability and inconsistencies of altmetrics,'' *Scientometrics*, vol. 115, no. 1, pp. 283–297, Apr. 2018.
[187] B. Hammarfelt, ''Using altmetrics for assessing research impact in the humanities,'' *Scientometrics*, vol. 101, no. 2, pp. 1419–1430, Nov. 2014.
[188] K. Delli, C. Livas, F. K. L. Spijkervet, and A. Vissink, ''Measuring the social impact of dental research: An insight into the most influential articles on the Web,'' *Oral Diseases*, vol. 23, no. 8, pp. 1155–1161, Nov. 2017.
[189] A. Amath, K. Ambacher, J. J. Leddy, T. J. Wood, and C. J. Ramnanan, ''Comparing alternative and traditional dissemination metrics in medical education,'' *Med. Edu.*, vol. 51, no. 9, pp. 935–941, Sep. 2017.
[190] D. Ponte and J. Simon, ''Scholarly communication 2.0: Exploring Researchers' opinions on Web 2.0 for scientific knowledge creation, evaluation and dissemination,'' *Serials Rev.*, vol. 37, no. 3, pp. 149–156, Sep. 2011.
[191] G. Eysenbach, ''Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact,'' *J. Med. Internet Res.*, vol. 13, no. 4, p. e123, 2011.
[192] R. Kwok, ''Research impact: Altmetrics make their mark,'' *Nature*, vol. 500, no. 7463, pp. 491–493, Aug. 2013.
[193] H. Piwowar, ''Value all research products,'' *Nature*, vol. 493, no. 7431, pp. 159–159, Jan. 2013.
[194] J. Chavda and A. Patel, ''Measuring research impact: Bibliometrics, social media, altmetrics, and the BJGP,'' *Brit. J. Gen. Pract.*, vol. 66, no. 642, pp. e59–e61, Jan. 2016.
[195] M. Mehrazar, C. C. Kling, S. Lemke, A. Mazarakis, and I. Peters, ''Can we count on social media metrics?: First insights into the active scholarly use of social media,'' in *Proc. 10th ACM Conf. Web Sci. (WebSci)*, 2018, pp. 215–219.
[196] M. El Tantawi, E. Bakhurji, A. Al-Ansari, A. AlSubaie, H. A. Al Subaie, and A. Alali, ''Indicators of adolescents' preference to receive oral health information using social media,'' *Acta Odontologica Scandinavica*, vol. 77, no. 3, pp. 213–218, Apr. 2019.
[197] K. Kircaburun, ''Effects of gender and personality differences on Twitter addiction among turkish undergraduates,'' *J. Edu. Pract.*, vol. 7, no. 24, pp. 33–42, 2016.
[198] J. C. Levenson, A. Shensa, J. E. Sidani, J. B. Colditz, and B. A. Primack, ''The association between social media use and sleep disturbance among young adults,'' *Preventive Med.*, vol. 85, pp. 36–41, Apr. 2016.
[199] J. E. Sidani, A. Shensa, B. Hoffman, J. Hanmer, and B. A. Primack, ''The association between social media use and eating concerns among US young adults,'' *J. Acad. Nutrition Dietetics*, vol. 116, no. 9, pp. 1465–1472, Sep. 2016.
[200] J. Foster, C. Martin, and S. Patel, ''The initial measurement structure of the home drinking assessment scale (HDAS),'' *Drugs, Edu., Prevention Policy*, vol. 22, no. 5, pp. 410–415, Sep. 2015.
[201] P. J. Lavrakas, *Encyclopedia of Survey Research Methods*. Newbury Park, CA, USA: Sage, 2008.
[202] R. M. Groves, F. J. Fowler, Jr., M. P. Couper, J. M. Lepkowski, E. Singer, and R. Tourangeau, *Survey Methodology*, vol. 561. Hoboken, NJ, USA: Wiley, 2011.
[203] L. Gideon, *Handbook of Survey Methodology for the Social Sciences*. New York, NY, USA: Springer, 2012.
[204] F. J. Fowler, Jr., *Survey Research Methods*. Newbury Park, CA, USA: Sage, 2013.
[205] C. M. Hawkins, M. Hunter, G. E. Kolenic, and R. C. Carlos, ''Social media and peer-reviewed medical journal readership: A randomized prospective controlled trial,'' *J. Amer. College Radiol.*, vol. 14, no. 5, pp. 596–602, May 2017.
[206] A. Gates, R. Featherstone, K. Shave, S. D. Scott, and L. Hartling, ''Dissemination of evidence in paediatric emergency medicine: A quantitative descriptive evaluation of a 16-week social media promotion,'' *BMJ Open*, vol. 8, no. 6, Jun. 2018, Art. no. e022298.
[207] L. A. Sabato, C. Barone, and K. McKinney, ''Use of social media to engage membership of a state health-system pharmacy organization,'' *Amer. J. Health-System Pharmacy*, vol. 74, no. 1, pp. e72–e75, Jan. 2017.
[208] M. L. Wang, M. E. Waring, D. E. Jake-Schoffman, J. L. Oleski, Z. Michaels, J. M. Goetz, S. C. Lemon, Y. Ma, and S. L. Pagoto, ''Clinic versus online social network–delivered lifestyle interventions: Protocol for the get social noninferiority randomized controlled trial,'' *JMIR Res. Protocols*, vol. 6, no. 12, p. e243, 2017.
[209] M. Nishiwaki, N. Nakashima, Y. Ikegami, R. Kawakami, K. Kurobe, and N. Matsumoto, ''A pilot lifestyle intervention study: Effects of an intervention using an activity monitor and Twitter on physical activity and body composition,'' *J. Sports Med. Phys. Fitness*, vol. 57, no. 4, pp. 402–410, 2017.
[210] B. L. Berg, *Qualitative Research Methods for the Social Sciences*. London, U.K.: Pearson Higher Ed, 2001.
[211] A. Bhattacherjee, *Social Science Research: Principles, Methods, and Practices*. Tampa, FL, USA: USF Tampa Library Open Access Collections, 2012.
[212] D. Lockyer, ''The emotive meanings and functions of English'diminutive'interjections in Twitter posts,'' *SKASE J. Theor. Linguistics*, vol. 11, no. 2, pp. 1–6, 2014.
[213] B. De Cock and A. Pizarro Pedraza, ''From expressing solidarity to mocking on Twitter: Pragmatic functions of hashtags starting with #jesuis across languages,'' *Lang. Soc.*, vol. 47, no. 2, pp. 197–217, Apr. 2018.
[214] T. Jones, ''Toward a description of African American vernacular English dialect regions using 'Black Twitter,''' *Amer. Speech*, vol. 90, no. 4, pp. 403–440, Nov. 2015.
[215] N. A. B. Muhamad, N. Idris, and M. A. Saloot, ''Proposal: A hybrid dictionary modelling approach for malay tweet normalization,'' *J. Phys., Conf. Ser.*, vol. 806, Feb. 2017, Art. no. 012008.
[216] P. Lamabam and K. Chakma, ''A language identification system for code-mixed english-manipuri social media text,'' in *Proc. IEEE Int. Conf. Eng. Technol. (ICETECH)*, Mar. 2016, pp. 79–83.
[217] C. D. Manning, C. D. Manning, and H. Schütze, *Foundations of Statistical Natural Language Processing*. Cambridge, MA, USA: MIT Press, 1999.
[218] G. G. Chowdhury, ''Natural language processing,'' *Annu. Rev. Inf. Sci. Technol.*, vol. 37, no. 1, pp. 51–89, 2005.
[219] A. S. Clark, C. Fox, and S. Lappin, *The Handbook of Computational Linguistics and Natural Language Processing*. Hoboken, NJ, USA: Wiley, 2010.
[220] P. M. Nadkarni, L. Ohno-Machado, and W. W. Chapman, ''Natural language processing: An introduction,'' *J. Amer. Med. Inform. Assoc.*, vol. 18, no. 5, pp. 544–551, 2011.
[221] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, ''Natural language processing (almost) from scratch,'' *J. Mach. Learn. Res.*, vol. 12 pp. 2493–2537, Aug. 2011.

[222] C. Anagnostopoulos, L. Gillooly, D. Cook, P. Parganas, and S. Chadwick, "Stakeholder communication in 140 characters or less: A study of community sport foundations," *VOLUNTAS, Int. J. Voluntary Nonprofit Organizations*, vol. 28, no. 5, pp. 2224–2250, Oct. 2017.

[223] B. Sundstrom and A. B. Levenshus, "The art of engagement: Dialogic strategies on Twitter," *J. Commun. Manage.*, vol. 21, no. 1, pp. 17–33, Feb. 2017.

[224] R. Thackeray, B. L. Neiger, S. H. Burton, and C. R. Thackeray, "Analysis of the purpose of state health Departments' tweets: Information sharing, engagement, and action," *J. Med. Internet Res.*, vol. 15, no. 11, p. e255, 2013.

[225] W. Shin, A. Pang, and H. J. Kim, "Building relationships through integrated online media: Global Organizations' use of brand Web sites, Facebook, and Twitter," *J. Bus. Tech. Commun.*, vol. 29, no. 2, pp. 184–220, Apr. 2015.

[226] W. Tao and C. Wilson, "Fortune1000 communication strategies on Facebook and Twitter," *J. Commun. Manage.*, vol. 19, no. 3, pp. 208–223, Aug. 2015.

[227] C. O'Connor, "'Appeals to nature' in marriage equality debates: A content analysis of newspaper and social media discourse," *Brit. J. Social Psychol.*, vol. 56, no. 3, pp. 493–514, Sep. 2017.

[228] P. J. Jacques and C. C. Knox, "Hurricanes and hegemony: A qualitative analysis of micro-level climate change denial discourses," *Environ. Politics*, vol. 25, no. 5, pp. 831–852, Sep. 2016.

[229] J. M. Smith and T. van Ierland, "Framing controversy on social media: #NoDAPL and the debate about the Dakota access pipeline on Twitter," *IEEE Trans. Prof. Commun.*, vol. 61, no. 3, pp. 226–241, Sep. 2018.

[230] K. Gupta, J. Ripberger, and W. Wehde, "Advocacy group messaging on social media: Using the narrative policy framework to study Twitter messages about nuclear energy policy in the united states," *Policy Stud. J.*, vol. 46, no. 1, pp. 119–136, Feb. 2018.

[231] B. Kalsnes, A. H. Krumsvik, and T. Storsul, "Social media as a political backchannel: Twitter use during televised election debates in Norway," *Aslib J. Inf. Manage.*, vol. 66, no. 3, pp. 313–328, May 2014.

[232] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," in *Mining Text Data*. Boston, MA, USA: Springer, 2012, pp. 415–463.

[233] D. Al-Moghrabi, A. Johal, and P. S. Fleming, "What are people tweeting about orthodontic retention? A cross-sectional content analysis," *Amer. J. Orthodontics Dentofacial Orthopedics*, vol. 152, no. 4, pp. 516–522, Oct. 2017.

[234] D. L. Jessup, M. Glover IV, D. Daye, L. Banzi, P. Jones, G. Choy, J.-A.-O. Shepard, and E. J. Flores, "Implementation of digital awareness strategies to engage patients and providers in a lung cancer screening program: Retrospective study," *J. Med. Internet Res.*, vol. 20, no. 2, p. e52, 2018.

[235] N. J. Reavley and P. D. Pilkington, "Use of Twitter to monitor attitudes toward depression and schizophrenia: An exploratory study," *PeerJ*, vol. 2, p. e647, Oct. 2014.

[236] J. K. Harris, N. L. Mueller, D. Snider, and D. Haire-Joshu, "Local health department use of Twitter to disseminate diabetes information, united states," *Preventing Chronic Disease*, vol. 10, May 2013.

[237] P. Robinson, D. Turk, S. Jilka, and M. Cella, "Measuring attitudes towards mental health using social media: Investigating stigma and trivialisation," *Social Psychiatry Psychiatric Epidemiology*, vol. 54, no. 1, pp. 51–58, Jan. 2019.

[238] K. Ostwald and J. Dierkes, "Canada's foreign policy and bureaucratic (un)responsiveness: Public diplomacy in the digital domain," *Can. Foreign Policy J.*, vol. 24, no. 2, pp. 202–222, May 2018.

[239] K. Mossberger, Y. Wu, and J. Crawford, "Connecting citizens and local governments? Social media and interactivity in major U.S. cities," *Government Inf. Quart.*, vol. 30, no. 4, pp. 351–358, Oct. 2013.

[240] J. H. Kim and H. W. Park, "Food policy in cyberspace: A Webometric analysis of national food clusters in South Korea," *Government Inf. Quart.*, vol. 31, no. 3, pp. 443–453, Jul. 2014.

[241] P. Gunawong, "Open government and social media: A focus on transparency," *Social Sci. Comput. Rev.*, vol. 33, no. 5, pp. 587–598, Oct. 2015.

[242] C. J. Schneider, "Police presentational strategies on Twitter in Canada," *Policing Soc.*, vol. 26, no. 2, pp. 129–147, Feb. 2016.

[243] H. Wang, F. Wang, and J. Liu, "Accelerating peer-to-peer file sharing with social relations: Potentials and challenges," in *Proc. IEEE INFOCOM*, Mar. 2012, pp. 2891–2895.

[244] A. Gruzd, B. Wellman, and Y. Takhteyev, "Imagining Twitter as an imagined community," *Amer. Behav. Scientist*, vol. 55, no. 10, pp. 1294–1318, Oct. 2011.

[245] S. Gilbert, "Learning in a Twitter-based community of practice: An exploration of knowledge exchange as a motivation for participation in #hcsmca," *Inf., Commun. Soc.*, vol. 19, no. 9, pp. 1214–1232, Sep. 2016.

[246] C. N. May, M. E. Waring, S. Rodrigues, J. L. Oleski, E. Olendzki, M. Evans, J. Carey, and S. L. Pagoto, "Weight loss support seeking on Twitter: The impact of weight on follow back rates and interactions," *Transl. Behav. Med.*, vol. 7, no. 1, pp. 84–91, Mar. 2017.

[247] J. Stragier, P. Mechant, L. De Marez, and G. Cardon, "Computer-mediated social support for physical activity: A content analysis," *Health Edu. Behav.*, vol. 45, no. 1, pp. 124–131, Feb. 2018.

[248] H. Boateng, "Customer knowledge management practices on a social media platform: A case study of MTN ghana and vodafone ghana," *Inf. Develop.*, vol. 32, no. 3, pp. 440–451, Jun. 2016.

[249] S. Nabareseh, E. Afful-Dadzie, and P. Klimek, "Leveraging fine-grained sentiment analysis for competitivity," *J. Inf. Knowl. Manage.*, vol. 17, no. 02, Jun. 2018, Art. no. 1850018.

[250] J. S. Ryu, "Consumer attitudes and shopping intentions toward pop-up fashion stores," *J. Global Fashion Marketing*, vol. 2, no. 3, pp. 139–147, Aug. 2011.

[251] B. Y. Liau and P. P. Tan, "Gaining customer knowledge in low cost airlines through text mining," *Ind. Manage. Data Syst.*, vol. 114, no. 9, pp. 1344–1359, Oct. 2014.

[252] L. de Vries, S. Gensler, and P. S. H. Leeflang, "Effects of traditional advertising and social messages on brand-building metrics and customer acquisition," *J. Marketing*, vol. 81, no. 5, pp. 1–15, Sep. 2017.

[253] R. B. Clayton, "The third wheel: The impact of Twitter use on relationship infidelity and divorce," *Cyberpsychol., Behav., Social Netw.*, vol. 17, no. 7, pp. 425–430, Jul. 2014.

[254] J. Phua, S. V. Jin, and J. J. Kim, "Uses and gratifications of social networking sites for bridging and bonding social capital: A comparison of Facebook, Twitter, Instagram, and Snapchat," *Comput. Hum. Behav.*, vol. 72, pp. 115–122, Jul. 2017.

[255] L. V. Huang and P. L. Liu, "Ties that work: Investigating the relationships among coworker connections, work-related facebook utility, online social capital, and employee outcomes," *Comput. Hum. Behav.*, vol. 72, pp. 512–524, Jul. 2017.

[256] M. F. Wright and Y. Li, "The associations between young adults' face-to-face prosocial behaviors and their online prosocial behaviors," *Comput. Hum. Behav.*, vol. 27, no. 5, pp. 1959–1962, Sep. 2011.

[257] M. Kim and J. Cha, "A comparison of facebook, Twitter, and LinkedIn: Examining motivations and network externalities for the use of social networking sites," *1st Monday*, vol. 22, no. 11, Oct. 2017. [Online]. Available: https://firstmonday.org/ojs/index.php/fm/article/view/8066

[258] X. F. Liu and C. K. Tse, "Impact of degree mixing pattern on consensus formation in social networks," *Phys. A, Stat. Mech. Appl.*, vol. 407, pp. 1–6, Aug. 2014.

[259] A. Duma and A. Topirceanu, "A network motif based approach for classifying online social networks," in *Proc. IEEE 9th IEEE Int. Symp. Appl. Comput. Intell. Informat. (SACI)*, May 2014, pp. 311–315.

[260] C. Li, H. Wang, and P. Van Mieghem, "Epidemic threshold in directed networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 88, no. 6, Dec. 2013, Art. no. 062802.

[261] Z. Zhang and Z. Wang, "The data-driven null models for information dissemination tree in social networks," *Phys. A, Stat. Mech. Appl.*, vol. 484, pp. 394–411, Oct. 2017.

[262] C. Jiang, Y. Chen, and K. J. R. Liu, "Evolutionary dynamics of information diffusion over social networks," *IEEE Trans. Signal Process.*, vol. 62, no. 17, pp. 4573–4586, Sep. 2014.

[263] J. Scott and P. J. Carrington, *The SAGE Handbook of Social Network Analysis*. Newbury Park, CA, USA: Sage, 2011.

[264] A. Marin and B. Wellman, "Social network analysis: An introduction," *The SAGE Handbook of Social Network Analysis*, vol. 11. London, U.K.: SAGE, 2011.

[265] R. S. Burt, M. Kilduff, and S. Tasselli, "Social network analysis: Foundations and frontiers on advantage," *Annu. Rev. Psychol.*, vol. 64, no. 1, pp. 527–547, Jan. 2013.

[266] O. Serrat, *Social Network Analysis*. Singapore: Springer, 2017, pp. 39–43, doi: 10.1007/978-981-10-0983-9_9.

[267] A. S. Alberga, S. J. Withnell, and K. M. von Ranson, "Fitspiration and thinspiration: A comparison across three social networking sites," *J. Eating Disorders*, vol. 6, no. 1, p. 39, Dec. 2018.

[268] J. Yoon and E. Chung, "Image use in social network communication: A case study of tweets on the Boston marathon bombing," *Inf. Res.*, vol. 21, no. 1, pp. 1–23, 2016.

[269] D. B. Branley and J. Covey, ''Pro-ana versus pro-recovery: A content analytic comparison of social media Users' communication about eating disorders on Twitter and Tumblr,'' *Frontiers Psychol.*, vol. 8, p. 1356, Aug. 2017.
[270] M. Romney, R. G. Johnson, and K. Roschke, ''Narratives of life experience in the digital space: A case study of the images in Richard Deitsch's single best moment project,'' *Inf., Commun. Soc.*, vol. 20, no. 7, pp. 1040–1056, Jul. 2017.
[271] T. Mitts, ''From isolation to radicalization: Anti-muslim hostility and support for ISIS in the west,'' *Amer. Political Sci. Rev.*, vol. 113, no. 1, pp. 173–194, Feb. 2019.
[272] M. W. Bauer and G. Gaskell, *Qualitative Researching With Text, Image and Sound: A Practical Handbook for Social Research*. Newbury Park, CA, USA: Sage, 2000.
[273] L. Spenceley, *Qualitative Researching With Text, Image and Sound: A Practical Handbook for Social Research* (Research Methods in Education: Qualitative Research in Education), L. Atkins and S. Wallace, Eds. Newbury Park, CA, USA: Sage, 2012, pp. 189–206.
[274] D. Camenga, K. M. Gutierrez, G. Kong, D. Cavallo, P. Simon, and S. Krishnan-Sarin, ''E-cigarette advertising exposure in e-cigarette naïve adolescents and subsequent e-cigarette use: A longitudinal cohort study,'' *Addictive Behaviors*, vol. 81, pp. 78–83, Jun. 2018.
[275] L. Montgomery, K. Heidelburg, and C. Robinson, ''Characterizing blunt use among Twitter users: Racial/Ethnic differences in use patterns and characteristics,'' *Substance Use Misuse*, vol. 53, no. 3, pp. 501–507, Feb. 2018.
[276] M. J. Krauss, R. A. Grucza, L. J. Bierut, and P. A. Cavazos-Rehg, '''Get drunk. Smoke weed. Have fun.': A content analysis of tweets about marijuana and alcohol,'' *Amer. J. Health Promotion*, vol. 31, no. 3, pp. 200–208, May 2017.
[277] M. J. Krauss, S. J. Sowles, M. Moreno, K. Zewdie, R. A. Grucza, L. J. Bierut, and P. A. Cavazos-Rehg, ''Hookah-related Twitter chatter: A content analysis,'' *Preventing Chronic Disease*, vol. 12, Jul. 2015. [Online]. Available: https://www.cdc.gov/pcd/issues/2015/15_0140.htm
[278] T. K. Mackey, J. Kalyanam, T. Katsuki, and G. Lanckriet, ''Twitter-based detection of illegal online sale of prescription opioid,'' *Amer. J. Public Health*, vol. 107, no. 12, pp. 1910–1915, Dec. 2017.
[279] H. Baer, ''Redoing feminism: Digital activism, body politics, and neoliberalism,'' *Feminist Media Stud.*, vol. 16, no. 1, pp. 17–34, Jan. 2016.
[280] P. Prasad, ''Beyond rights as recognition: Black Twitter and posthuman coalitional possibilities,'' *Prose Stud.*, vol. 38, no. 1, pp. 50–73, Jan. 2016.
[281] K. M. Bivens and K. Cole, ''The grotesque protest in social media as embodied, political rhetoric,'' *J. Commun. Inquiry*, vol. 42, no. 1, pp. 5–25, Jan. 2018.
[282] A. Woloshyn, '''Welcome to the Tundra': Tanya Tagaq's creative and communicative agency as political strategy,'' *J. Popular Music Stud.*, vol. 29, no. 4, Dec. 2017, Art. no. e12254.
[283] A. I. Khan, ''A rant good for business: Communicative capitalism and the capture of anti-racist resistance,'' *Popular Commun.*, vol. 14, no. 1, pp. 39–48, Jan. 2016.
[284] T. Espinha, A. Zaidman, and H.-G. Gross, ''Web API growing pains: Stories from client developers and their code,'' in *Proc. Softw. Evol. Week, IEEE Conf. Softw. Maintenance, Reengineering, Reverse Eng. (CSMR-WCRE)*, Feb. 2014, pp. 84–93.
[285] M. D. P. Salas-Zárate, G. Alor-Hernández, R. Valencia-García, L. Rodríguez-Mazahua, A. Rodríguez-González, and J. L. López Cuadrado, ''Analyzing best practices on Web development frameworks: The lift approach,'' *Sci. Comput. Program.*, vol. 102, pp. 1–19, May 2015.
[286] Y. Qiu, ''The openness of open application programming interfaces,'' *Inf., Commun. Soc.*, vol. 20, no. 11, pp. 1720–1736, Nov. 2017.
[287] K. Theimer, ''What is the meaning of archives 2.0?'' *Amer. Archivist*, vol. 74, no. 1, pp. 58–68, Apr. 2011.
[288] J. Littman, D. Chudnov, D. Kerchner, C. Peterson, Y. Tan, R. Trent, R. Vij, and L. Wrubel, ''API-based social media collecting as a form of Web archiving,'' *Int. J. Digit. Libraries*, vol. 19, no. 1, pp. 21–38, Mar. 2018.
[289] L. Chang, W. Li, L. Qin, W. Zhang, and S. Yang, ''pSCAN: Fast and exact structural graph clustering,'' *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 2, pp. 387–401, Jan. 2017.
[290] S. Rayana and L. Akoglu, ''Less is more: Building selective anomaly ensembles,'' *ACM Trans. Knowl. Discov. Data*, vol. 10, no. 4, p. 42, 2015.
[291] J. P. Fairbanks, R. Kannan, H. Park, and D. A. Bader, ''Behavioral clusters in dynamic graphs,'' *Parallel Comput.*, vol. 47, pp. 38–50, Aug. 2015.
[292] F. Bistaffa and A. Farinelli, ''A COP model for graph-constrained coalition formation,'' *J. Artif. Intell. Res.*, vol. 62, pp. 133–153, May 2018.
[293] D. Chakrabarti and C. Faloutsos, ''Graph mining: Laws, generators, and algorithms,'' *ACM Comput. Surveys*, vol. 38, no. 1, pp. 2–es, Jun. 2006.
[294] L. Tang and H. Liu, ''Graph mining applications to social network analysis,'' in *Managing and Mining Graph Data*. Boston, MA, USA: Springer, 2010, pp. 487–513.
[295] U. Kang and C. Faloutsos, ''Big graph mining: Algorithms and discoveries,'' *ACM SIGKDD Explorations Newslett.*, vol. 14, no. 2, pp. 29–36, Apr. 2013.
[296] K. Li, K. Darr, and F. Gao, ''Enriching classroom learning through a microblogging-supported activity,'' *E-Learning Digit. Media*, vol. 15, no. 2, pp. 93–107, Mar. 2018.
[297] K. Hull and J. E. Dodd, ''Faculty use of Twitter in higher education teaching,'' *J. Appl. Res. Higher Edu.*, vol. 9, no. 1, pp. 91–104, Feb. 2017.
[298] T. Trust and E. Pektas, ''Using the ADDIE model and universal design for learning principles to develop an open online course for teacher professional development,'' *J. Digit. Learn. Teacher Edu.*, vol. 34, no. 4, pp. 219–233, Oct. 2018.
[299] J. Carpenter, ''Preservice teachers' microblogging: Professional development via Twitter,'' *Contemp. Issues Technol. Teacher Edu.*, vol. 15, no. 2, pp. 209–234, 2015.
[300] J. Colwell and A. C. Hutchison, ''Considering a Twitter-based professional learning network in literacy education,'' *Literacy Res. Instruct.*, vol. 57, no. 1, pp. 5–25, Jan. 2018.
[301] T. Zhu, H. Gao, Y. Yang, K. Bu, Y. Chen, D. Downey, K. Lee, and A. N. Choudhary, ''Beating the artificial chaos: Fighting OSN spam using its own templates,'' *IEEE/ACM Trans. Netw.*, vol. 24, no. 6, pp. 3856–3869, Dec. 2016.
[302] J. Zhang, R. Zhang, Y. Zhang, and G. Yan, ''The rise of social Botnets: Attacks and countermeasures,'' *IEEE Trans. Dependable Secure Comput.*, vol. 15, no. 6, pp. 1068–1082, Nov. 2018.
[303] S. Lee and J. Kim, ''Early filtering of ephemeral malicious accounts on Twitter,'' *Comput. Commun.*, vol. 54, pp. 48–57, Dec. 2014.
[304] M. Torky, A. Meligy, and H. Ibrahim, ''Recognizing fake identities in online social networks based on a finite automaton approach,'' in *Proc. 12th Int. Comput. Eng. Conf. (ICENCO)*, Dec. 2016, pp. 1–7.
[305] S. Lee and J. Kim, ''WarningBird: A near real-time detection system for suspicious URLs in Twitter stream,'' *IEEE Trans. Dependable Secure Comput.*, vol. 10, no. 3, pp. 183–195, May 2013.
[306] Y.-M. Kim and D. Delen, ''Medical informatics research trend analysis: A text mining approach,'' *Health Informat. J.*, vol. 24, no. 4, pp. 432–452, Dec. 2018.

AMIR KARAMI is currently an Assistant Professor with the College of Information and Communications, a Faculty Associate with the Arnold School of Public Health, and the Social Media Core Director with the Big Data Health Science Center (BDHSC), University of South Carolina. His research was published in different journals such as *Information Processing and Management*, *Journal of Biomedical Informatics*, the *International Journal of Fuzzy Systems*, the *International Journal of Information Management*, and conferences, such as the Association for Information Science and Technology (ASIS&T) Annual Meeting, iConference, the Hawaii International Conference on System Science (HICSS), and the International Conference on Data Mining (ICDM). His research interests include social media analysis, text mining, computational social science, and medical/health informatics.

MORGAN LUNDY is currently pursuing the dual master's degrees with the College of Arts and Sciences and the College of Information and Communication, University of South Carolina. Her research interests include social media analysis, health informatics, and data science.

FRANK WEBB is currently pursuing the degree with the South Carolina Honors College, University of South Carolina. His work has been published with the Southern Association for Information Systems (SAIS) and the Annual Meeting of Association for Information Science and Technology (ASIS&T). His research interests include social media analysis, health informatics, and data science.

YOGESH K. DWIVEDI is currently a Professor of digital marketing and innovation, the Founding Director of the Emerging Markets Research Centre (EMaRC), and the Co-Director of Research with the School of Management, Swansea University, U.K. He has published more than 300 articles in a range of leading academic journals and conferences that are widely cited (more than 14 thousand times as per Google Scholar). His research interests include interface of information systems (IS) and marketing, focusing on issues related to consumer adoption and diffusion of emerging digital innovations, digital government, and digital marketing, particularly in the context of emerging markets. He is an Associate Editor of *European Journal of Marketing*, *Government Information Quarterly*, and the *International Journal of Electronic Government Research*. He is a Senior Editor of *Journal of Electronic Commerce Research*. He is also the Editor-in-Chief leading the *International Journal of Information Management*.