Do news media and citizens have the same agenda on COVID-19? An empirical comparison of Twitter posts

Chunjia Han, Mu Yang* and Athena Piterou
Department of Systems Management and Strategy, University of Greenwich, UK

Abstract

This study analyses the agenda setting on social media in the COVID-19 pandemic by exploiting one of the disruptive technologies, big data analytics. Our purpose is to examine whether the agenda of news organisations matches the public agenda on social media in crisis situations, and to explore the feasibility and efficacy of applying big data analytics on social media data. To this end, we used an unsupervised machine learning approach, structural topic modelling and analysed 129,965 tweets posted by UK news media and citizens during April 2 and 8, 2020. Our study reveals a wide diversity of topics in the tweets generated by both groups and finds only a small number of topics are similar, indicating different agendas set in the pandemic. Moreover, we show that citizen tweets focused more on expressing feelings and sharing personal activities while news media tweets talked more about facts and analysis on COVID-19. In addition, our results find that citizens responded more significantly to breaking news. The findings of the study contribute to the agenda setting literature and offer valuable practical implications.

Keywords: COVID-19; big data analytics; social media; news media; citizen; agenda setting

*Dr. Mu Yang is the corresponding author.

1 Introduction

Since December 2019, the coronavirus disease 2019 (COVID-19) epidemic has swept the world, causing significant impact on society and economy (WHO, 2020; Nagano et al., 2020). The impact has attracted great attention from news media: the COVID-19 pandemic has become the main topic of news in the last several months (Kouzy et al., 2020). Previous research on crisis communication finds that the way people seek information on crisis, evaluate and make crisis-related decisions is highly related to the content covered by news media (An & Gower, 2009). Indeed, news coverage is commonly considered to have a great influence on people's perception and behaviours in crisis events such as public health emergencies (Coombs, 2006; Francken et al., 2009).

With the fast development of the Internet and disruptive technologies, the sources from which people obtain new information have been changing (Yang & Han, 2019). Social media, such as Twitter has become an important tool for information dissemination and sharing (Bruns et al., 2013). Since the outbreak of COVID-19, the number of tweets about COVID-19 has been increasing at a strikingly rapid pace (Singh et al., 2020). These tweets include the ones generated by citizens and also the ones posted by news organisations via their organisational Twitter accounts. Adopting Twitter for disseminating news online can help news organisations deliver COVID-19 related information to citizens in a timelier manner. Moreover, unlike traditional news media (e.g., newspaper, television), the communication functions on Twitter such as retweeting, favouriting, and commenting can further facilitate the information dissemination (Kim et al., 2016). Furthermore, social media has been found to be able to stimulate interests around certain topics (Takahashi et al., 2015; Grover et al., 2019), which may help address public concerns, improve public wellbeing and facilitate the implementation of the measures introduced by government for containing the outbreak of COVID-19 (Cinelli et al., 2020).

According to the agenda setting theory (McCombs & Shaw, 1972), there is a strong

correlation between media coverage on certain issues and the perceived importance of these issues by the public, i.e., citizens (Tewksbury & Scheufele, 2009). In particular, the public agenda is considered to be heavily affected by press published by news organisations (Dearing et al., n.d.). Here, an agenda is a set of issues that can be ranked hierarchically according to importance. In the case of COVID-19, agenda setting can be observed regarding the COVID-19 related prominent issues across the public, media and society. Although social media has become a popular method for citizens to access information, the rise of social media is found to modify rather than eliminate the agenda setting role of news media (Sayre et al., 2010). News media are believed to still set the public agenda as social media is seen as the online versions of traditional news media and attract more public attention when it comes to online news (Hindman, 2009). However, although the general public and news media all create and consume social media content, their motivations for utilising social media are different. According to the uses and gratifications theory which addresses how individual users actively choose to consume media, individuals use social media to satisfy many of their combined needs, such as needs for information, socialisation and emotional support (Urista et al., 2009). Unlike individual users, news organisations have adopted social media in order to draw online consumers and access a wider audience (Armstrong & Gao, 2010).

Therefore, the question arises to what extent the content posted on social media by the general public are influenced by the agenda published by news organisations on social media in crisis situations. Is the public merely responding to media concerns on social media or is it also shaping the public agenda on issues? As news media and social media both play important roles for information dissemination during crises, it becomes important and necessary to investigate and answer these questions. However, to the best of our knowledge, there is no study focused on these questions especially in the context of the COVID-19 pandemic. Therefore, our research aims to fill this knowledge gap by investigating whether the topics that the general public post on social media are in line with the news topics published by news organisations on social

media, and assess the differences if inconsistency exists. We select Twitter as the study case in this research because it is one of the largest social media platforms, containing a huge amount of citizens' conversation about the COVID-19 pandemic (E. Chen et al., 2020). Moreover, Twitter is currently the most popular tool adopted by news organisations for content dissemination (Armstrong & Gao, 2010), making Twitter a suitable social media platform for our study.

A big challenge of analysing the social media data and answering the research questions is how to extract valuable insights from a large amount of content data. Thanks to the fast development of disruptive technologies such as big data analytics, which allows us to analyse a large amount of unstructured social media data, gaining insights and important findings in real-time, with a good level of speed and accuracy (Hashem et al., 2016; Blazquez & Domenech, 2018). In particular, we utilise an unsupervised text analysis approach, Structural topic modelling (STM), and analyse 129,965 Twitter posts collected by a Python-programmed web scrapper between April 2 and 8, 2020 with a focus on UK news media and citizens.

This study demonstrates how disruptive technologies such as big data analytics can be applied for COVID-19 analysis. There are a number of interesting findings disclosed in this research. Firstly, this research fills the research gap on how the value of unstructured user-generated data can be leveraged for investigating the difference of the general public and news media's response to crises with the use of big data analytics. It demonstrates the feasibility and effectiveness of using unsupervised machine learning approach, STM to gain insights from the large amount of text data on social media. Secondly, a number of topics on the COVID-19 pandemic are extracted from the STM model including four categories of the topics posted by news organisations and three categories from citizen generated tweets, demonstrating the agenda difference between the general public and news organisations. Thirdly, our study fills up the research gap by revealing that only a small number of topics are found in common in news media tweets and citizens' tweets, and most of these common topics do not account

for similar shares. This finding suggests that the agendas of news media and general public on social media platforms are not well matched during the pandemic. Fourthly, from observing the variation of the common topics' proportions over time, we find that both news media and citizens show changing patterns of concerns on breaking news. However, the changing patterns are not the same: citizens' response to breaking news are more significant. Apart from the academic contributions, this study offers valuable practical implications. The analysis results can be applied to guide practitioners in the news industry on the development of their social media strategies, help policy makers better understand citizens' concerns on the COVID-19 pandemic, and facilitate the use of social media platforms on navigating people's interests and improving public wellbeing.

2 Literature review

2.1 Social media as sources of information and support in crises

In crisis situations individuals have increased needs of information and social support. Social media serves as a tool of crisis communication in situations of high uncertainty such as pandemics, natural disasters and terrorist attacks. Crises are characterised by their unprecedented occurrence, uncertainty, lack of information and the need for sensemaking (Mirbabaie & Zapatka, 2017; Brachten et al., 2018; Stieglitz et al., 2018). Reuter et al. (2018) investigate the emerging field of crisis informatics, which investigates how ICT and social media are used before, during and after emergency and crisis events. Social media use increases in a crisis (Austin et al., 2012) as people try to cope with uncertainty and deal with disaster creatively (Panagiotopoulos et al., 2016). In a crisis, audiences may turn to traditional media for educational purposes and to social media for insider information and to check in with family and friends (Austin

et al., 2012). Social media may be considered a source of more reliable and unfiltered information (Procopio & Procopio, 2007), which may not be available from other sources (Sutton et al., n.d.).

Social media use during natural disasters can be nuanced with users also employing humour and sharing photos such as in the aftermath of Hurricane Sandy (Murthy & Gross, 2017). Analysis of tweets on Australian floods identified various uses including sharing information and personal experiences, and expressing gratitude (Shaw et al., 2013). The role of social media in expressing gratitude and connecting communities while healing from a disaster was also addressed by Glasgow et al. (2016). There are a number of studies exploring tweets about COVID-19 noting issues such as conspiracy theories and misinformation (e.g. E. Chen et al., 2020; Kouzy et al., 2020; Singh et al., 2020). In the case of Zika, a comparison of tweets posted by the public and the Centre for Disease Control identified different concerns (Glowacki et al., 2016). During the 2009 H1N1 epidemic twitter was mainly used to share information from credible sources, with tweeting activity peaking after major news stories (Chew & Eysenbach, 2010).

2.2 The use of social media by individuals and news organisations

According to uses and gratifications theory, individuals use social media that satisfy many of their combined needs such as needs for information, socialisation and emotional support (Urista et al., 2009). Audiences may also choose to engage with media that reinforce their current beliefs and ways of thinking (Austin et al., 2012). Individuals as consumers tend to create their own content to satisfy social and ego-defensive needs rather than for utilitarian knowledge seeking (Daugherty et al., 2008). Young adults mainly use communication technologies for socialisation and for emotional needs, and to a lesser degree for utilitarian reasons (Behairy et al., 2006). Yoo et al. (2014)

examined how social influences, including social conformity and social values affect twitter use. They find that both hedonic and utilitarian features of social networking sites affect their rate of adoption, however, social influences have been overlooked. Yet, G. M. Chen (2011) argues that the primary motivation of using twitter is utilitarian as it is an information-oriented social networking site.

News organisations have adopted social media to attract online consumers and increase their audience (Armstrong & Gao, 2010). Twitter has been used for breaking up-to date information in crisis situations, such as the London riots in 2011 (Vis, 2013). The use of twitter may challenge the gatekeeping role of traditional journalism and news media (Armstrong & Gao, 2010; Lasorsa et al., 2012). Boersma & Graham (2013) showed that journalists are increasingly using Twitter as a source of newsworthy information or to illustrate a story. Still, they are more likely to cite tweets from official sources (Moon & Hadley, 2014). Bane (2019) remarks that traditional news organisations are more likely to use twitter to quote official sources and for opinion comments. Although evidence suggests that news media do not make use of the connective or technological capabilities of twitter to a full extent (Engesser & Humprecht, 2015), Hermida (2010) argues that the increasing use of social media in news communication has created new awareness systems -described as ambient journalism- that involve the always connected, digital dissemination of information from multiple sources.

2.3 Agenda setting on social media

McCombs & Shaw (1972) established agenda setting research by drawing on the distinction between the press telling the readers *what to think* and *what to think about*. Agenda setting addresses the latter question. The theory assumes a strong correlation between media coverage on certain issues and the perceived importance of these issues by the public (Tewksbury & Scheufele, 2009). Dearing et al. (n.d.) distinguish between the media agenda, the public agenda and the policy agenda. An agenda is

a set of issues that can be ranked by their importance. Agenda setting is viewed as a political process which is necessary in all societies, in order to prioritise problems. Agenda setting relates to the importance of issues, while agenda framing defines what attributes of an issue are emphasised (Chong & Druckman, 2007). Gamson (1992) described framing as a "signature matrix" that includes various condensing symbols (such as catchphrases, taglines, visual images) and reasoning devices. Twitter hashtags can be included as symbols in this signature matrix.

To an extent traditional news media have lost some of their power in setting the public agenda since the public can access information from multiple sources, including social media: the media agenda cannot be equated with the public agenda. However, it is believed that traditional news media (i.e., print and broadcast) still set the public agenda as it is their online versions that attract more public attention when it comes to online news (Hindman, 2009). Arguably, the rise of social media has modified rather than eliminated the agenda setting role of traditional media (Sayre et al., 2010).

There is some extant research comparing agenda setting in traditional and social media. For instance, Meraz (2009) analysed hyperlinking to question whether the social media agenda on blogs is set by traditional media. Sayre et al. (2010) analysed agenda setting on traditional, online and social media in the case of Proposition 8 in California. However, it is still unclear how news media affect the general public agenda setting on social media platforms. With the increasing engagement of both the news media and the general public on social media, it becomes important to investigate the agenda setting role of news media organisations on social media platforms. Moreover, in crises such as the COVID-19 pandemic, when social media are a main channel for information dissemination and public communication, it becomes essential to clarify how news media affect the general public agenda setting via social media platforms. This would help us develop a better understanding on how both news media and social media can be applied to deal with crises. This study aims to fill the research gap by answering this question.

2.4 Disruptive technologies in social media analysis

Disruptive technologies are crucial for social science studies relating to COVID-19, as they are able to provide efficient, effective and low-cost analysis. Pioneer studies have applied disruptive technologies such as deep learning (Abdel-Basset et al., 2021) to process and analyse data and conclude important findings on COVID-19 infection (Abdel-Basset, Chang, & Mohamed, 2020; Baudier et al., 2021). A framework using disruptive technologies for COVID-19 analysis is developed in Abdel-Basset et al. (2020)'s research. Nine disruptive technologies are discussed including artificial intelligence (AI), industry 4.0, Internet of Things (IoT), Internet of Medical Things (IoMT), big data, virtual reality (VR), Drone technology and Autonomous Robots, 5 G, and blockchain. This study provides an idea on how different organisations can benefit from the application of these disruptive technologies for fast decision making and reduction of the negative impact of COVID-19. For example, it is considered that two disruptive technologies visualisation and analytics play an essential role in the effective development and presentation of complex work in healthcare applications (Chang, 2018b). Additionally, another disruptive technology, computational intelligence has been widely adopted in medical research (Chang, 2018a).

As one of the disruptive technologies for COVID-19 related analysis (Abdel-Basset, Chang, & Nabeeh, 2020), big data analytics enables agenda research using social media data (Russell Neuman et al., 2014; Yang & Han, 2021). A review of the contribution of big data analytics in social media research identifies a range of text mining techniques including natural language processing, sentiment analysis, and social network analysis (Ghani et al., 2019). Using sentiment analysis for example, integrating sentiment analysis and other novel disruptive technologies can facilitate studies in different disciplines, especially in social media analysis (Xu et al., 2020; Chang et al., 2020; Karyotis et al., 2018). Such disruptive techniques are pertinent to agenda setting and agenda framing research as they can help rank the relative importance of issues, identify sym-

bols used to frame agendas and model the relation between different issues and different actors posting on social media. There is a wide range of topic modelling techniques, such as Latent Dirichlet allocation (LDA), have been used to analyse user-generated content online. The most relevant study is on tracking disease outbreaks using twitter data and LDA (Missier et al., 2016). In this research, we choose STM (M. E. Roberts et al., 2014) which is similar to LDA but has the advantage of allowing incorporating the metadata of data sample to explain topical prevalence (Büschken & Allenby, 2016; M. E. Roberts et al., 2016). This approach has been seen in analysing customer online reviews (Hu et al., 2019; Han & Yang, 2020).

3 Data and methods

Using unsupervised text analysis methods, this study examines the Twitter posts related to the COVID-19 pandemic in order to assess the differences in the discussion topics between news organisations and citizens on social media. A unified framework of extracting topics from rich social media data (e.g., Twitter posts) is presented in Figure 1, which summaries the data collection and analysis process carried out in our study.

3.1 Data acquisition and preparation

We choose UK as a test case which is of both theoretical and practical importance. The news industry in the UK are well developed and most news organisations have organisational social media accounts. A number of studies (Armstrong & Gao, 2010; Vis, 2013; Messner et al., 2012) have looked into how news organisations used Twitter for promoting political events in the context of UK. Moreover, extant literature on the theory of agenda setting have focused on the politics in the UK setting, e.g., (Cushion et al., 2018). We crawl all COVID-19 related posts generated from UK accounts

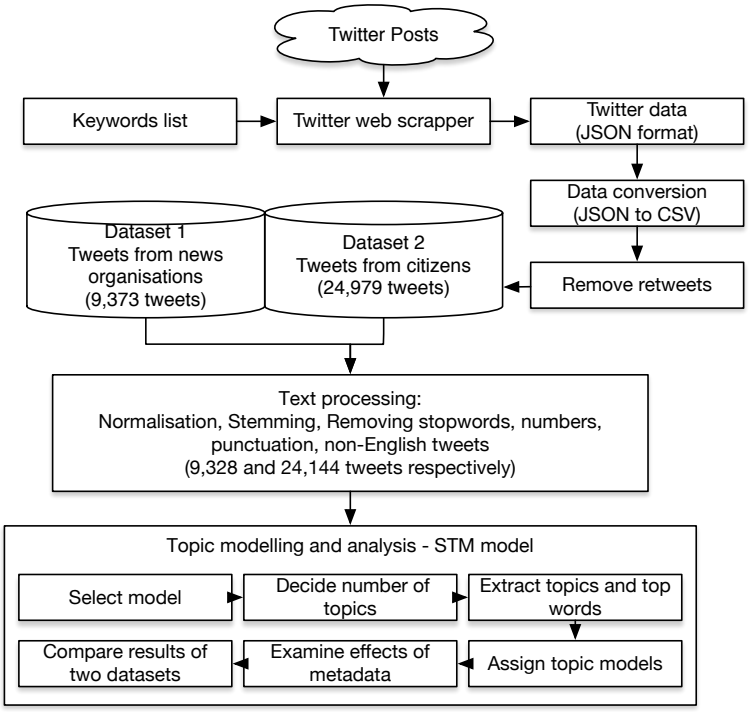

Figure 1: Overall framework for Twitter data collection and topic analysis

from Twitter, we set our test period from April 2 to April 8 2020. This is a good methodological choice because the aim of the research is to examine whether news organisations and citizens have different interests and concerns towards the pandemic, and how they respond to breaking news. The duration of one week is sufficient to detect the difference if exists due to the tremendous number of posts, and is able to capture the response to breaking news due to the time effect of breaking news. The duration is also widely adopted by similar studies, such as (Durahim & Coşkun, 2015; Paul et al., 2017; Panagiotopoulos et al., 2016). Moreover, at the chosen time period there was a strong sense of emergence about the pandemic and two pieces of breaking news happened, making the collected dataset useful to study crisis situations while later the pandemic had become the new normal (Buheji et al., 2020).

In order to collect appropriate sample data from the Twitter platform, we firstly defined a keywords list which specifies which Twitter postings are relevant. More specifically, we included texts such as “covid”, “corona”, “coronavirus” and COVID-19

related hashtags in the list, and specified the country setting as "GB" since we focus on UK related posts and time duration as one week, from April 2 to April 8, 2020. By developing a Python-based web scrapper, a total of 129,965 tweets were collected with the specification defined in the keywords list. The collected data were in JSON format and were converted into CSV files for analysis. The data columns include tweet generated time, text content, hashtags, URLs, source, retweet count, favourites count, follower count, user screen name, location, language, etc.

Next, we removed user IDs to preserve privacy and anonymity. This process is in line with the ethical framework for big data analysis research (Chang, 2021). We then cleaned the dataset by removing retweets, leaving us 34,352 tweets for the study. It is worth noting that the removed retweets are the ones that have exactly duplicate content. If people retweet a post but add content to it (e.g., expressing their attitude), then we do not consider it as a retweet and it was not removed from the dataset. There are two reasons we remove retweets: firstly, this study focuses on the textual content of tweets with the objective of identifying the differences in the discussion content between news organisations and citizens. Second, as Twitter bots have been frequently and widely used by companies, organisations and individuals to enable automatic retweet actions (Dickerson et al., 2014), removing retweets can eliminate the effects of Twitter bots and capture the true discussion of news organisations and citizens. We then separated the dataset by extracting tweets posted by news organisations. As suggested by a prior study (Armstrong & Gao, 2010), eight UK's most popular news organisations are chosen in the study, which are BBC News, BBC Breaking News, BBC World, Daily Mirror, Guardian, Independent, Telegraph, and The Times. This step gives us 9,373 tweets posted by eight UK news media accounts and 24,979 tweets from citizens. We define the news media tweets dataset as Dataset 1, and the citizen generated tweets dataset as Dataset 2.

The final step of the data acquisition and preparation is text processing, including normalisation, stemming and character removing which are consistent with similar

processing efforts applied to topic extraction research, such as (Berliner et al., 2018; Han & Yang, 2020). More specifically, words in the column of tweets text content were normalised by transforming all letters to lower case in order to obtain a more uniform form and reduce the size of the vocabulary. Then, the words were stemmed by removing affixes, followed by the character removing step where stopwords, numbers and punctuation are removed. Finally, non-English tweets were filtered out from both datasets. We used the *tidytext* Package (Silge & Robinson, 2016) in *R* programming language to automate the process.

3.2 Model setup

Topic models, as one stream of unsupervised text analysis methods have been found suitable for analysing user-generated content, such as customer online reviews (Büschken & Allenby, 2016). The methods examine the co-occurrence relationship among words and output the collections of words with high probability of co-occurrence, i.e., the topics. We choose STM (M. E. Roberts et al., 2014) among other approaches in this method category because of its advantage of allowing incorporating the metadata of data sample to explain topical prevalence (Büschken & Allenby, 2016; M. E. Roberts et al., 2016), for example, in our case metadata refer to the information associated with Twitter posts, e.g., tweets created time, favourites count, source, etc. The key process of STM can be summaries as follows (M. E. Roberts et al., 2014).

We apply STM to all the tweets collected for the study. In other words, we view the whole set of the tweets as a document. In STM, a document is defined as a mixture over topics, meaning that a document is composed of multiple topics (M. E. Roberts et al., 2014). Thus, the whole set of the tweets is a mixture over topics, and a topic is a mixture over words where each word has a probability of belonging to a topic. The STM is a hierarchical model in which a document d's prevalence of each topic (denoted by $d(\theta_d)$) is drawn from a logistic-normal distribution whose mean is a function of

document covariates X_d:

$$\theta_d \sim LogisticNormal(X_{d\gamma}, \Sigma)\,.$$

Then, given the topic-prevalence vector, one specific topic, $z_{d,n}$ is associated with the position which needs to be filled through the following process:

$$z_{d,n} \sim Multinomial(\theta_d)\,.$$

Next, the words of each document $w_{d,n}$ are assigned to the topics:

$$w_{d,n} \sim Multinomial(\beta_{d,z})$$

where $\beta_{d,z}$ is the probability of choosing vocabulary word w to fill a position in document d given topic z. The *stm* package (M. Roberts et al., 2018) in R is used to set up the model for our analysis where tweets text content are inputs as documents, prevalence function is set as follows.

$$prevalence \sim favourites\ count + s(tweets\ created\ time)\,, \tag{1}$$

where s is the smooth function of time, and *favourites count* is one of the topical prevalence covariates indicating how many likes a tweet had been received. It is worth noting that the covariate is replaced by other metadata items (e.g., source, follower count) when checking the robustness of the results. As the results are nearly the same, we only report the results under the setting of *favourites count* for simplicity.

Next, we decide on the number of topics K which is an important parameter of STM and helps to achieve substantive interpretation of the outcomes of the modelling (Li et al., 2013). Using function *searchK* from the *stm* and *furrr* (Vaughan & Dancho, 2018) packages in R, we evaluate the models trained on a sparse matrix in parallel with a range

of different values of K, i.e., from 5 to 30. Based on the trained models, we compute a number of model fitting indicators including the semantic coherence of the topics, held-out likelihood and residuals. By comparing the indicators, $K = 10$ is selected for both datasets. With $K = 10$, we have semantic coherence locally maximised at -164, held-out likelihood maximised at -7.34 and residuals at a relatively low value at 32.0.

4 Results and discussion

In this section, we present the results of topic extraction from the STM model, analyse the different motivations news organisations and citizens use social media by assessing the topic differences, and investigate whether news organisations influenced citizens' discussion in the COVID-19 crisis from the angle of agenda setting theory (Dearing et al., n.d.).

Our topic modelling approach is quite successful at identifying distinct and internally coherent topics from tweets content, and the obtained topics indicate both news media and citizens talked about a wide range of topics. Table 1 and 2 summarise the results. The second and third columns are the outputs of our STM model, showing the topic proportions and top seven words for each topic. The top words are the ones that have the highest probability of appearing in the topic but least probability in other topics. The first column are the topic labels assigned by two researchers in social science based on the top words, close reading of tweets examples in each topic and two studies on Covid-19 public discussion (Stokes et al., 2020; Xue et al., 2020). The topics are grouped into different categories based on the semantics of topic labels and the categories presented in these two studies.

Table 1: Topic summary - news media tweets

Topic Label	Topic Proportions	Top Words
Topic Category: COVID-19 update		
Topic 7: Death news	11.7%	die, will, death, hospit, live, find, warn
Topic 9: International covid news	10.6%	say, test, trump, video, claim, like, posit
Topic 6: Case number update	10.1%	nhs, face, mirrorceleb, two, year, number, keep
Topic Category: Advice and measures		
Topic 10: Safety advice	12.3%	home, how, can, safe, man, toll, make
Topic 1: Methods for slowing spread*	8.7%	show, spread, tech, get, break, use, theori
Topic 2: Rules introduced by industries/companies	8.5%	help, staff, take, first, countri, rule, virus
Topic Category: Breaking news		
Topic 4: Boris covid condition*	9.7%	peopl, bori, johnson, care, time, worker, intens
Topic 8: Queen's speech	9.1%	home, govern, nurs, queen, watch, now, labour
Topic Category: Society efforts		
Topic 3: Healthcare workers*	9.8%	nurs, crisi, world, health, need, patient, doctor
Topic 5: Societal support*	9.5%	new, day, worker, help, call, support, cut

Topics with the * mark are the common topics appear in both datasets.

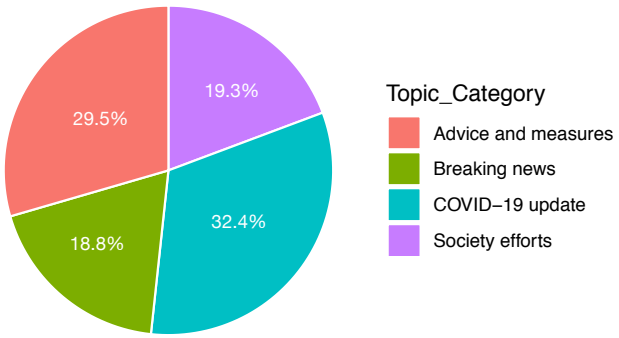

Figure 2: Topic categories for news organisations' generated tweets

4.1 Prevalent topics from news organisations

As shown in Table 1, news organisations posted tweets across a wide diversity of topics between April 2 and 8, 2020, covering four categories: COVID-19 update, Advice and measures, Breaking news and Society efforts. The percentage of each topic category is presented in Figure 2. From the topic proportions (see the second column of Table 1), all topics account for similar share, indicating that news organisations pay generally equal attention to COVID-19 related news.

The top three topics with the highest proportions are Safety advice (12.3%), Death news (11.7%) and International covid news (10.6%), accounting for 34.6% of all tweets. The Safety advice topic contains tweets about suggestions to people on how to stay

safe during the pandemic. For example,

> Quote: How to make sure your post is as safe as possible during coronavirus lockdown [from Daily Mirror]

The topic of Death news focuses on the news update about celebrities and workers death, which is reflected in the top words, such as "die", "death" and tweets examples. The topic International covid news relates to the news around the world, for example, the measures introduced by other countries, speech by other countries' leaders. The remaining seven topics with a proportion ranging from 8.5% to 10.1% are about the update of case numbers, situation of frontline healthcare workers, UK Prime Minister Boris Johnson's COVID-19 condition, the Queen's speech on COVID-19, societal and industry actions and rules as well as the approaches for slowing the spread of the virus. Among these topics, UK Prime Minister Boris Johnson's COVID-19 condition and the Queen's speech on COVID-19 are grouped into the Breaking news category. As noted by Vis (2013) that Twitter is a useful tool for breaking up to date news in emergencies, the identification of these two topics confirms this conclusion and shows UK news organisations used Twitter during the test period when there was high demand for news updates.

4.2 Prevalent topics from citizens

In terms of the citizen generated tweets, the identified topics are organised into three categories: Express feelings, Discussion of COVID-related issues and Sharing activities as shown in Table 2. Unlike the topics identified from news media tweets, these topics account for unequal shares with the top two topics Thank workers and Wishes for Boris take almost one third of all tweets.

The largest topic Thank workers accounting for 15.3% of all tweets, primarily relates to people's appreciation and thankfulness feelings for healthcare and community workers. As the study period of April 2 to April 8, 2020 was a critical week for the UK in terms of dealing with the pandemic, the results suggest that people use Twitter to

Table 2: Topic summary - citizen generated tweets

Topic Label	Topic Proportions	Top Words
Topic Category: Express feelings		
Topic 6: Thank workers*	15.3%	nhs, work, thank, support, staff, worker, nurs
Topic 9: Wishes for Boris*	13.8%	borisjohnson, hope, keep, safe, bori, wish, best
Topic 2: Family and friends	9.6%	love, take, famili, friend, away, long, lost
Topic Category: Discussion		
Topic 5: Social distance	10.2%	socialdistanc, one, stay, stayhom, walk, exercis, street
Topic 7: Sports	9.9%	now, player, live, watch, tri, footbal, leagu
Topic 8: Quarantine/lockdown measures	9.7%	lockdown, time, look, stayhomestaysaf, morn, unit, quarantin
Topic 10: Methods for slowing spread*	9.4%	peopl, feel, report, case, identifi, slow, spread
Topic 3: Government strategies	8.1%	govern, back, death, social, distanc, explor, follow
Topic 1: Covid testing	6.1%	test, stayhomesavel, make, know, matthancock, surviv, dailybrief
Topic Category: Sharing activities		
Topic 4: Activities at home	7.9%	day, today, week, isolation, sing, garden, amazonprim

Topics with the * mark are the common topics appear in both datasets.

support workers who are at the frontline and to keep up morale. This finding is in line with earlier studies that suggested the importance of sharing gratitude on social media platforms following disasters and emergencies (Shaw et al., 2013; Glasgow et al., 2016). The second largest topic, Wishes for UK PM Boris Johnson, accounts for 13.8% of all tweets comprises wishing for Boris to recover soon, suggesting the public were sharing a common emotional response. The third largest topic Social distance concerns about the social distancing measures introduced by government (10.2%). By close reading of tweet examples in this topic, it is found that people also talked about their own experience on social distancing apart from generally discussing the issues around the topic. Similar to our findings, Australian users were found sharing personal experience of natural disasters (i.e., floods) on Twitter (Shaw et al., 2013). Our finding further supports the conclusions of extant literature on the motivation of citizens use social media during crisis.

Besides these three topics, people expressed their feelings on caring about their families and friends (9.6%) and shared the activities they did at home during the lockdown (7.9%). Tweets examples for these five topics are presented as follows.

> (Thank workers): Thanks to @thurrockcouncil #WasteCollection crews for their services. #ThankYou #YourAreOurHeroes #COVID-19 #coronavirus #WestThurrock #Grays #Thurrock #StayHomeSaveLives
>
> (Wishes for Boris): Sending best wishes to our Prime Minister @BorisJohnson @10DowningStreet I hope you have a full and speedy recovery. The country is behind you and you're in the care of our amazing @NHSuk All

the best from Cornwall Sir, rest up and get better! #Covid19 #BorisJohnson

(Social distance): I did not have to worry this morning. #socialdistancing #roundwoodpark #covid_19 #oneplus7pro #nhs @ Roundwood Park

(Family and friends): I love video calls too but I am such a hugger I am really missing that physical contact with friends and family #shuvonshuvoff #kindness #hugging #coronavirus #StayHomeSaveLives

(Activities at home): Finally did my gardening so much of it #gardening #home #living alone #lockdown #coronavirus #summer #plants #mine

The rest of the tweets focused on the discussion of COVID-19 related topics, such as the impact of the pandemic on sports (9.9%), introduced quarantine and lockdown measures (9.7%), methods for slowing the virus spread (9.4%), government approaches and strategies (8.1%), and virus testing (6.1%). These topics are generally consistent with other studies on identifying concerns of Tweeters (Abd-Alrazaq et al., 2020). The percentages of tweets that each topic category covers are depicted in Figure 3.

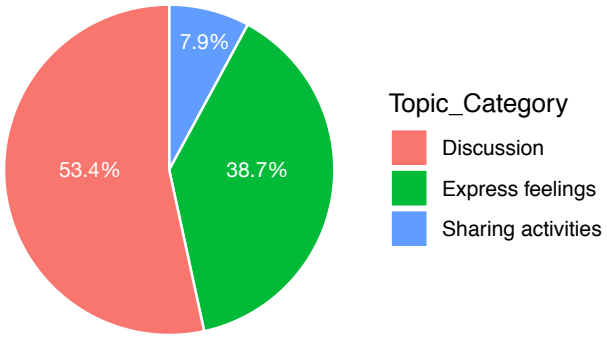

Figure 3: Topic categories for citizen generated tweets

4.3 Comparison of topics between news media tweets and citizen tweets

An interesting finding is that the tweets posted by news organisations are all about reporting facts and analysis on COVID-19 related issues; however, a large proportion of citizen generated tweets focus on personal feeling expression and sharing personal

activities. As shown in Figure 3, people posted tweets to express their feelings quite often, and liked to share their activities during the lockdown, accounting for 38.7% and 7.9% respectively of all tweets. Moreover, discussing COVID-19 related topics which takes the majority percentage (i.e., 53.4%), suggesting that Twitter is still seen as an information medium (G. M. Chen, 2011). However, these discussions are widely based on personal own experience rather than logical analysis such as news media generated tweets. This is reflected in both top words, such as "walk", "street", "live", "watch", "look", "morn[ing]", "feel", and a substantial number of tweets examples, such as the tweet quotation we presented in previous section on the topic Social distance. This finding shows that the role of Twitter in the COVID-19 pandemic is to address both emotional and information needs, which confirms the conclusions of previous study (Urista et al., 2009). The discussion on the issues such as lockdown measures and methods for slowing spread can be seen as an attempt of collective sensemaking (Stephens et al., 2020).

A small number of similar topics occurring in news media tweets and citizen tweets are found in our analysis, and are highlighted using star marks in Table 1 and 2. According to agenda setting theory discussed in Section 2, citizens respond to the news ranked highly in the media agenda but they may not allocate the same level of attention. Moreover, the media agenda could modified by social media activity: citizens may also frame some of the topics according to their personal experience (Chong & Druckman, 2007). For instance, discussion on quarantine measures and social distancing spills into the discussion of activities at home. In addition, most of these common topics do not account for similar shares. More specifically, the two topics - Healthcare workers and Societal support in news media tweets are corresponding to the topic Thank workers in citizen generated tweets. The former two topics account for 19.3% in total of all news media tweets while the latter accounts for 15.3% of all citizen generated tweets. The changes in the prevalence of the three topics are shown in Figure 4: the topic prevalences were all fluctuated during the test period.

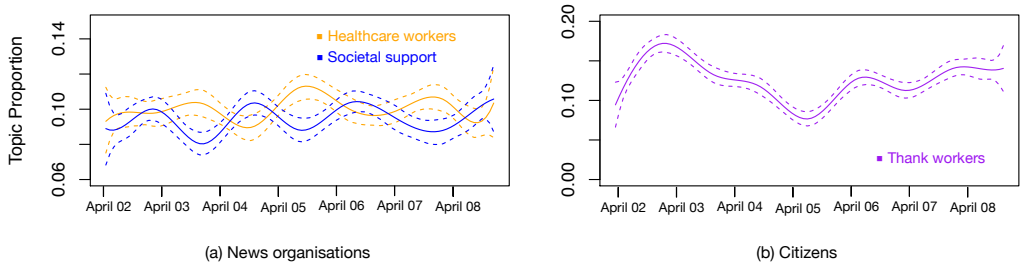

Figure 4: Trends over time for the common topic estimated by STM model: *Healthcare workers & Societal support* vs. *Thank workers* (The prevalence of each topic is plotted as a smooth function of tweet posted time as shown in Eq. 1, with 95% confidence intervals where the dotted curves imply.)

The other common topic, Boris covid condition identified from news media tweets corresponds to the topic Wishes for Boris in citizen generated tweets. Boris covid condition takes 9.7% share while the share that Wishes for Boris takes is much higher, i.e., 13.8%. Figure 5 plots the topic trend over the test period, and shows the topic proportion changes are different: the topic proportion for news media tweets fluctuated during the week while citizens' topic proportion increases dramatically in the later days of the week. The case of the H1N1 outbreak also showed a great increase in public's tweeting activity after major news stories (Chew & Eysenbach, 2010). The last common topic, Methods for slowing spread accounts for similar share in both types of tweets. The topic proportion for news media tweets is 8.7% and for citizen generated tweets is 9.4%. Furthermore, the trend of this topic is relatively flat during the test period as shown in Figure 5. This finding suggests that although news organisations and citizens have similar concerns on certain topics, they give these topics different levels of attention and the trends in these topics are different.

4.4 Different response to breaking news

The variation over time in the prevalence of the identified two common topics Boris covid condition/Wishes for Boris and Methods for slow spreading are plotted and presented in Figure 5. The proportions of the topic Methods for slow spreading are relatively flat over time for both news media tweets and citizen generated tweets, reflecting stable interests on the topic. However, when there is breaking news such as when UK PM Boris Johnson was infected with coronavirus, both news media and citizens show changing patterns of concerns on this topic. The proportion of the both topics trended up and down over time, reflecting that both news media and people have an attention on the breaking news and the attention changes when there is a change in Boris's COVID-19 infection condition.

Further, citizens' response to the breaking news are more significant. As shown in Figure 5(b), an increase in the topic proportion is observed when Boris was admitted to hospital. The topic proportion reached its highest level (i.e., nearly 25% of all tweets) when Boris was moved to ICU, and then started trending downwards. It is interesting to see that the tweets posted by news organisations on this topic does not show similar pattern. Instead, the topic proportion is not observed increasing much when the key events happened (see Figure 5(a)). This finding might indicate that news organisations treated this news as facts, while for the public there was an emotional response and need for social support with the feeling of the country in crisis.

Unlike Boris-related tweets, the other topic belonging to the Breaking news category, Queen's speech did not raise much attention from citizens. The number of related tweets is too small to form a topic. One example from citizen-generated tweets relating to this topic is

> Quote: Has the Queen found or funded a cure? Bought PPE or ventilators for #NHS or set aside a palace for #Coronavirus patients? No.....

This tweet is categorised by our model as a discussion around COVID-19. On the contrary, when looking at the news media tweets examples under the Queen's speech

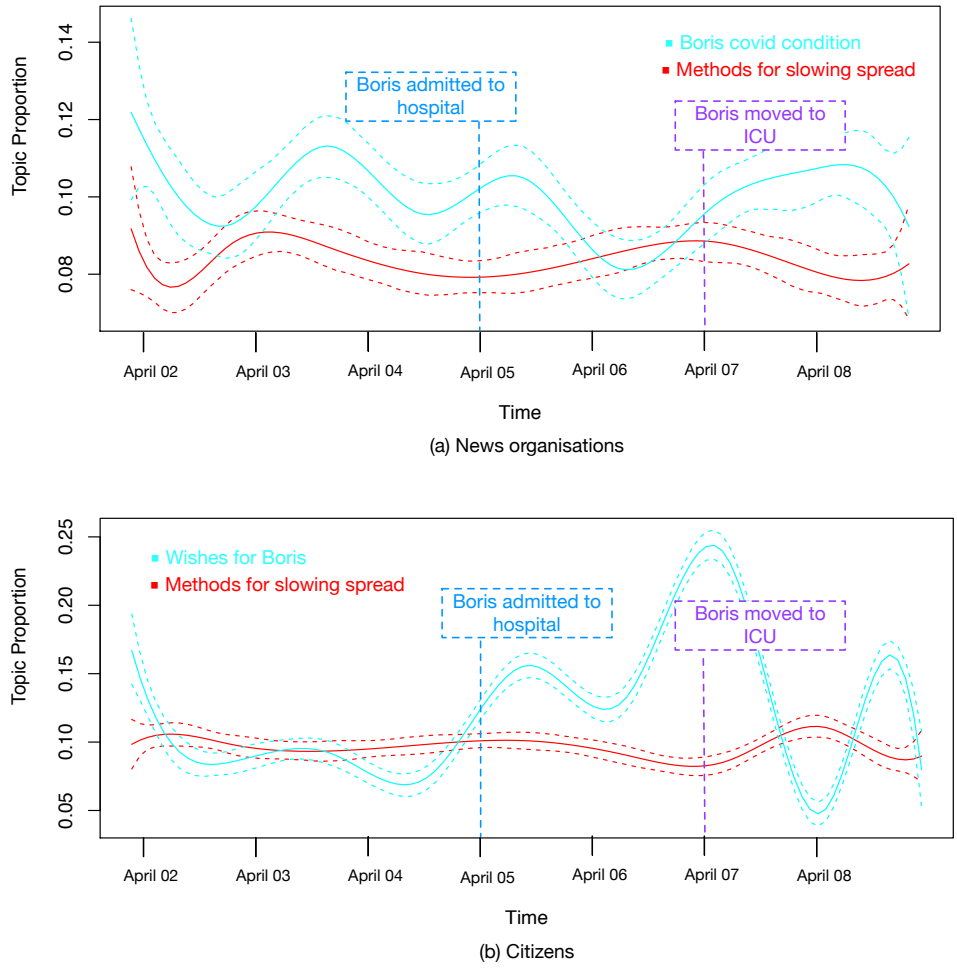

Figure 5: Trends over time in two common topics' proportions estimated by STM model: *Boris covid condition/Wishes for Boris* and *Methods for slowing spread* (The prevalence of each topic is plotted as a smooth function of tweet posted time as shown in Eq. 1, with 95% confidence intervals where the dotted curves imply.)

topic, we find that these tweets focus more on disseminating the information about when and how the Queen will give a speech, and use Twitter to drive the traffic to their websites by including the links of the news webpage. An example is listed below.

> Quote: Coronavirus: Queen to address nation in Sunday broadcast https://t.co/m1Uw4cNrtv [from BBC News]

This finding is in line with a number of studies, such as (Canter, 2013) which suggest

that using social media to drive more traffic to their online site is one of the main reasons news organisations engage with Twitter. Furthermore, this finding extends earlier studies on agenda setting (Bruns et al., 2013) which assume topics receiving a lot of media coverage would be more likely to migrate from the media to the public agenda: we show this may not be the case on social media. The issues discussed on social media are much broader, allowing for more expansive discussion of different attributes of a topic (Fortunato & Martin, 2016). Therefore, the tweets generated by citizens show a dramatic response to the Boris breaking news but not to the news of Queen's speech.

5 Conclusion

This research provides useful insights into the difference of topic contents generated by citizens and news organisations about the COVID-19 pandemic on social media. We explore how disruptive technology can help with the research by utilising an unsupervised text analysis approach, STM to analyse Twitter posts generated by citizens and news organisations respectively in the context of UK. Important findings are identified in this research which exhibits significant contributions on both theoretical and practical implications. Firstly, this study demonstrates the feasibility and effectiveness of using unsupervised machine learning approach, structural topic modelling to gain insights from the large amount of text data on social media. In particular, the procedures and methodologies that are introduced and developed in this research fill the research gap of how the value of unstructured user-generated content can be leveraged for investigating the difference of the general public and news media's response to crises.

Secondly, in line with other studies of Twitter post content analysis (e.g., Abd-Alrazaq et al., 2020), a number of topics on the COVID-19 pandemic are extracted from the STM model. Four categories are found for the tweets posted by news organi-

sations which are COVID-19 update, Advice and measures, Breaking news and Society efforts. In terms of citizen generated tweets, three categories covering Express feelings, Discussion and Sharing activities are identified. From these categories and topic labels, we show that tweets posted by news organisations are all about reporting facts or analysis on COVID-19 related issues, while a large proportion of citizen generated tweets focus on personal feeling expression and sharing personal activities. This finding provides a new evidence on the debate of whether social media are used by the general public for emotional needs or information needs in crises (Glasgow et al., 2016).

Thirdly, the research findings help to clarify the effect of news media on public agenda setting via social media platforms in the pandemic. Previous studies received conflicted conclusion on the relationship between news media agenda and public agenda on social media platforms. According to the agenda setting theory, news media coverage on certain issues are expected to affect the perceived importance of these issues by the public (Tewksbury & Scheufele, 2009). However, studies based on the uses and gratifications theory indicate that individuals and news media have different motivations when using social media (Urista et al., 2009). Our study helps to clarify the conclusion and fills the research gap by revealing that only a small number of topics are found common in news media tweets and citizens' tweets, and most of these common topics do not account for similar shares. This finding suggests that the agendas of news media and general public on social media platforms are not well matched during the pandemic. This finding provides new evidence for both researchers and practitioners to understand how the rise of social media affects the news media's influence on the public's agenda setting during crises.

Fourthly, from observing the variation of the common topics' proportions over time, we find that both news media and citizens show changing patterns of concerns on breaking news. However, the changing patterns are not the same, with citizens' response to breaking news more significant. This finding may indicate that comparing with news media, citizens are more sensitive to key events and the proportion of their posts

on relevant topics increases faster on social media. This partially supports previous findings on tweeting activity peaking after major news stories (Chew & Eysenbach, 2010).

The findings of this study contribute to the theoretical literature on exploring disruptive technology on COVID-19 analysis. The study applies unsupervised machine learning approach in textual data analysis and reveals the interest difference between news media and citizens as well as the way that news organisations use social media in the COVID-19 crisis. They can also be applied to guide practitioners in the news industry on the development of their social media strategies, help policy makers better understand citizens' concerns on the COVID-19 pandemic, and facilitate the use of social media platforms on navigating people's interests and improving public wellbeing. Despite its contributions, this study also has limitations that should be addressed in future research. The Twitter data we applied in this research are from a short period, it could be beneficial if a whole picture of the public and news media's response during the whole pandemic period can be captured. Future study could examine longer time period to investigate the changes in the society response to the pandemic, how the agenda setting behaviours of citizens and news organisations change over time. In addition, the citizen tweets analysed in the study contains tweets posted by other organisations except the news organisations. We did not exclude the tweets posted by other organisations because they help to form a complete picture of how the public respond to the pandemic. It is difficult to distinguish which posts are from organisations considering a large number of organisations are small and operated by individual people and may be used for personal purposes (Armstrong & Gao, 2010). However, looking into the tweets from organisations such as NHS may generate more interesting insights. Also, because this research is focused on Twitter posts written in English in the context of UK, future study could extend the study to more countries and examine the specific language use in tweets which may provide additional insights. Finally, investigating the effect of news media on citizens' discussion is an interesting direction

and is left for future work.

References

Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hamdi, M., & Shah, Z. (2020). Top concerns of tweeters during the covid-19 pandemic: infoveillance study. *J Med Internet Res*, *22*(4), e19016.

Abdel-Basset, M., Chang, V., Hawash, H., Chakrabortty, R. K., & Ryan, M. (2021). Fss-2019-ncov: A deep learning architecture for semi-supervised few-shot segmentation of covid-19 infection. *Knowl Based Syst*, *212*, 106647.

Abdel-Basset, M., Chang, V., & Mohamed, R. (2020). Hsma_woa: A hybrid novel slime mould algorithm with whale optimization algorithm for tackling the image segmentation problem of chest x-ray images. *Appl. Soft Comput.*, *95*, 106642.

Abdel-Basset, M., Chang, V., & Nabeeh, N. A. (2020). An intelligent framework using disruptive technologies for covid-19 analysis. *Technol. Forecast. Soc. Change*, 120431.

An, S.-K., & Gower, K. K. (2009). How do the news media frame crises? a content analysis of crisis news coverage. *Public Relat. Rev.*, *35*(2), 107–112.

Armstrong, C. L., & Gao, F. (2010). Now tweet this: How news organizations use twitter. *E. News*, *4*(4), 218–235.

Austin, L., Fisher Liu, B., & Jin, Y. (2012). How audiences seek out crisis information: Exploring the social-mediated crisis communication model. *J. Appl. Commun. Res.*, *40*(2), 188–207.

Bane, K. C. (2019). Tweeting the agenda: How print and alternative web-only news organizations use twitter as a source. *Journal. Pract.*, *13*(2), 191–205.

Baudier, P., Kondrateva, G., Ammi, C., Chang, V., & Schiavone, F. (2021). Patients' perceptions of teleconsultation during covid-19: A cross-national study. *Technol. Forecast. Soc. Change*, *163*, 120510.

Behairy, N., Mukherjee, S., Venkatesh, B. E., et al. (2006). Technology-based communication patterns of youth. *ACR North American Advances*.

Berliner, D., Bagozzi, B. E., & Palmer-Rubin, B. (2018). What information do citizens want? evidence from one million information requests in mexico. *World Dev.*, *109*, 222–235.

Blazquez, D., & Domenech, J. (2018). Big data sources and methods for social and economic analyses. *Technol. Forecast. Soc. Change*, *130*, 99–113.

Boersma, M., & Graham, T. (2013). Twitter as a news source: how dutch and british newspapers use twitter in their news coverage, 2007–2012. *Journal. Pract.*, *7*(4), 446–464.

Brachten, F., Mirbabaie, M., Stieglitz, S., Berger, O., Bludau, S., & Schrickel, K. (2018). Threat or opportunity?-examining social bots in social media crisis communication. *arXiv preprint arXiv:1810.09159*.

Bruns, A., Highfield, T., & Burgess, J. (2013). The arab spring and social media audiences: English and arabic twitter users and their networks. *Am. Behav. Sci.*, *57*(7), 871–898.

Buheji, M., Ahmed, D., et al. (2020). Planning for'the new normal': Foresight and management of the possibilities of socio-economic spillovers due to covid-19 pandemic. *Bus. Manag. and Strategy*, *11*(1), 160–179.

Büschken, J., & Allenby, G. M. (2016). Sentence-based text analysis for customer reviews. *Mark. Sci.*, *35*(6), 953–975.

Canter, L. (2013). The interactive spectrum: The use of social media in uk regional newspapers. *Convergence*, *19*(4), 472–495.

Chang, V. (2018a). Computational intelligence for medical imaging simulations. *J. Med. Syst.*, *42*(1), 1–12.

Chang, V. (2018b). An overview, examples, and impacts offered by emerging services and analytics in cloud computing virtual reality. *Neural. Comput. Appl.*, *29*(5), 1243–1256.

Chang, V. (2021). An ethical framework for big data and smart cities. *Technol. Forecast. Soc. Change*, *165*, 120559.

Chang, V., Liu, L., Xu, Q., Li, T., & Hsu, C.-H. (2020). An improved model for sentiment analysis on luxury hotel review. *Expert Syst.*, e12580.

Chen, E., Lerman, K., & Ferrara, E. (2020). Covid-19: The first public coronavirus twitter dataset. *arXiv preprint arXiv:2003.07372*.

Chen, G. M. (2011). Tweet this: A uses and gratifications perspective on how active twitter use gratifies a need to connect with others. *Comput. Hum. Behav.*, *27*(2), 755–762.

Chew, C., & Eysenbach, G. (2010). Pandemics in the age of twitter: content analysis of tweets during the 2009 h1n1 outbreak. *PloS one*, *5*(11), e14118.

Chong, D., & Druckman, J. N. (2007). A theory of framing and opinion formation in competitive elite environments. *J. Commun.*, *57*(1), 99–118.

Cinelli, M., Quattrociocchi, W., Galeazzi, A., Valensise, C. M., Brugnoli, E., Schmidt, A. L., ... Scala, A. (2020). The covid-19 social media infodemic. *arXiv preprint arXiv:2003.05004*.

Coombs, W. T. (2006). The protective powers of crisis response strategies: Managing reputational assets during a crisis. *J. Promot. Manag.*, *12*(3-4), 241–260.

Cushion, S., Kilby, A., Thomas, R., Morani, M., & Sambrook, R. (2018). Newspapers, impartiality and television news: Intermedia agenda-setting during the 2015 uk general election campaign. *Journal. Stud.*, *19*(2), 162–181.

Daugherty, T., Eastin, M. S., & Bright, L. (2008). Exploring consumer motivations for creating user-generated content. *J. Interact. Advert.*, *8*(2), 16–25.

Dearing, J. W., Rogers, E. M., & Rogers, E. (1996). *Agenda-setting*. SAGE Publications.

Dickerson, J. P., Kagan, V., & Subrahmanian, V. (2014). Using sentiment to detect bots on twitter: Are humans more opinionated than bots? In *2014 ieee/acm international conference on advances in social networks analysis and mining (asonam 2014)* (pp. 620–627).

Durahim, A. O., & Coşkun, M. (2015). # iamhappybecause: Gross national happiness through twitter analysis and big data. *Technol. Forecast. Soc. Change*, *99*, 92–105.

Engesser, S., & Humprecht, E. (2015). Frequency or skillfulness: How professional news media use twitter in five western countries. *Journal. Stud.*, *16*(4), 513–529.

Fortunato, J. A., & Martin, S. E. (2016). The intersection of agenda-setting, the media environment, and election campaign laws. *J. Inf. Policy*, *6*(1), 129–153.

Francken, N., Minten, B., & Swinnen, J. F. (2009). Media, monitoring, and capture of public funds: evidence from madagascar. *World Dev.*, *37*(1), 242–255.

Gamson, W. A. (1992). *Talking politics*. Cambridge university press.

Ghani, N. A., Hamid, S., Hashem, I. A. T., & Ahmed, E. (2019). Social media big data analytics: A survey. *Comput. Hum. Behav.*, *101*, 417–428.

Glasgow, K., Vitak, J., Tausczik, Y., & Fink, C. (2016). "with your help... we begin to heal": Social media expressions of gratitude in the aftermath of disaster. In *International conference on social computing, behavioral-cultural modeling and prediction and behavior representation in modeling and simulation* (pp. 226–236).

Glowacki, E. M., Lazard, A. J., Wilcox, G. B., Mackert, M., & Bernhardt, J. M. (2016). Identifying the public's concerns and the centers for disease control and prevention's reactions during a health crisis: An analysis of a zika live twitter chat. *Am. J. Infect. Control*, *44*(12), 1709–1711.

Grover, P., Kar, A. K., Dwivedi, Y. K., & Janssen, M. (2019). Polarization and acculturation in us election 2016 outcomes–can twitter analytics predict changes in voting preferences. *Technol. Forecast. Soc. Change*, *145*, 438–460.

Han, C., & Yang, M. (2020). Revealing airbnb user concerns on different room types. *Ann. Tour. Res.*(in press).

Hashem, I. A. T., Chang, V., Anuar, N. B., Adewole, K., Yaqoob, I., Gani, A., ... Chiroma, H. (2016). The role of big data in smart city. *Int. J. Inf. Manage.*, *36*(5), 748–758.

Hermida, A. (2010). Twittering the news: The emergence of ambient journalism. *Journal. Pract.*, *4*(3), 297–308.

Hindman, D. B. (2009). Mass media flow and differential distribution of politically disputed beliefs: The belief gap hypothesis. *Journal. Mass Commun. Q.*, *86*(4), 790–808.

Hu, N., Zhang, T., Gao, B., & Bose, I. (2019). What do hotel customers complain about? text analysis using structural topic model. *Tour. Manag.*, *72*, 417–426.

Karyotis, C., Doctor, F., Iqbal, R., James, A., & Chang, V. (2018). A fuzzy computational model of emotion for cloud based sentiment analysis. *Inf. Sci.*, *433*, 448–463.

Kim, E. H.-J., Jeong, Y. K., Kim, Y., Kang, K. Y., & Song, M. (2016). Topic-based content and sentiment analysis of ebola virus on twitter and in the news. *J. Inf. Sci.*, *42*(6), 763–781.

Kouzy, R., Abi Jaoude, J., Kraitem, A., El Alam, M. B., Karam, B., Adib, E., ... Baddour, K. (2020). Coronavirus goes viral: quantifying the covid-19 misinformation epidemic on twitter. *Cureus*, *12*(3).

Lasorsa, D. L., Lewis, S. C., & Holton, A. E. (2012). Normalizing twitter: Journal. pract. in an emerging communication space. *Journal. Stud.*, *13*(1), 19–36.

Li, K., Zhang, Q., Kwong, S., Li, M., & Wang, R. (2013). Stable matching-based selection in evolutionary multiobjective optimization. *IEEE Trans. Evol. Comput.*, *18*(6), 909–923.

McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. *Public Opin. Q.*, *36*(2), 176–187.

Meraz, S. (2009). Is there an elite hold? traditional media to social media agenda setting influence in blog networks. *J. Comput. Mediat. Commun.*, *14*(3), 682–707.

Messner, M., Linke, M., & Eford, A. (2012). Shoveling tweets: An analysis of the microblogging engagement of traditional news organizations. In *International symposium on online journalism* (Vol. 2, pp. 74–87).

Mirbabaie, M., & Zapatka, E. (2017). Sensemaking in social media crisis communication–a case study on the brussels bombings in 2016. In *Proceedings of the 25th european conference on information systems* (p. 2169-2186).

Missier, P., Romanovsky, A., Miu, T., Pal, A., Daniilakis, M., Garcia, A., ... da Silva Sousa, L. (2016). Tracking dengue epidemics using twitter content classification and topic modelling. In *international conference on web engineering* (pp. 80–92).

Moon, S. J., & Hadley, P. (2014). Routinizing a new technology in the newsroom: Twitter as a news source in mainstream media. *J. Broadcast. Electron. Media*, *58*(2), 289–305.

Murthy, D., & Gross, A. J. (2017). Social media processes in disasters: Implications of emergent technology use. *Soc. Sci. Res.*, *63*, 356–370.

Nagano, H., de Oliveira, J. A. P., Barros, A. K., & Junior, A. d. S. C. (2020). The 'heart kuznets curve'? understanding the relations between economic development and cardiac conditions. *World Dev.*, *132*, 104953.

Panagiotopoulos, P., Barnett, J., Bigdeli, A. Z., & Sams, S. (2016). Social media in emergency management: Twitter as a tool for communicating risks to the public. *Technol. Forecast. Soc. Change*, *111*, 86–96.

Paul, D., Li, F., Teja, M. K., Yu, X., & Frost, R. (2017). Compass: Spatio temporal sentiment analysis of us election what twitter says! In *Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining* (pp. 1585–1594).

Procopio, C. H., & Procopio, S. T. (2007). Do you know what it means to miss new orleans? internet communication, geographic community, and social capital in crisis. *J. Appl. Commun. Res.*, *35*(1), 67–87.

Reuter, C., Hughes, A. L., & Kaufhold, M.-A. (2018). Social media in crisis management: An evaluation and analysis of crisis informatics research. *Int. J. Hum-Comput. Int.*, *34*(4), 280–294.

Roberts, M., Stewart, B., & Tingley, D. (2018). *stm: An r package for the structural topic model.* (`http://www.structuraltopicmodel.com`)

Roberts, M. E., Stewart, B. M., & Airoldi, E. M. (2016). A model of text for experimentation in the social sciences. *J. Am. Stat. Assoc.*, *111*(515), 988–1003.

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., ... Rand, D. G. (2014). Structural topic models for open-ended survey responses. *Am. J. Pol. Sci.*, *58*(4), 1064–1082.

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

Sayre, B., Bode, L., Shah, D., Wilcox, D., & Shah, C. (2010). Agenda setting in a digital age: Tracking attention to california proposition 8 in social media, online news and conventional news. *Policy Internet*, *2*(2), 7–32.

Shaw, F., Burgess, J., Crawford, K., & Bruns, A. (2013). Sharing news, making sense, saying thanks: Patterns of talk on twitter during the queensland floods. *Australian J. Commun.*, *40*(1), 23–40.

Silge, J., & Robinson, D. (2016). tidytext: Text mining and analysis using tidy data principles in r. *J. Open Source Softw.*, *1*(3), 37.

Singh, L., Bansal, S., Bode, L., Budak, C., Chi, G., Kawintiranon, K., ... Wang, Y. (2020). A first look at covid-19 information and misinformation sharing on twitter. *arXiv preprint arXiv:2003.13907*.

Stephens, K. K., Jahn, J. L., Fox, S., Charoensap-Kelly, P., Mitra, R., Sutton, J., ... Meisenbach, R. J. (2020). Collective sensemaking around covid-19: Experiences, concerns, and agendas for our rapidly changing organizational lives. *Manag. Commun. Q.*, *34*(3), 426–457.

Stieglitz, S., Mirbabaie, M., & Milde, M. (2018). Social positions and collective sense-making in crisis communication. *Int. J. Hum-Comput. Int.*, *34*(4), 328–355.

Stokes, D. C., Andy, A., Guntuku, S. C., Ungar, L. H., & Merchant, R. M. (2020). Public priorities and concerns regarding covid-19 in an online discussion forum: longitudinal topic modeling. *J. Gen. Intern. Med.*, *35*(7), 2244–2247.

Sutton, J. N., Palen, L., & Shklovski, I. (n.d.). *Backchannels on the front lines: Emergency uses of social media in the 2007 southern california wildfires.* University of Colorado.

Takahashi, B., Tandoc Jr, E. C., & Carmichael, C. (2015). Communicating on twitter during a disaster: An analysis of tweets during typhoon haiyan in the philippines. *Comput. Hum. Behav.*, *50*, 392–398.

Tewksbury, D., & Scheufele, D. A. (2009). News framing theory and research. In *Media effects* (pp. 33–49). Routledge.

Urista, M. A., Dong, Q., & Day, K. D. (2009). Explaining why young adults use myspace and facebook through uses and gratifications theory. *Hum. Commun.*, *12*(2), 215–229.

Vaughan, D., & Dancho, M. (2018). *furrr: Apply mapping functions in parallel using futures.* (`https://cran.r-project.org/web/packages/furrr/index.html`)

Vis, F. (2013). Twitter as a reporting tool for breaking news: Journalists tweeting the 2011 uk riots. *Digit. Journal*, *1*(1), 27–47.

WHO. (2020). Coronavirus disease 2019 (covid-19): situation report. Retrieved from `https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports`

Xu, Q., Chang, V., & Hsu, C.-H. (2020). Event study and principal component analysis based on sentiment analysis–a combined methodology to study the stock market with an empirical study. *Inf. Syst. Front.*, *22*(5), 1021–1037.

Xue, J., Chen, J., Hu, R., Chen, C., Zheng, C., Su, Y., & Zhu, T. (2020). Twitter discussions and emotions about the covid-19 pandemic: Machine learning approach. *J Med Internet Res*, *22*(11), e20550.

Yang, M., & Han, C. (2019). Stimulating innovation: managing peer interaction for idea generation on digital innovation platforms. *J. Bus. Res.*, *125*, 456-465.

Yang, M., & Han, C. (2021). Revealing industry challenge and business response to covid-19: a text mining approach. *Int. J. Contemp. Hosp. Manag.*(in press).

Yoo, J., Choi, S., Choi, M., & Rho, J. (2014). Why people use twitter: social conformity and social value perspectives. *Online Inf. Rev.*, *38*(2), 265–283.

Author biography

- Dr. Chunjia Han (c.han@gre.ac.uk) is Senior Lecturer in Business Planning and Innovation with over 10 years of experience of working with user innovation, open innovation and crowdsourcing, both as a practitioner and an academic. He has been involved in EU and UK funded projects on innovation management.
- Dr. Mu Yang (m.yang@gre.ac.uk) is Senior Lecturer in Business Strategy and Innovation with expertise in data analytics, digital innovation and user behaviour analysis. She has been working on a number of digital innovation projects supported by European Research Council, EPSRC UK and EU Horizon 2020.
- Dr. Athena Piterou (a.piterou@gre.ac.uk) is Senior Lecturer in Sustainability at the University of Greenwich with expertise in innovation studies and social network analysis. She has published in the area of renewable energy. She has been involved in EU funded projects regarding small business innovation and community renewable energy. More recently she has worked on a Newton Fund project regarding networks in the creative industries.