# Chapter 14
# Text Mining Tweets on Post-COVID-19 Sustainable Tourism: A Social Media Network and Sentiment Analysis

**Dongdong Wu, Hui Li, Yueqing Li, and Yuhong Wang**

**Abstract** The primary purpose of this chapter is to get to know the public attitude towards sustainable tourism after COVID-19 and its polarity or emotion. Using Twitter Archiving Google Sheet, 6718 tweets were collected from July 11 to August 10, 2021, with the hashtags #covid19 and #tourism, #sustainabletourism or #ecotourism or #responsibletourism. Tableau and Gephi were used to visualise and aggregate the social media network. Using R Studio, the word frequency, association and sentiment analysis were carried out. The main findings are as follows: (1) retweets take most of all data; (2) media accounts are more visible and active than individual ones in the community network; (3) the "trust" emotion and "anticipation" emotion are dominant in the tweets. Besides, this chapter also tried to use related social behaviour theories to explain the observed social media user behaviours. Practical implications have also been provided to dissolve people's psychological and emotional problems and enhance people's confidence in tourism recovery.

**Keywords** Post-COVID-19 · Sustainable tourism · Text mining · Social media network · Twitter · Sentiment analysis

D. Wu · H. Li (✉)
College of Tourism and Service Management, Nankai University,
Tianjin, People's Republic of China
e-mail: dwu@mail.nankai.edu.cn; lihuihit@nankai.edu.cn

Y. Li
Department of Industrial and Systems Engineering, Lamar University, Beaumont, TX, USA
e-mail: yueqing.li@lamar.edu

Y. Wang
School of Business, Jiangnan University, Wuxi, Jiangsu, People's Republic of China
e-mail: yuhongwang@jiangnan.edu.cn



K. Dube et al. (eds.), *COVID-19, Tourist Destinations and Prospects for Recovery*, https://doi.org/10.1007/978-3-031-22257-3_14

## 14.1 Introduction

The COVID-19 pandemic has wreaked havoc on the tourism industry, particularly on tourists' fears and risk perceptions (Balasubramanian et al., 2021). In the age of big data, public perceptions towards post-COVID-19 sustainable tourism is an important area of investigation, which could influence the development direction of e-tourism with information and communication technologies (ICTs). For example, public opinions and perceptions could spread rapidly with the help of ICTs and influence actual individual behaviours or community involvement. To explore the recovery of tourism in the post-COVID-19 era, sustainability may be an important research topic (Dube et al., 2021). The unique features of social media are that it connects contributors and audiences through conversations that can occur in real-time (Segerberg & Bennett, 2011). Social media application programming interfaces make it possible to gather data around specific hashtags and keywords and reveal the hashtag public that forms around current issues or events. Twitter, for example, is less controlled than surveys, but it represents a much larger sample and provides an almost real-time index of perception (Weller et al., 2014). With a focus on post-COVID-19 sustainable tourism-related hashtags, all the above features make Twitter the ideal place for real-time collection and new data analysing and conducting real-life research.

However, the data produced by social media, huge in size and noisy, needs to be summarized and converted into interpretable forms so that the information contained can be utilized for practical purposes (Wu et al., 2020). Sentiment analysis (or opinion mining) has been one of the most active fields for natural language processing (NLP) research. Since its inception, the NLP research has expanded significantly (Liu, 2015). Balasubramanian et al. (2021) examined post-COVID-19 tourist concerns using sentiment analysis and topic modelling, but semi-structured survey data could be subject to subjectivity. Besides, automated sentiment analysis has been used to extract public opinion from various sources (Kiritchenko et al., 2014). As a typical user-generated content (UGC), tweets (maximum 140 characters) are short and informal and often contain slang and emotion. Thus, a higher sentiment analysis accuracy for tweets is frequently easier to achieve. In this context, our primary purpose is to get to know the public attitude toward sustainable tourism after COVID-19 and its polarity or emotion.

Despite Twitter's potential as a data source, only very few researchers have researched its use for analysing sustainable tourism after COVID-19. Twitter data can supplement and extend information on big data and tourism research. This chapter contributes to further exploring the tweets mining of post-COVID-19 sustainable tourism through social media networks and sentiment analysis. Specific research questions include the following: (1) Why does a specific type of information spread widely via Twitter? (2) What type of sentiments are expressed towards the topic-related tweets? (3) How do social behaviour theories explain the observed social media user behaviours? In the next two sections, the literature review and methodology are introduced, respectively. Then, the result and discussions are provided regarding the research topic. Finally, we conclude this research and provide policy implications for the recovery of tourism.

## 14.2 Literature Review

### *14.2.1 Study on COVID-19 and Sustainable Tourism*

While the impact of the COVID-19 pandemic has been thoroughly examined, there has been very little attention paid to the social media network and sentimental reaction, particularly on post-COVID-19 sustainable tourism. COVID-19 is widely recognised as a challenge or even a game-changer for travel and tourism (Higgins-Desbiolles, 2020; Nhamo et al., 2020). Serval studies have investigated the sustainable and competitive issues of tourism destinations (Wu & Li, 2021; Wu et al., 2022b), but the situation has changed in the aftermath of the COVID-19 pandemic crisis. Some studies focused on the recovery of tourism-related industries, such as national COVID-19 exit strategies for tourism (Collins-Kreiner & Ram, 2021) and the impact of public health emergencies on hotel demand (He et al., 2022).

Sustainable tourism has been deemed a driving force of the tourism industry in the post- COVID-19 scenario (Palacios-Florencio et al., 2021). Several studies discussed whether a shift of tourism towards a higher level of sustainability or not under the COVID-19 crisis (Romagosa, 2020; Tauber & Bausch, 2022). However, it is rare to find the analysis of post-COVID-19 sustainable tourism-related topics based on social media. Academics have found that social media data is of great value. It reflects what people think, and the data is large enough to access. For example, Kim and Chae (2018) explored the relationship between social media usage and hotel performance using firm-generated tweets. Therefore, text information from social media (i.e. Twitter) can be used to supplement data sources and provides direct insights into public perceptions of post-COVID-19 sustainable tourism.

### *14.2.2 Study on Social Media Network and Sentiment Analysis*

Social networks represent a challenging emerging sector in the context of big data: The natural language expressions of people can be easily reported through short text messages, rapidly creating unique content of huge dimensions that must be efficiently and effectively analysed to create actionable knowledge for decision-making processes (Pozzi et al., 2016). Text mining includes the collection, analysis and study of frequencies of words and recognition of patterns to support visualisation and predictive analytics. It has been applied in a wide variety of fields, such as job advertisements (Pejic-Bach et al., 2020), customer service (Mahr et al., 2019), customer complaints (Hu et al., 2019), tourist perceived experience (Li et al., 2020) and online hotel or travel reviews (Berezina et al., 2016; Guerreiro & Rita, 2020; Hou et al., 2019), amongst others.

In addition, limited research has been conducted using sentiment analysis in sustainable tourism studies in a post-COVID-19 context. Sentiment analysis is an analysis of the information extracted to identify reactions, situations, contexts and emotions (Thelwall, 2019). In this chapter, it refers to the practice of applying NLP and text analysis techniques to identify and extract subjective information from post-COVID-19 sustainable tourism-related tweets. It has been used to analyse short informal texts and monitor emotions, such as on Weibo and WeChat, but data from the Twitter platform (with a wider coverage) has not been considered (Kirilenko et al., 2018; Mohammad et al., 2015). For post-COVID-19 sustainable tourism research, it is crucial to gain important insights from opinions expressed online, particularly from social media blogs (i.e. Tweets).

While the statistical representation can effectively capture information on different parameters in a database, text mining aims to capture and provide a coherent overview of content in various areas (Cambria et al., 2017). In social media studies of hospitality and tourism, Xiang et al. (2017) comparatively explore data quality. Database research frequently suffers from the inability to record public exposure, which can cause an estimate to be inaccurate. Online surveys and questionnaires can gather subjective results, but they cannot reflect the actual sentiment and public perceptions (Qian et al., 2019). Hence, the visualisation of user interaction and mapping of community networks using UGC from social media shall be an exploratory vision.

## 14.3 Methodology

Data collection and preprocessing are necessary initial steps for text mining. Then, we focus on the research topic through social media networks and sentiment analysis. Figure 14.1 explains the methodology and framework of this research, which is designed by the authors.

### *14.3.1 Data Collection*

Twitter Archiving Google Sheet (TAGS) was adopted to collect the Twitter tweets of research interest. Tweets can be sampled in real-time over a predefined period using user-defined hashtag search terms. Sampled from July 11, 2021, to August 10, 2021, we collected 6718 tweets (5987 unique tweets) with a focus on post-COVID-19 sustainable tourism using the hashtags #covid19 and #tourism, #sustainabletourism or #ecotourism or #responsibletourism. Google spreadsheet stores

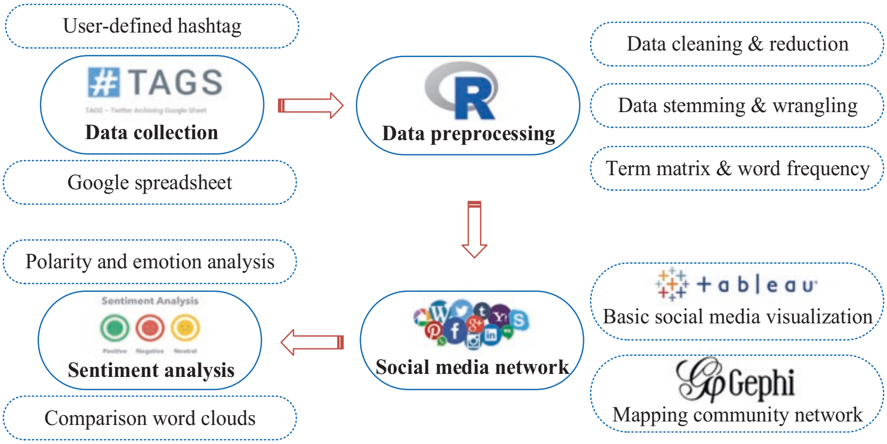


**Fig. 14.1** Research methodology and framework

the following related data: Twitter ID and username, tweet content and generated time, user follower and user-friend count, etc.

### *14.3.2 Data Preprocessing*

In order to transform the unstructured raw text into a format for identifying patterns, preprocessing is necessary (Feldman & Sanger, 2007). R Studio is a free and open-source integrated development environment for R, and its featured 'packages' are utilised for data preprocessing.

**Data cleaning and reduction** The 'tm' package uses the 'Corpus' function to create a corpus (Feinerer & Hornik, 2019). To clean the corpus, unwanted characters (such as '&amp', 'RT|via', 'http\\w+', etc.) were removed. Then, independent terms (words) are created, and tokenization is conducted by splitting text into tokens. The entire text is converted to lowercase, removing URLs, emojis, non-English words, punctuations, numbers and whitespace. With the help of the 'SnowballC' package (Bouchet-Valet, 2019), stop words (e.g., 'the' and 'a') are removed using the stop words dictionary.

**Data stemming and wrangling** Stemming collects similar words in one word to improve the accuracy of the mined text by removing suffixes and reducing words to their fundamental forms. Data wrangling is also carried out, which promoted the process of cleaning up data and removing redundancy. 'tidyr' and 'dplyr' packages, great and simple data wrangling tools, were used to implement this process.

**Term matrix and word frequency** Using the stemmed terms, a document-term matrix (DTM) has been created in which the rows match documents, i.e. tweets, and the columns reflect the terms., i.e. words (Liu, 2015). The DTM describes the frequency of terms that occur in a collection of documents. The frequency of words indicates that from the word that is used most often in the dataset to the one that is least used when compiling DTM occurrences.

### *14.3.3 Social Media Network*

A business data analysis tool, Tableau Desktop Professional Edition, was used to achieve meaningful results from the tweets. The basic statistics of tweets, including tweet content, temporal trends, retweets, tweet reach, user activity, user visibility and secondary hashtags, are examined using Tableau.

Gephi, an open-source network visualisation software, was adopted to map the virtual clusters (Qian et al., 2019; Wu et al., 2020). The number of interactions between users determines the distance between users and followers. The linkages between Twitter users are generated as a spatial map with the help of the Forceatlas2 algorithm, which embeds in the network software (Jacomy et al., 2014).

### *14.3.4 Sentiment Analysis*

The 'syuzhet' package is loaded into R Studio to analyse the polarities, which can be converted into categorical variables 'positive', 'neutral' and 'negative' (Jockers, 2017). The 'get_nrc_sentiment' function in the 'syuzhet' library is used to get an analysis of emotions using the text that is being tested and divide the emotions into eight categories. The function calculates the presence of emotions: 'anger', 'anticipation', 'disgust', 'fear', 'joy', 'sadness', 'surprise', 'trust' and overall 'negative' and 'positive' sentiment scores.

The 'ggplot2' package (Wickham & Chang, 2019), known as the grammar of graphics, is used to further visualize our results from the sentiment analysis. The 'word cloud' package is used to find out the words that were most commonly associated with each emotion, as well as with each polarity. The analysis of independent word associations with the 'findAssocs' function provides information about post-COVID-19 sustainable tourism that goes beyond typical experimental results.

## 14.4 Result and Discussions

### *14.4.1 Social Media Visualization Analysis*

Three types of tweets, i.e. @mention (14.34%), original (31.47%) and retweet (54.19%), are presented in Fig. 14.2 during the study period. *Agenda setting theory* examines why information about some issues, but not others, is available to the

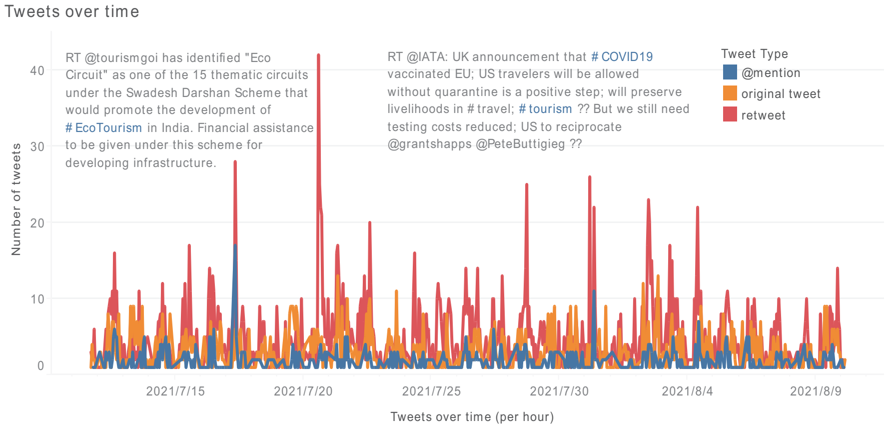


**Fig. 14.2** The temporal trends of tweets per hour

public and how public opinion is changed (McCombs & Valenzuela, 2020). Therefore, mass communication can influence public perceptions of a specific theme. Retweets take most of all data, which ensures increased exposure to valuable information. The fluctuation of these three is synchronous, basically as a whole. Spikes indicate specific events of interest for the topic.

During the study period, the type of 'retweet' is the most, compared with '@ mention' and 'original tweet'. On July 17th, the widely retweeted tweet posts the news that financial support will help ecotourism development in India. On July 20th, another hot retweet says that the United Kingdom (UK) hopes to reduce the costs of COVID-19 vaccination to preserve livelihoods in travel. The above tweet texts are inserted in Fig. 14.2.

Table 14.1 shows the top five messages that are most worthy of sharing, and the range of propagation and most followed users, respectively. Retweet indicates the frequency of retweeting activity on a specific tweet. While tweet reach represents the number of followers that each tweet potentially could have reached. We can find that most tweets appeared prominently in both lists: they have achieved wide reach because they have been retweeted by many users with substantial follower networks.

The tweet texts listed in Table 14.1 highlight the popular topics about post-COVID-19 sustainable tourism that is eco-tourism destination recommendations, concerns about COVID-19 vaccination, suggestions for sustainable tourism and attention to the practice of ecological tourism construction. *Planned behaviour theory* points out that the best way to predict and explain a person's behaviours is through one's behavioural intentions (Ajzen, 1991). These popular topics may be an indication of the way things are going.

As shown in Fig. 14.3, the distribution of the secondary hashtags by day indicates the different preferences and habits of Twitter users over time. The topic of daily tweets during the study period can be represented by several secondary hashtags represented by stacked colour blocks. That is to say, secondary hashtags

**Table 14.1** Top five retweets and tweet reach

| Retweet | Tweet reach | Text |
|---|---|---|
| 98 (1) | 23,382 (3) | RT @user: Odisha, an eco-tourism destination that the nature-lover in you can't afford to miss! #EcoTourism #DekhoApnaDesh |
| 72 (2) | 7483 (5) | RT @user: UK announcement that #COVID19 vaccinated EU; US travellers will be allowed without quarantine is a positive step; will preserve livelihoods in #travel; #tourism?? But we still need testing costs reduced; US to reciprocate |
| 67 (3) | 97,359 (1) | RT @user: We can all take small steps this summer?? to reduce?? and help make #tourism greener! Here are our 5 tips to make your holidays more #sustainable?? #EUClimatePact #MyWorldOurPlanet #sustainabletourism |
| 63 (4) | 41,590 (2) | RT @user: The user has identified "Eco Circuit" as one of the 15 thematic circuits under the Swadesh Darshan Scheme that would promote the development of #EcoTourism in India. Financial assistance to be given under this scheme for developing infrastructure |
| 60 (5) | 12,725 (4) | RT @user: Nakta Pahad in McCluskieganj located on Chhotanagpur Plateau is every traveller's dream pursuit! #EcoTourism #DekhoApnaDesh |

Note: The numbers in parentheses indicate the order of the tweets

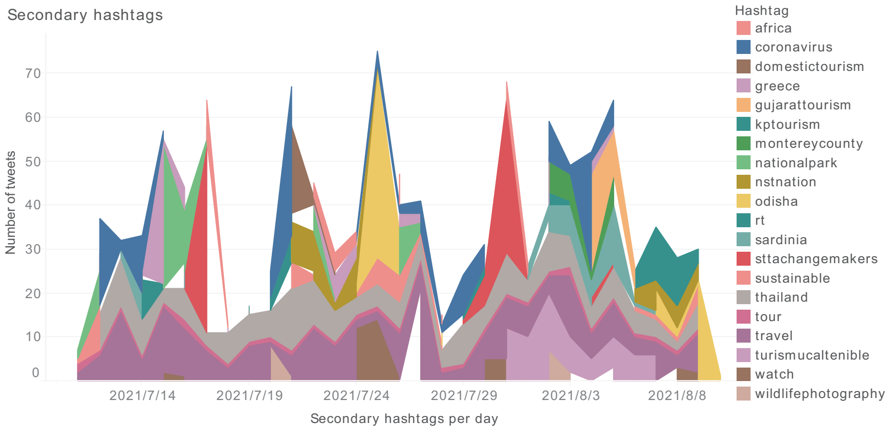


**Fig. 14.3** The distribution of the top 20 secondary hashtags by day

can explain events or topics related to the topic of this research, and the degree to which they are related. As can be seen from Fig. 14.3, the contributions of hashtags # 'travel' and # 'Thailand' to the topic during the study period are continuous. We will dig deeper into this connection in the following paragraphs.

### *14.4.2 Mapping the Community Network*

As shown in Fig. 14.4, the visible and active users are the statistic of the number of tweets from different users. The most followed users may not appear as the most active or most visible, but their large follower numbers may make them influential.

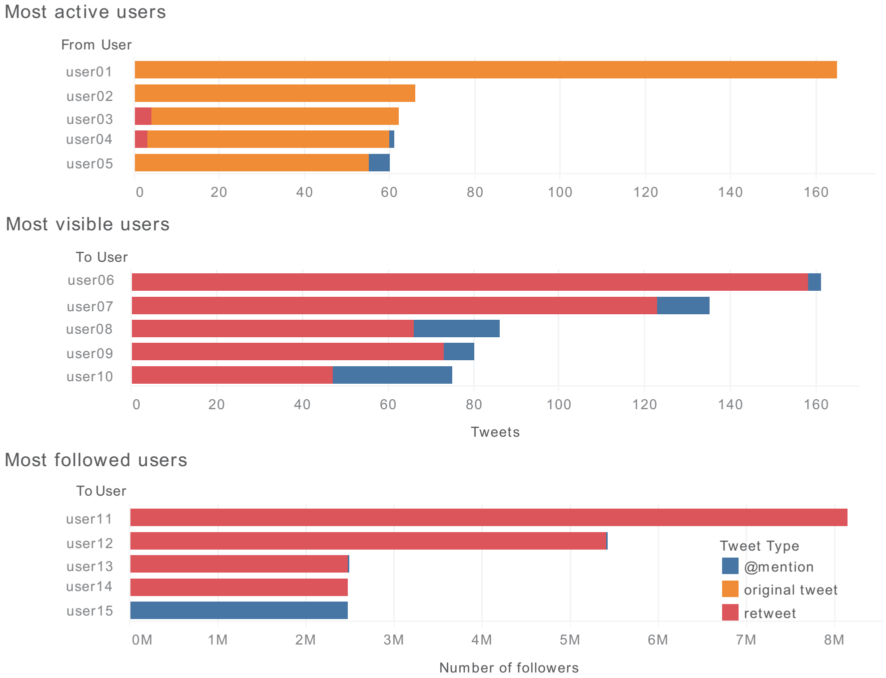


**Fig. 14.4** The most visible, active and followed Twitter users

During the study period, the most active user is ‘user01’, which is a personal account spreading original tweets about travel experiences in Bangkok. User with ID being ‘user06’ is most visible, which is a media account aiming at propagandising the geography, humanity and tourism culture of India. The most followed user is ‘user11’, which is a marketing account focusing on recommending daily doses of travel, style, art, culture, food and healthy living.

*Social network theory* points out that people tend to think and behave similarly because they are connected (Wasserman & Faust, 1994). Figure 14.5 shows some major network trends based on the visualisation of the sustainable tourism community. The dots represent user accounts, and the edges represent the connections. A key concept of social network theory is ‘centrality’ and examines how being at the central point in a social network is the most ‘popular’.

Six major clusters are visualised, which take ‘gtpgr’, ‘traveltomorrow’, ‘sttakenya’, ‘incredibleindia’, ‘gujarattourism’ and ‘odhiamboatieno’ as the core, respectively. These are either media accounts which post news about sustainable tourism or individual users who recommend eco-tourism destinations. These user networks are also embedded in Fig. 14.5, representing different roles in exploring the topic of post-COVID-19 sustainable tourism.

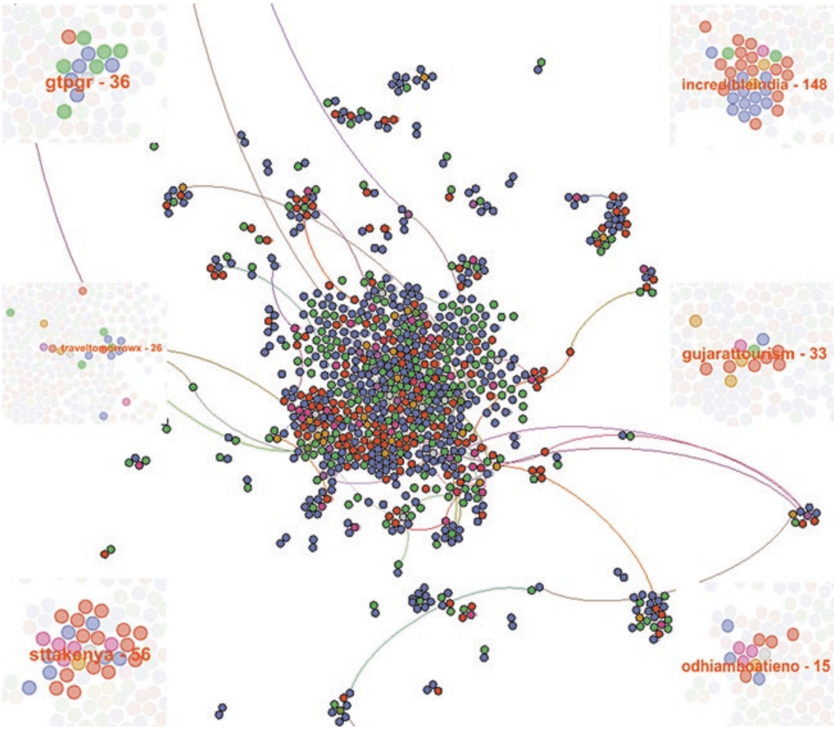


**Fig. 14.5** Sustainable tourism network community map

### *14.4.3 Word Frequency and Association*

The frequency of the word is the times that word is used in the dataset. The left side of Fig. 14.6 shows the word cloud, which represents the frequency of words that appeared top 100 in the tweet texts. The words 'tourism' (3909 times) and 'covid' (3009 times) are highly associated with the defined hashtags directly. Other words like 'ecotourism', 'travel', 'sustainabletourism', 'coronavirus' and 'responsible-tourism' focus on the research topic well.

The association analysis based on the word 'delta' with other terms is carried out to dig deeper into the 'connection' mentioned above. The right side of Fig. 14.6 shows the correlation coefficients ranked top 10. It is evident that the delta variant leads to travel restrictions and an increase in COVID-19 cases in Thailand. The analysis of word frequency and association can help better understand how a pandemic affects public attention and perceived risks. COVID-19 poses a major obstacle to travel, and the following tweets can describe the causality properly.

'Covid cases pass 750,000 as rate dips Sunday.'
'100,000 retailers might cease operations without government support.'
'Thailand recorded over 20,000 Covid-19 cases for the fourth day in a row on Saturday.'

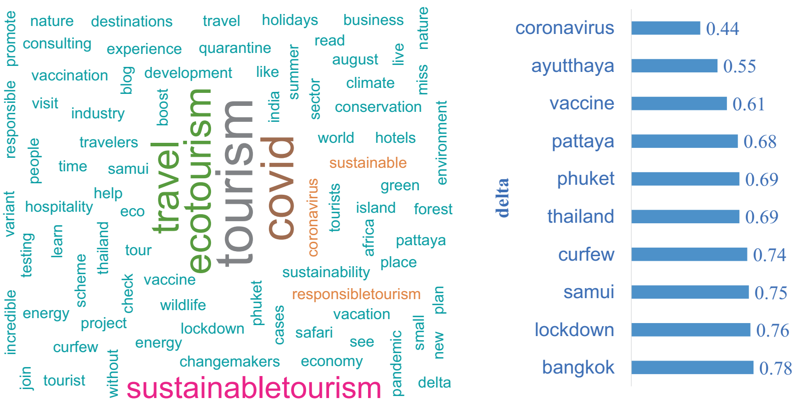


**Fig. 14.6** The word cloud and association of stemmed terms in tweets

#Thailand #Bangkok #Phuket #Samui #Pattaya #Ayutthaya #covid19 #coronavirus #lockdown #vaccine #travel #tourism #delta #curfew

### *14.4.4 Polarity and Emotion Analysis*

Polarity in sentiment analysis refers to identifying sentiment orientation (positive, neutral and negative) in tweet texts. The sentiment score of neutral polarity (1186) is slightly higher than negative polarity (1009), and the sentiment score of positive polarity (4523) is the highest. However, the meaning of the crucial term “neutral sentiment” remains different, and the proper explanations still need to be explored (Liu, 2015).

Emotion analysis with the ‘nrc’ lexicon is shown on the right side of Fig. 14.7. The ‘trust’ emotion and ‘anticipation’ emotion are dominant in the tweets, with a score of 4435 and 4138, respectively. Followed by the ‘joy’ and ‘fear’ emotions with scores of 3461 and 2047, respectively. *Social learning theory* (Bandura & Walters, 1977) believes that individuals can influence each other, and all the observed perceptions (including emotional contagion) from social media users will generate the dominant emotions. Through text mining and emotion analysis, we can perceive that the public’s attitude towards post-COVID-19 sustainable tourism is positive (i.e. trust, anticipation) as a whole.

The dominant words in different categories can give a specific cognitive sentiment intuitively. The words ‘covid’, ‘pandemic’, ‘pollution’ and ‘costs’ are obvious on the left side of Fig. 14.7. The words mentioned above mainly represent the ‘fear’, ‘sadness’ and ‘disgust emotions vividly. In addition, it also reflects the public’s sensitivity and aversion to the bad situation of COVID-19. This illustrates that the vision of sustainable tourism is intertwined with the intermittent spread of COVID-19. Sustainable tourism amid the pandemic will be a long quest.

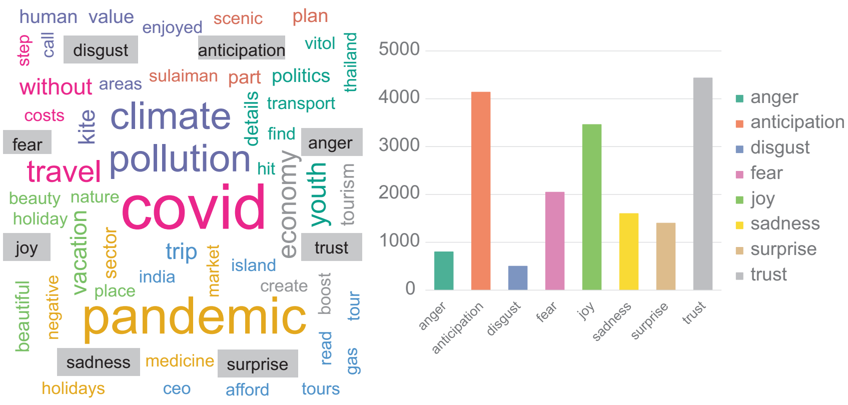


**Fig. 14.7** Sentiment scores and word clouds are associated with each emotion

## 14.5 Conclusion and Policy Implications

### *14.5.1 Conclusion*

COVID-19 has affected the tourism industry in many ways. At the same time, travel and tourism activities have also enabled the pandemic to become a global issue (Dube et al., 2021). This research aimed to understand public perceptions on the trends of post-COVID-19 sustainable tourism and uncover meaningful themes of concerns posted by Twitter users during the post-COVID-19 era. The result also confirms that the sentiment influences people's activities on Twitter, which is also correlated to the news coverage of popular tweets and the daily number of infections (Zhang et al., 2021). In addition, the analysis results can be used as a case to demonstrate that social-behavioural theory can well explain specific social media user behaviours.

This research remained some limitations that can be improved in future work. Firstly, there is necessary to add more related hashtags and lengthen the study period to enrich the tweets' dataset. Besides, how the sentiment score varies across time is also a deserved topic to study. Secondly, there is a need to further explore NLP to understand the semantics of text in different and complex contexts. It is prospective to conduct aspect-based sentiment analysis including sensitivity to context, subjectivity and tone, comparisons, irony and sarcasm. Thirdly, diverse machine learning methods can be applied to learn the correct sentiment value from a set of input features. Combining advanced machine learning techniques to create classifiers and train tweet samples is a burgeoning research direction (Hofmann & Chisholm, 2016).

### 14.5.2 Policy Implications

Since the COVID-19 outbreak, social media has become the most important channel for the public to obtain information (Yu et al., 2021). The exploration of social media networks and the sentiment analysis of the public can provide policy enlightenment for the recovery of tourist destinations. Government departments of tourism destinations should pay attention to collecting and releasing information related to the event and accumulating data that can reflect the emotional response of large-scale groups towards COVID-19 (Hou et al., 2020).

Tourism destination managers should pay attention to the gathering of social media networks and play a leading role to spread positive images of tourism destinations. In the post-COVID-19 era, individuals' emotional responses to the epidemic are crucial to the potential impact of tourism destinations. Further measures should be taken to reduce people's anger and other negative emotions, dissolve people's psychological and emotional problems and enhance people's confidence in tourism recovery. Besides, nudging people to think about accuracy is a simple way to improve choices about what to share on social media (Pennycook et al., 2020).

Social media is not only a channel for releasing important information but also a gathering platform for public opinions (Cinelli et al., 2020). The interactive nature of social media and the sudden generation of epidemic events lead to the rapid spread of public opinions on specific social media. In response to public opinions about the emergence of epidemic events, official institutions or social organizations must actively respond to the public. Through the sound management mechanism for epidemic events to realize the monitoring, early warning and response to social media public opinion can effectively resolve public opinions.

**Acknowledgements** The authors thank editors and reviewers for their helpful comments and suggestions. This work is based on the short abstract presented by the first author at the Ideas Fair of TTRA 2022 Annual International Conference (Wu et al., 2022a) and was supported by the National Natural Science Foundation of China (Nos. 71971124, 71932005); the Liberal Arts Development Fund of Nankai University (No. ZB21BZ0106); the One Hundred Talents Program of Nankai University (No. 63223067) and the Program of Overseas Studies for Postgraduates by Jiangnan University.

## References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Balasubramanian, S., Kaitheri, S., Nanath, K., Sreejith, S., & Paris, C. M. (2021). Examining post-COVID-19 tourist concerns using sentiment analysis and topic modeling. In W. Wörndl, C. Koo, & J. L. Stienmetz (Eds.), *Information and communication technologies in tourism 2021*. Springer, Cham. https://doi.org/10.1007/978-3-030-65785-7_54

Bandura, A., & Walters, R. H. (1977). *Social learning theory*. Prentice-Hall.

Berezina, K., Bilgihan, A., Cobanoglu, C., & Okumus, F. (2016). Understanding satisfied and dissatisfied hotel customers: Text mining of online hotel reviews. *Journal of Hospitality Marketing & Management, 25*(1), 1–24. https://doi.org/10.1080/19368623.2015.983631

Bouchet-Valet, M. (2019). *SnowballC: Snowball stemmers based on the C libstemmer UTF-8 library* (Version 0.6.0). Available at http://CRAN.Rproject.org/package=SnowballC. Last accessed 24 Nov 2019.

Cambria, E., Das, D., Bandyopadhyay, S., & Feraco, A. (2017). *A practical guide to sentiment analysis*. Springer. https://doi.org/10.1007/978-3-319-55394-8

Cinelli, M., Quattrociocchi, W., Galeazzi, A., Valensise, C. M., Brugnoli, E., Schmidt, A. L., Zola, P., Zollo, F., & Scala, A. (2020). The COVID-19 social media infodemic. *Scientific Reports, 10*(1), 1–10. https://doi.org/10.1038/s41598-020-73510-5

Collins-Kreiner, N., & Ram, Y. (2021). National tourism strategies during the Covid-19 pandemic. *Annals of Tourism Research, 89*, 103076. https://doi.org/10.1016/j.annals.2020.103076

Dube, K., Nhamo, G., & Chikodzi, D. (2021). COVID-19 cripples global restaurant and hospitality industry. *Current Issues in Tourism, 24*(11), 1487–1490. https://doi.org/10.1080/13683500.2020.1773416

Feinerer, I., & Hornik, K. (2019). *tm: Text Mining Package* (Version 0.7-7). Available at https://CRAN.R-project.org/package=tm. Last accessed 24 Nov 2019.

Feldman, R., & Sanger, J. (2007). *The text mining handbook: Advanced approaches in analyzing unstructured data*. Cambridge University Press. https://doi.org/10.1017/CBO9780511546914

Guerreiro, J., & Rita, P. (2020). How to predict explicit recommendations in online reviews using text mining and sentiment analysis. *Journal of Hospitality and Tourism Management, 43*(4), 269–272. https://doi.org/10.1016/j.jhtm.2019.07.001

He, L. Y., Li, H., Bi, J. W., Yang, J. J., & Zhou, Q. (2022). The impact of public health emergencies on hotel demand – Estimation from a new foresight perspective on the COVID-19. *Annals of Tourism Research, 94*, 103402. https://doi.org/10.1016/j.annals.2022.103402

Higgins-Desbiolles, F. (2020). The "war over tourism": Challenges to sustainable tourism in the tourism academy after COVID-19. *Journal of Sustainable Tourism, 29*(4), 551–569. https://doi.org/10.1080/09669582.2020.1803334

Hofmann, M., & Chisholm, A. (2016). *Text mining and visualization: Case studies using open-source tools*. CRC Press.

Hou, Z., Cui, F., Meng, Y., Lian, T., & Yu, C. (2019). Opinion mining from online travel reviews: A comparative analysis of Chinese major OTAs using semantic association analysis. *Tourism Management, 74*, 276–289. https://doi.org/10.1016/j.tourman.2019.03.009

Hou, Z., Du, F., Jiang, H., Zhou, X., & Lin, L. (2020). *Assessment of public attention, risk perception, emotional and behavioural responses to the COVID-19 outbreak: Social media surveillance in China*. Available at SSRN, https://doi.org/10.2139/ssrn.3551338

Hu, N., Zhang, T., Gao, B., & Bose, I. (2019). What do hotel customers complain about? Text analysis using structural topic model. *Tourism Management, 72*, 417–426. https://doi.org/10.1016/j.tourman.2019.01.002

Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS One, 9*(6), e98679. https://doi.org/10.1371/journal.pone.0098679

Jockers, M. (2017). *syuzhet: Extracts sentiment and sentiment-derived plot arcs from text* (Version 1.0.4). Available at https://CRAN.R-project.org/package=syuzhet. Last accessed 24 Nov 2019.

Kim, W.-H., & Chae, B. K. (2018). Understanding the relationship among resources, social media use and hotel performance: The case of Twitter use by hotels. *International Journal of Contemporary Hospitality Management, 30*(9), 2888–2907. https://doi.org/10.1108/IJCHM-02-2017-0085

Kirilenko, A. P., Stepchenkova, S. O., Kim, H., & Li, X. R. (2018). Automated sentiment analysis in tourism: Comparison of approaches. *Journal of Travel Research, 57*(8), 1012–1025. https://doi.org/10.1177/0047287517729757

Kiritchenko, S., Zhu, X., & Mohammad, S. M. (2014). Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research, 50*, 723–762. https://doi.org/10.1613/jair.4272

Li, C., Li, H., Liu, Y., & Liang, S. (2020). A research on the differences of inbound tourist perceived experience based on text mining. *Nankai Business Review, 23*(1), 28–39. (in Chinese).

Liu, B. (2015). *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press. https://doi.org/10.1017/CBO9781139084789

Mahr, D., Stead, S., & Odekerken-Schröder, G. (2019). Making sense of customer service experiences: A text mining review. *Journal of Services Marketing, 33*(1), 88–103. https://doi.org/10.1108/JSM-10-2018-0295

McCombs, M., & Valenzuela, S. (2020). *Setting the agenda: Mass media and public opinion*. Wiley.

Mohammad, S. M., Zhu, X., Kiritchenko, S., & Martin, J. (2015). Sentiment, emotion, purpose, and style in electoral tweets. *Information Processing & Management, 51*(4), 480–499. https://doi.org/10.1016/j.ipm.2014.09.003

Nhamo, G., Dube, K., & Chikodzi, D. (2020). *Counting the cost of COVID-19 on the global tourism industry*. Springer Nature. https://doi.org/10.1007/978-3-030-56231-1

Palacios-Florencio, B., Santos-Roldán, L., Berbel-Pineda, J. M., & Castillo-Canalejo, A. M. (2021). Sustainable tourism as a driving force of the tourism industry in a post-COVID-19 scenario. *Social Indicators Research, 158*(3), 991–1011. https://doi.org/10.1007/s11205-021-02735-2

Pejic-Bach, M., Bertoncel, T., Meško, M., & Krstić, Ž. (2020). Text mining of industry 4.0 job advertisements. *International Journal of Information Management, 50*, 416–431. https://doi.org/10.1016/j.ijinfomgt.2019.07.014

Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological Science, 31*(7), 770–780. https://doi.org/10.1177/0956797620939054

Pozzi, F. A., Fersini, E., Messina, E., & Liu, B. (2016). *Sentiment analysis in social networks*. Elsevier Science. Available at https://www.perlego.com/book/1830588/sentiment-analysis-in-social-networks-pdf

Qian, C., Li, Y., & Wang, Y. (2019). Text mining tweets on driving safety and cellphone use. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 63*(1), 1537–1540. https://doi.org/10.1177/1071181319631464

Romagosa, F. (2020). The COVID-19 crisis: Opportunities for sustainable and proximity tourism. *Tourism Geographies, 22*(3), 690–694. https://doi.org/10.1080/14616688.2020.1763447

Segerberg, A., & Bennett, W. L. (2011). Social media and the organization of collective action: Using Twitter to explore the ecologies of two climate change protests. *The Communication Review, 14*(3), 197–215. https://doi.org/10.1080/10714421.2011.597250

Tauber, V., & Bausch, T. (2022). Will COVID-19 boost sustainable tourism: Wishful thinking or reality? *Sustainability, 14*(3), 1686. https://doi.org/10.3390/su14031686

Thelwall, M. (2019). Sentiment analysis for tourism. In M. Sigala, R. Rahimi, & M. Thelwall (Eds.), *Big data and innovation in tourism, travel, and hospitality*. Springer. https://doi.org/10.1007/978-981-13-6339-9_6

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge University Press. https://doi.org/10.1017/CBO9780511815478

Weller, K., Bruns, A., Burgess, J., Mahrt, M., & Puschmann, C. (2014). *Twitter and society*. Peter Lang Publishing Group. Available at https://eprints.qut.edu.au/66321/

Wickham, H., & Chang, W. (2019). *ggplot2: Create elegant data visualisations using the grammar of graphics* (Version 3.2.1). Available at https://CRAN.R-project.org/package=ggplot2. Last accessed 24 Nov 2019.

Wu, D., & Li, H. (2021). Benchmark analysis of tourism sustainable development using meta-frontier data envelopment analysis. In *Proceedings of the 3rd international Tourism and Retail Service Management Conference (TRMC 2021)*. Sunway University, Kuala Lumpur, Malaysia.

Wu, D., Li, Y., & Wang, Y. (2020). *Text mining public perception on autonomous driving*. Available at SSRN, https://doi.org/10.2139/ssrn.4062130

Wu, D., Li, H., Li, Y., & Wang, Y. (2022a). Text mining tweets on post COVID-19 sustainable tourism through social media network and sentiment analysis. Travel and Tourism ResearchAssociation: Advancing Tourism Research Globally (TTRA 2022). Available at https://scholarworks.umass.edu/ttra/2022/researchabstract/57/

Wu, D., Li, H., & Wang, Y. (2022b). Measuring sustainability and competitiveness of tourism destinations with data envelopment analysis. *Journal of Sustainable Tourism.* https://doi.org/10.1080/09669582.2022.2042699

Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management, 58*, 51–65. https://doi.org/10.1016/j.tourman.2016.10.001

Yu, M., Li, Z., Yu, Z., He, J., & Zhou, J. (2021). Communication related health crisis on social media: A case of COVID-19 outbreak. *Current Issues in Tourism, 24*(19), 2699–2705. https://doi.org/10.1080/13683500.2020.1752632

Zhang, Q., Yi, G. Y., Chen, L.-P., & He, W. (2021). *Text mining and sentiment analysis of COVID-19 tweets*. Available at arXiv, https://arxiv.org/abs/2106.15354

**Dongdong Wu** is a PhD candidate at the College of Tourism and Service Management, Nankai University, China. He is also a short-term visiting student at the Department of Industrial and Systems Engineering, Lamar University. His research focuses on sustainable tourism, tourism firm management, text mining and event study.

**Hui Li** is a Young Chang-Jiang Professor and Associate Dean at the College of Tourism and Service Management, Nankai University, China. His research focuses on tourism big data mining and prediction, tourism firm management and governance.

**Yueqing Li** is an Associate Professor at the Department of Industrial and Systems Engineering, Lamar University, USA. His research focuses on big data analytics, human factors and ergonomics, human-computer interaction and neuroergonomics.

**Yuhong Wang** is a Professor at the Department of Management Science and Engineering, School of Business, Jiangnan University, China. His research focuses on big data management and application, uncertainty prediction and decision-making, and grey system theory.