

Revealing industry challenge and business response to Covid-19: a text mining approach

Abstract

Purpose

This study aims to conduct a ‘real-time’ investigation with user-generated content on Twitter to reveal industry challenges and business responses to the Covid-19 pandemic. Specifically, using the hospitality industry as an example, the study analyses how Covid-19 has impacted the industry, what are the challenges and how the industry has responded.

Design/methodology/approach

With 94,340 tweets collected between October 2019 and May 2020 by a programmed web scraper, unsupervised machine learning approaches such as structural topic modelling are applied.

Findings

The results show that: (1) despite the adverse consequences from the pandemic, the hospitality industry has shown increasing interests in finding ways to survive, such as looking into novel technologies and adopting new business strategies; (2) the pandemic has created an opportunity for organisations to jump out from their daily business operations and rethink about the future development of the industry; (3) the Covid-19 impact is not only shown on the reduction in the job demand but also a change in the demand structure of the job market; (4) the use of novel text mining approaches on unstructured social media data is effective in identifying industry-level challenge and response to public emergencies.

Originality

This study contributes to the literature on business response during crises providing for the first time a study of utilising unstructured content on social media for industry-level analysis in the hospitality context.

Keywords: Covid-19, business challenge, hospitality industry, topic modelling, social media, user-generated content

1 Introduction

The coronavirus disease 2019 (Covid-19) has had a major impact on the economy. It is found that six weeks of social distancing would bring GDP down by 5.6% (Barrot et al., 2020). Although the service demand in specific sectors such as healthcare has dramatically increased during the pandemic, many other sectors such as hospitality, tourism and airline have seen the demand dropped down dramatically (del Rio-Chanona et al., 2020). In order to navigate the economic disruptions resulting from Covid-19 and get business survive and revive in a post-Covid-19 world, governments have been trying hard to provide support to business. For example, the UK government launched the “eat out to help out” discount (BBC News, 2020). At the same time, organisations and industry leaders have suggested innovative approaches for business to manage and organise key resources (Budhwar & Cumming, 2020; Verbeke, 2020). A number of industries have implemented novel approaches (e.g., virtual property viewing in the real estate industry) to ensure the safe return of service provision and started to re-design business strategies (e.g., menu redesign in the hospitality industry) in order to stabilise business and promote consumer trust and confidence (Rivera, 2020). A study by Shankar (2020) pointed that “this is a key moment for the management to look at challenges and opportunities in the business in the short to medium term”; indeed, the development of new business strategies needs to be built on a deep understanding of industry challenges and opportunities caused by Covid-19 (Budhwar & Cumming,

2020; Fong et al., 2020).

Researchers have developed and refined many methods to investigate crisis impact and industry challenges. The most common methods rely on case studies, interviews, or surveys (Timoshenko & Hauser, 2019). This set of methods have been implemented in previous studies on a wide range of global crises such as the 2003 severe acute respiratory syndrome (SARS) outbreak, the 2008 global economic crisis, and the 2015 Middle East Respiratory Syndrome (MERS) outbreak (see Chien & Law, 2003; A. Lee & Cho, 2016). Further to the research about Covid-19, traditional methods continued to be applied to play a main role in studying business challenge and response. For example, in Bartik et al. (2020)'s research, they conducted a survey of more than 5,800 small businesses to investigate the impact of Covid-19 on small business outcomes and how small businesses had been adjusting to the pandemic. Moreover, working on the survey data collect from 512 Small and medium-sized enterprises (SMEs), Le et al. (2020) evaluated how policy related factors affect the survival and development of SMEs during the pandemic.

While most of prior studies rely on data collected from surveys, it is worth to pay attention to the issues associated with this approach. Traditional methods such as surveys are normally expensive and time-consuming, as it takes time to design, distribute questionnaires and collect responses (Timoshenko & Hauser, 2019). Thus, findings and conclusions are sometimes captured in delay (Stamolampros et al., 2019). This could be a big issue for studies about crisis such as the Covid-19 pandemic, as they emphasise research timeliness in order to support organisations on fast response and decision-making (Shankar, 2020). Moreover, the data collection from surveys could be biased as participants may be reluctant to speak the truth (Stamolampros et al., 2019). Therefore, it becomes necessary to develop a novel approach to undertake a 'real-time' investigation on crisis impact and business response.

With the development of digital technologies, massive amounts of unstructured data are generated every moment by individuals and organisations on social media

platforms such as Twitter and Facebook. Most importantly, these are ‘real-time’ data which can be applied to disclose the latest information and opinions raised by a crowd of people and organisations (Chae, 2015; Marlowe et al., 2017). While previous studies on crises have started to investigate how to tap into and leverage the value of user-generated content, most of them developed their research from the sociology angles. Prior studies are interested in the topics such as information spread and detecting people’s attitudes regarding crises (Alshaabi et al., 2020; Kouzy et al., 2020). Few studies have investigated how organisations and industries utilise this ‘real-time’ data to analyse business challenge and response in crisis period. One important reason is pointed out by Terpstra et al. (2012): although the information being shared on social media during crises is valuable, filtering and understanding the information content of messages is highly challenging due to their unstructured text formation and massive size (Kwon et al., 2020). The development of machine learning approaches addresses the challenges and makes tapping the value of user-generated data on social media possible. Therefore, by applying unsupervised machine learning approaches, this research aims to present the value of unstructured user-generated data on emergency research and the complementary benefits they offer for both academic research and business practice.

As one of the most severely affected industries by the pandemic (Rivera, 2020), the hospitality industry is examined in this study. In particular, this study aims to answer the following research questions:

RQ1: How to utilise unstructured textual data collected from social media platform to investigate ‘real-time’ industry-level response to public emergencies?

RQ2: How to extract topics and analyse the challenges and response from the hospitality industry to the Covid-19 pandemic?

RQ3: What are the challenges, concerns and response to the Covid-19 pandemic in the hospitality industry?

In this study, we develop a Python-based web scraper which collected 141,516 Twitter posts between October 1, 2019 and May 31, 2020 using hashtags #hospitalityin-

dustry and #hospitality and keyword hospitality industry. The collected dataset represents a large amount of intelligence regarding the discussion about the hospitality industry. Unsupervised machine learning approaches including the structural topic modelling (M. E. Roberts et al., 2014) and network analysis (Blei & Lafferty, 2006) are applied to our dataset and are utilised to extract the topics from the Twitter discussion, depict the correlation among topics and analyse the changing trend of the topics hidden in the unstructured big data.

2 Overview of relevant literature

The pandemic has negatively affected every single industry around the world, among which hospitality industry is one of the most severely affected industries by the outbreak (Jones & Comfort, 2020). Travel restriction, lockdown and stay-at-home orders restrict people’s mobility by rigid enforcement, which directly caused the severe disruption on the business of the hospitality industry (Dolnicar & Zare, 2020). By collecting primary data from surveys, a number of studies have been developed with aims to understand the Covid-19 impact and provide practice suggestions to the hospitality industry. Through a questionnaire survey of 23 hotels in China with 1,594 collected responses, Zhang et al. (2020) examined the effect of hotel safety leadership on employee safety behaviour during the COVID-19 pandemic. Moreover, Jia et al. (2019) investigated the influence of socially responsible human resource management on employee fears of external threats during the COVID-19 outbreak based on a survey of 408 employees in Chinese hospitality and tourism firms. With a total of 170 U.S. adult consumers recruited through Amazon Mechanical Turk, Huang & Liu (2020) examined the effectiveness of donation appeals containing warmth focused versus competence-focused messages in hospitality CSR marketing during the COVID-19 pandemic.

There are a number of studies attempted to tap into the value of unstructured data on social media, however, all of them focus on the understanding of customers’

opinions and behaviours in the epidemic. For example, Yu et al. (2020) reviewed 10,132 online comments related to Covid-19 and identified the key themes mostly discussed by tourists are risk perception and public health.

Previous research suggest that social media platforms such as Twitter are effective and efficient in terms of analysing issues related to crisis events (Abd-Alrazaq et al., 2020; Alshaabi et al., 2020). Further to the research specifically on Covid-19, Ahmed et al. (2020) performed social network analysis and content analysis of Twitter data from a 7-day period to investigate the drivers of the 5G conspiracy theory and strategies to deal with such misinformation. Rufai & Bunce (2020) utilised content analysis to explore the role of Twitter as adopted by the Group of Seven (G7) world leaders in response to Covid-19. By using 14 different trending hashtags and keywords related to the Covid-19 pandemic, Kouzy et al. (2020) analysed the magnitude of misinformation spread on Twitter. Except for investigating information spread, Twitter is a good option for organisational and industrial related research as most professionals and organisations own Twitter accounts which have been widely used for posting information and communication (Vis, 2013). Various opinions and discussions can be observed on Twitter which helps to set up a more comprehensive view (Bigné et al., 2019; Malhotra et al., 2012). Further, because the posts on Twitter are frequently updated and rapidly disseminated, Twitter textual data can be applied to reflect ‘real-time’ opinions and concerns raised by people and organisations (Chae, 2015; Marlowe et al., 2017).

Moreover, a few studies have demonstrated the suitability of using Twitter textual data to identify discussion topics and concerns which are raised by people and organisations in a certain industry. For example, in Chae (2015)’s research, the topics which are widely discussed and shared in the supply chain industry are extracted through a content analysis on Twitter posts and metadata collected using the hashtag #supplychain. A few studies in the hospitality industry have started to tap into the value of Twitter textual data. Most of them are developed based on the application of sentiment analysis: Philander & Zhong (2016) created a sentiment index and

demonstrated the application of sentiment analysis on Twitter data to build low-cost and real-time measures of customer perceptions in the hospitality industry. Kirilenko & Stepchenkova (2017) provided a comprehensive description of the public discourse about Sochi 2014 Winter Olympics on Twitter and found the differences in pre- and post-Games sentiment towards issues surrounding the Olympics. Guercini et al. (2014) processed 67,953 publicly shared tweets to identify customer sentiment about the services of four airline companies. However, extant studies mainly focus on using Twitter data to understand customers' concerns and demands such as (Ćurlin et al., 2019) which looks into the tourism industry. To the best of our knowledge, few study has investigated how to extract topics from Twitter posts in the hospitality context and how to gain insights about industry-related concerns. This study aims to fill the research gap by utilising text mining approaches to analyse the challenges and response from the hospitality industry to the Covid-19 pandemic.

3 Data and methods

3.1 Sample and preprocessing

To collect the tweets discussing hospitality-related topics and events, we programmed a web scraper using Python. The scraper can take keyword and hashtag searches as input for specifying which tweets to collect. By conducting a series of keyword and hashtag searches, we find hashtags #hospitality, #hospitalityindustry and keyword hospitality industry are the most prevalent ones used on Twitter, we thus use them in our data collection. By setting language as English and posting date between October 1, 2019 and May 31, 2020 with a focus on the first pandemic wave, a sample of 141,516 tweets are collected by our scraper. Although the coronavirus was reported and officially named at the end of December 2019, we collected data from October 2019 to understand what people talked about before the pandemic; then, we are able to compare the

topic changes in people’s discussions due to the pandemic. Moreover, no geographical limitation was applied to the data collection in order to investigate our research problem from a global picture.

After removing retweets, the final dataset contains 94,340 tweets. It is worth noting that the dataset contains the tweets posted by ordinary users and also users who own business accounts. Although the study focuses on business response to the pandemic, we did not remove the tweets posted by ordinary users in our analysis. The reason is that ordinary users’ reactions and discussions are also important intelligence for understanding the industry challenge as well as how they feel about the business response made to the pandemic in the hospitality industry. Table 1 presents the summary statistics of the dataset.

The collected dataset include the tweets text content and their metadata, such as tweets created time, hashtags, whether a tweet has media, whether a tweet has been replied, URLs, retweet count, likes count, follower count, user id, user screen name, etc. The text of each tweet is considered as a *document* and the overall tweets are seen as a whole *corpus* which is a collection of individual documents. We applied a set of common natural language processing techniques widely used by text analysis research (e.g., Berliner et al., 2018) to pre-process the documents, making them compatible with the unsupervised text mining techniques used below. Our pre-processing steps include (i) normalisation, (ii) stemming and (iii) character removing. More specifically, words in the tweets text content are 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 are stemmed by removing affixes, followed by the character removing step where stopwords, numbers and punctuation are removed. After the pre-processing, a corpus with 94,340 processed tweets documents and 47,259 unique words (i.e., vocabulary) is finalised.

3.2 Model set-up

Topic models, as one stream of unsupervised machine learning approach has been found suitable for analysing user-generated content, such as customer online reviews (Büschken & Allenby, 2016). This stream of 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 Structural Topic Modelling (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 the twitter posts, e.g., tweets created time, likes count, retweet count, whether a tweet has media, whether a tweet had been replied, etc. Figure 1 presents the plate diagram of STM and visualises the process for text analysis. The key process of STM can be summaries as follows (M. E. Roberts et al., 2014).

The STM is a hierarchical model in which a document d 's prevalence of each topic (denoted by θ_d) is drawn from a logistic-normal distribution whose mean is a function of document covariates X_d . Here a document d represents a tweet in our study.

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

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 where n is the index of each word in the document d :

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

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

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

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

$$\text{prevalence} \sim \text{likes count} + s(\text{tweets created time}),$$

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

Next, we decide on the number of topics K which is an important user-specified 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 . By quantitatively evaluating the model outputs using three criteria: (i) the semantic coherence of topics, (ii) held-out likelihood and (iii) residuals across a range of different topic numbers, an optimal number of topics for our corpus is found in the range of $K = [30, 40]$. We follow the extant applications (Berliner et al., 2018) on guarding against overfitting and by applying the one standard error rule (Reinsel, 2003), $K = 30$ is selected.

3.3 Topics

Our model identifies 30 topics which best characterise the content of 94,340 tweets during the eight-month period between October 1, 2019 and May 31, 2020. Each topic represents an underlying word distribution where every word in our documents is given a probability of assignment to that topic, as described in the previous section. The model outputs a list of top words for each topic which are the words that have the highest probability of appearing in the topic but least probability in other topics. Two extra researchers in hospitality research were recruited in our project for topic label assignment. By looking at the top words and close reading of a substantial number of tweets examples for each topic, the labels were discussed and assigned by the research team. We report these labels, along with each topic’s proportion and top seven words in Table 2.

3.4 Network analysis

To examine the relationship between the topics identified by the STM analysis, we conduct a network analysis which has been applied in a number of content analysis studies for modelling the connections between discussion topics and categories based on words co-occurrence (e.g., Ibrahim & Wang, 2019). Network analysis is defined as a set of techniques to depict the relations among entities and can help researchers to analyse the structures that emerge from the recurrence of these relations (Chiesi, 2015). In this study, network analysis is performed to detect the correlation between the identified topics and to investigate the strength of the correlations. The analysis is based on the results generated from our STM model. More specifically, we computed the Pearson correlation of the maximum a posteriori probability (MAP) estimates for the document-topic proportions θ which yields the marginal correlation of the variational distribution (P. M. Lee, 1989). By setting a threshold as 0.01, the edge between two topics is dropped if the correlation coefficient falls below the threshold.

Figure 2 illustrates the network analysis result, wherein the nodes represent the identified topics, and the edges indicate the correlation. The size of nodes represents the topic proportion with the biggest node Topic 9 Covid-19 impact, and the smallest node Topic 10 on furniture-related discussion. The length of the edges represents the strength of the correlation: the shorter the edge, the stronger the correlation. If two topics are not connected, the correlation is below the threshold 0.01. Three main topic clusters are formed by grouping together topics with correlation value over 0.4.

4 Results and analysis

In this section, we firstly overview the identified topics and their connections, then provide greater detail on the topic clusters and analyse the changes in topic proportions over time affected by the Covid-19 pandemic.

As shown in Figure 2, 19 topics are connected with Topic 30 Generic industry discussion which contains tweets on the discussion about the hospitality industry in general. An example tweet in this topic is presented as follows.

“Looking to learn more about the best industry in the world? Take a few moments to download, like, and share The Hospitality Spirit podcast. #Travel #Tourism #Hospitality #Podcasting”

Except for Topic 30, the rest topics are more specific, related to certain areas, such as Topic 24 on the opening of new hotels, Topic 29 about catering jobs opportunities. Moreover, most of them are grouped into three clusters based on their correlation with other topics. Topic 28 IIHM-related is not connected with any other topic. The tweets in this topic are found related to a hotel management institute (called IIHM) in India.

The topic cluster on the left side of Figure 2 is the biggest cluster, including eight topics which account for 34.2% of all tweets in our sample dataset. This cluster is highly related to the Covid-19 pandemic as two of the eight topics are specifically

about the impact of Covid-19 (i.e., Topic 9) and how to cope with the pandemic (i.e., Topic 16). Interestingly, the cluster at the top of Figure 2 groups together the topics on job advertisements and food service. Six topics fall into this cluster and account for 22.5% of all tweets. The last cluster which is positioned on the right side of Figure 2 contains six topics and is related to hospitality event hosting and planning. This cluster accounts for 18.1% of all tweets in our sample dataset. In order to show the variation in the prevalence of each topic over time, we plot the prevalence of each topic as a smooth function of tweets' created time and present the results in Figure 3.

The following sections describe the three topic clusters and several selected topics in greater detail. We devote particular attention to the topics which had a big change in their prevalence between October 1, 2019 and May 31, 2020.

4.1 Topic cluster related to Covid-19

A recent study assessed the impact of the Covid-19 pandemic on the hospitality industry through interpretation of a human rights lens (Baum & Hai, 2020). The study shows that people's rights to participate in hospitality and tourism have been challenged as never before due to the events including borders closing, curtailment of travel, closure of tourism facilities, and rescinding of consumer protection rights relating to areas such as refunds, compensation and insurance (Murray, 2020). Yu et al. (2020) have observed an increase in the communication of hospitality related issues discussed on social media. Our results confirm the findings. More specifically, Topic 9 - Covid-19 impact is detected as the largest topic, accounting for 7.3% of all tweets. It primarily relates to the discussion on how the Covid-19 pandemic had been influencing the hospitality industry. It is found that this topic has a huge and dramatic increase in its proportion starting from mid January and reached over 10% of proportion in April, indicating people's increasing interests and concerns on the Covid-19 pandemic in hospitality contexts.

Further, as shown in Figure 2, Topic 9 is a central node which connects to other five topics directly within the cluster. These five topics are Topic 8: Sector changes, Topic 11: Technology trend, Topic 16: Coping with Covid-19, Topic 18: Sector support, and Topic 20: Advice and experience sharing for success. An example of each of the five topics are presented as follows.

[Topic 8 Sector changes] “We’re in the midst of a cultural change about women in leadership – it’s an on-going negotiation on many levels, and that’s challenging for women building careers and their managers. – Peggy Berg of @CastellProject #hospitality”

[Topic 11 Technology trend] “How #AugmentedReality Is Revolutionizing the #Hospitality Industry (From Inside Out)”

[Topic 16 Coping with Covid-19] “Get caught up on all @CornellSHA C&I webinars at... Watch for insights on how the #hospitality industry is Coping with Covid-19, and what industry leaders are doing to move forward with hope and determination.”

[Topic 18 Sector support] “A number of support measures are being administered by local authorities, including #business rate relief for #retail, #hospitality and #leisure businesses, and a £10,000 grant for businesses receiving Small Business Rate Relief”

[Topic 20 Advice and experience sharing for success] “If increasing revenue for your hotel is your primary concern, then improving #guestexperience should be your foremost focus. Read more to know how it helps you: ... #GuestLoyalty #hoteliers #Hotelogix #HotelManagement #Hospitality”

It is found that four of these five topics (i.e., Topic 8, 11, 16, and 18) which are closely connected with Topic 9 Covid-19 impact also show an upward trend over time (see Figure 3). More specifically, the proportion of Topic 8 Sector changes did not

change much at the beginning but started increasing from March and has been keeping increasing. This finding is in line with recent studies which suggest that discussions are around the changes in hospitality industry due to the pandemic (Yu et al., 2020).

Topic 11 Technology trend and Topic 18 Sector support presented an upward trend between January and early April but then went downwards slightly between April and May, and then went upwards again from late May. In terms of Topic 16 Coping with Covid-19, as shown in our analysis, the changing pattern in its topic proportion matched with the changing pattern of Topic 9 Covid-19 impact: the proportion reached its highest (i.e., almost 5.0%) quickly in March, but started going downwards from May. This is in line with the finding of Tsionas (2020), who finds that as the closure of tourism facilities and restaurants posed great difficulty for the industry, business owners and industry workers tried to seek ways to reduce the negative impact and receive governmental subsidies.

Topic 24 New hotel opening which is connected to the cluster but sits outside of the cluster, accounts for 5.3% of all tweets in our sample dataset. It has been trending downwards from late December, while another topic - Topic 14 Travel booking which has direct connection with Topic 24 only experienced a downward trend between mid December and March and has been keeping increasing strongly since April. The tweets in Topic 14 are primarily hotel and restaurant booking advertisements. , for example:

“Recapture your cancelled bookings with GuestTalk and promote your staycation to guests, giving them something to look forward to. See how we can help you#hospitality #staycation #travel

The results of the prevalence change in Topic 24 (New hotel opening) and 14 (Travel booking) may suggest that as influenced by the Covid-19 pandemic, fewer new hotels were launched, but hotels and restaurants were posting more advertisements when the lockdown policy was eased in some countries in order to attract more customers and save their business.

4.2 Topic cluster related to employment and food services

This is the second largest topic cluster, containing seven topics, and are notable for their focus on job opportunities in hospitality industry (i.e., Topic 5 Manager job ads, Topic 7 General job ads, Topic 29 Catering job ads) and food services (i.e., Topic 3 Restaurant survival guide, Topic 13 Food delivery, Topic 25 Sports event). Topic 23 Hotel for sale is included in the cluster but has a correlation with only Topic 7 General job ads.

As analysed by Dube et al. (2020), sit-in guests in restaurants dropped to zero in many countries because of the social distancing initiatives, movement restrictions and lockdowns, which led to an unprecedented loss of employment and revenue. Topic 5, 7 and 29 are all related to job opportunities, where Topic 5 focuses on Manager-level jobs accounting for 4.5% of all tweets, Topic 7 focuses on various levels of jobs (such as hotel receptionist and customer service assistant) accounting for 1.7% of all tweets and Topic 29 is about catering jobs with topic proportion of 4.6%. Topic 7 and 29 have been showing a downward trend since the outbreak of the Covid-19. Moreover, the percentage of the decrease is greater for Topic 29 Catering job ads compared with Topic 7 General job ads, indicating that the demand of catering jobs had been decreasing more significantly. Chetty et al. (2020) observed similar situations where businesses laid off most of their low-income employees during the pandemic through analysing anonymised data from private companies. In addition, our analysis finds the decrease in Topic 7's proportion has been slowing down from mid April. Interestingly, the prevalence of Topic 5 Manager job ads shows a complete different changing pattern: its topic proportion started going upwards from February and reached its highest in May, accounting for 6.3% of all tweets. It might suggest that although the Covid-19 pandemic has a negative influence on the hospitality job market, its impact on the job demand at manager level is different.

Apart from Topic 7 (General job ads) and 29 (Catering job ads), Topic 5 (Manager

job ads) is correlated with Topic 13 Food delivery which discussed the food delivery services provided by restaurants, pubs and cafes. An example of Topic 13 is presented in the following.

“Pizza Punks of Glasgow now Delivering with Sweetheat Heated Delivery Bags #sweetheatheateddeliverybags #homedelivery #takeaway #fastfood #restaurant #hospitality #chef #pizzeria #glasgow #pizza #scotlandpic”

As shown in Figure 3, the proportion of the topic has been increasing since December 2019 over the whole period of Covid-19 pandemic, indicating a great interest in food delivery services. This finding confirms the assumption raised in Gössling et al. (2020)’s study, which estimated that the easing of social distancing might benefit take-away and fast food business over fine-dining restaurants. Unlike Topic 13, Topic 26 Drinks which is also in the cluster and accounts for 2.6% of all tweets in our sample dataset, shows a downward trend over the examination period.

The rest topics in this cluster are Topic 3 Restaurant survival guide and Topic 23 Hotel for sale, accounting for 1.3% and 2.4% respectively of all tweets in our sample datasets. Most tweets in these two topics are related to Topic 30 Generic industry discussion but focus more on how restaurants could survive during the crisis and the news on what hotels are for sale. Two representative examples of Topic 3 and 23 are presented as follows.

[Topic 3 Restaurant survival guide] “With survival on the line, restaurants may rethink financing - #restaurants #hospitality #restaurantbusiness”

[Topic 23 Hotel for sale] “Hopefully the industry will be open soon and new owners can be found. Three Scottish hotels up for sale amid ‘dramatic’ impact of Covid-19 #stocktakerguy #hospitality #hospitalityproud #unitedwestand #hotels”

In terms of the change in their prevalence over time, Topic 3 showed an upward trend from November 2019 to February 2020, and then started going downwards until

late May. However, Topic 23 has a relatively slow downward trend, but had an increase in the topic proportion in May 2020.

4.3 Topic cluster related to event planning and hosting

The study by Chetty et al. (2020) suggested that the Covid-19 pandemic affected all parts of the hospitality value chain. Moreover, cancelling or postponing events such as major conferences, festivals and sports, and bans on gatherings of people over certain numbers had been immediately felt by the hospitality industry (Gössling et al., 2020). The third cluster relates to event planning, hosting and special day celebration. As shown in Figure 2, it consists of six topics: Topic 22 Event hosting, Topic 2 Event planning, Topic 12 Special day celebration, Topic 25 Sports event, Topic 19 Booking offers and Topic 21 Business running tips. Among all these topics, Topic 22 and 12 are connected with over three other topics within the cluster while the rest topics have only one or two connections.

Interestingly, the prevalence of Topic 22 (Event hosting), 12 (Event planning) and 19 (Booking offers) trended downwards over the examination period between October 2019 and May 2020, in line with the findings presented by Gössling et al. (2020). Moreover, the percentage of the decrease is greater between mid January and late May, indicating that the Covid-19 pandemic further affected events hosting. With regard to Topic 2 (Event planning), 21 (Business running tips) and 25 (Sports event), their prevalence fluctuated over the examination period but showed a downward trend during the Covid-19 pandemic. The results suggest that the topics in this cluster received less interests on Twitter during the examination period, which aligns with the findings of Chetty et al. (2020).

In addition, Topic 12 Special day celebration is correlated with Topic 17 Guest experience which sits outside the cluster. It is found that both Topic 17 and its connected topic - Topic 4 Guest praise have been experiencing a decrease in their topic

prevalence, indicating fewer discussions were on guest experience as well as the services provided by hotels and restaurants. This might be caused by the closure of hotels and restaurants due to the lockdown policy introduced by many countries.

5 Discussion and conclusions

5.1 Conclusions

This study aims to conduct a ‘real-time’ investigation with user-generated content on Twitter to identify the main concerns in the hospitality industry during the Covid-19 pandemic, and gain a better understanding of the business challenges and responses from the industry. With 94,340 tweets collected between October 2019 and May 2020 by a programmed web scraper, unsupervised machine learning approaches such as structural topic modelling are applied in this research. The analysis results demonstrate the feasibility and effectiveness of applying the structural topic modelling approach to Twitter unstructured data, and disclose a number of novel findings which have important meanings for both hospitality industry and academia.

5.2 Theoretical implications

This research exhibits significant contributions on theoretical implications in the research area of hospitality and tourism management. Firstly, it demonstrates the feasibility and effectiveness of using unsupervised machine learning approach, structural topic modelling to gain insights from the large amount of user-generated content in hospitality research. In particular, the procedures and methodologies that are introduced and developed in this research fill the research gap of how the value of unstructured ‘real-time’ data can be leveraged for investigating the industry’s response to crises.

Secondly, this study answers a number of key research questions regarding the effects of the Covid-19 pandemic on the hospitality industry. Our results show that the

information and discussions about the topic Covid-19 impact have increased significantly during the examination period, indicating a rise in the concerns on how the pandemic has been affecting the hospitality industry. We further find that four topics (i.e., Coping with Covid-19, Technology trend, Sector changes, and Sector support) which are closely connected to the topic of Covid-19 impact also exhibited an upward trend during the examination period. These results extend on current studies on how the hospitality industry has been responding to the pandemic (Martins et al., 2020; Kim et al., 2020) and what efforts have been made for overcoming the challenges caused by the pandemic (Rivera, 2020). In particular, increasing attention were paid to how novel technologies (e.g., Virtual reality, Internet of Things) could be used to solve industry challenges. This result is consistent with the finding of Jiang & Wen (2020) which argued that to thrive post-Covid-19, hospitality business need to gain new strengths and have plan to compete in a new world.

Thirdly, this study extends the conclusions of current studies by providing new insights. In particular, complicated results are found about the impact of Covid-19 on job market in the hospitality industry. It is widely believed that the Covid-19 outbreak severely affected the hospitality job demand (Baum & Hai, 2020). Moreover, an industry investigation estimates there could be tens of millions of people lost their jobs and the negative impact may continue even after the end of the pandemic (McKinsey & Company, 2020). Evidences are found in our research to support the conclusion of the negative impact of Covid-19 on job market. In particular, we find that the proportions of the topics of Catering job ads and General job ads have reduced significantly since the outbreak of Covid-19. However, the proportion of Topic 5 Manager job ads exhibited a rapid growth starting from the beginning of the Covid-19 pandemic, which creating new insights on the impact of the pandemic on employment (Chetty et al., 2020; Guerrieri et al., 2020; Montenovo et al., 2020).

We further investigated the reason that might trigger the growth, and find that this topic is closely connected to the topic of Food delivery which also experienced

a significant growth during the examination period. This finding may indicate that the increase in the demand of manager-level jobs might be related to the growth of food delivery business. As restaurants and hotels are temporarily closed following the lockdown and stay-at-home orders adopted by many countries (Gössling et al., 2020), a large number of business in the hospitality industry have changed their strategies and started to provide food delivery services. The rapid change of business strategy may bring management challenges particularly for the restaurants and hotels which did not have management experience on food delivery service. Therefore, it becomes necessary for them to enhance their management capability by recruiting staff with management experience.

5.3 Practical implications

Firstly, this study provides solutions for hospitality industry practitioners to transform social media data into useful business insights. For the case of Covid-19 pandemic analysed in the research, the findings help the practitioners to understand what the industry has focused on and what are the opportunities for the industry with regards to the pandemic.

Secondly, this study detects significant change of focus in the hospitality industry. More specifically, we find that people and organisations have switched their attention from the topics which are more related to daily operations to the topics focusing on the macro level of the hospitality industry. For example, the discussion on the topics of Interior design and Guest experience have reduced on Twitter since the Covid-19 outbreak, while the tweets related to the topics of Technology trend and Sector changes increased at the same time. These findings may indicate that the impact of Covid-19 has pushed people and organisations in hospitality industry to pay more attention to the macro-level topics instead of the micro-level ones. The result deepens our understanding of the role of crises on creating opportunities for industries to rethink

about the future development (Melvin & Taylor, 2009; Devece et al., 2016).

Thirdly, this study discloses other areas in the hospitality industry which have been affected by the Covid-19 pandemic, and identifies evidence of industry recovery. In particular, the information and discussions about topics such as New hotel opening, Event hosting and Event planning have dramatically reduced on Twitter since the Covid-19 outbreak. Furthermore, we find that the topics such as Marketing news and Travel booking started trending upwards from May after the significant reduction in their topic proportions between the beginning of the Covid-19 outbreak (i.e., January) and April 2020. This finding could be considered as an evidence showing recovery signs of the hospitality industry from the pandemic.

5.4 Limitations and future research

Despite the valuable insights obtained by the study, there are several limitations. We did not look at the types of twitter account which may generate different types of information and discussion. Moreover, the use of twitter data alone cannot capture a whole picture of business response from the industry to the pandemic. This research can be extended to analyse different twitter accounts and incorporate qualitative research involving stakeholders of the industry as well as customers to investigate the Covid-19 impact more comprehensively. Finally, due to the rise in global hospitality and tourism, examining the specific language use in tweets and linking it with the identified topics may provide additional insights.

References

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

- Ahmed, W., Vidal-Alaball, J., Downing, J., & Seguí, F. L. (2020). Covid-19 and the 5g conspiracy theory: social network analysis of twitter data. *Journal of Medical Internet Research*, *22*(5), e19458.
- Alshaabi, T., Minot, J., Arnold, M., Adams, J. L., Dewhurst, D. R., Reagan, A. J., ... Dodds, P. S. (2020). How the world's collective attention is being paid to a pandemic: Covid-19 related 1-gram time series for 24 languages on twitter. *arXiv preprint arXiv:2003.12614*.
- Barrot, J.-N., Grassi, B., & Sauvagnat, J. (2020). Sectoral effects of social distancing. *Available at SSRN*.
- Bartik, A. W., Bertrand, M., Cullen, Z. B., Glaeser, E. L., Luca, M., & Stanton, C. T. (2020). *How are small businesses adjusting to covid-19? early evidence from a survey* (Tech. Rep.). National Bureau of Economic Research.
- Baum, T., & Hai, N. T. T. (2020). Hospitality, tourism, human rights and the impact of covid-19. *International Journal of Contemporary Hospitality Management*.
- BBC News. (2020). *Chancellor gives diners 50% off on eating out*. (<https://www.bbc.co.uk/news/business-53337170>)
- Berliner, D., Bagozzi, B. E., & Palmer-Rubin, B. (2018). What information do citizens want? evidence from one million information requests in mexico. *World Development*, *109*, 222–235.
- Bigné, E., Oltra, E., & Andreu, L. (2019). Harnessing stakeholder input on twitter: A case study of short breaks in spanish tourist cities. *Tourism Management*, *71*, 490–503.

- Blei, D., & Lafferty, J. (2006). Correlated topic models. *Advances in neural information processing systems*, 18, 147.
- Budhwar, P., & Cumming, D. (2020). New directions in management research and communication: Lessons from the covid-19 pandemic. *British Journal of Management*, 31(3), 441.
- Büschken, J., & Allenby, G. M. (2016). Sentence-based text analysis for customer reviews. *Marketing Science*, 35(6), 953–975.
- Chae, B. K. (2015). Insights from hashtag# supplychain and twitter analytics: Considering twitter and twitter data for supply chain practice and research. *International Journal of Production Economics*, 165, 247–259.
- Chetty, R., Friedman, J. N., Hendren, N., Stepner, M., et al. (2020). *How did covid-19 and stabilization policies affect spending and employment? a new real-time economic tracker based on private sector data* (Tech. Rep.). National Bureau of Economic Research.
- Chien, G. C., & Law, R. (2003). The impact of the severe acute respiratory syndrome on hotels: a case study of hong kong. *International journal of hospitality management*, 22(3), 327–332.
- Chiesi, A. M. (2015). Network analysis.
- Ćurlin, T., Jaković, B., & Miloloža, I. (2019). Twitter usage in tourism: Literature review. *Business Systems Research Journal*, 10(1), 102–119.
- del Rio-Chanona, R. M., Mealy, P., Pichler, A., Lafond, F., & Farmer, D. (2020). Supply and demand shocks in the covid-19 pandemic: An industry and occupation perspective. *arXiv preprint arXiv:2004.06759*.

- Devece, C., Peris-Ortiz, M., & Rueda-Armengot, C. (2016). Entrepreneurship during economic crisis: Success factors and paths to failure. *Journal of Business Research*, 69(11), 5366–5370.
- Dolnicar, S., & Zare, S. (2020). *Coronavirus and airbnb: Disrupting the disruptor*. SocArXiv. doi: 10.31235/osf.io/t9n6q
- Dube, K., Nhamo, G., & Chikodzi, D. (2020). Covid-19 cripples global restaurant and hospitality industry. *Current Issues in Tourism*, 1–4.
- Fong, L. H. N., Law, R., & Ye, B. H. (2020). Outlook of tourism recovery amid an epidemic: Importance of outbreak control by the government. *Annals of Tourism Research*, 102951.
- Gössling, S., Scott, D., & Hall, C. M. (2020). Pandemics, tourism and global change: a rapid assessment of covid-19. *Journal of Sustainable Tourism*, 1–20.
- Guercini, S., Misopoulos, F., Mitic, M., Kapoulas, A., & Karapiperis, C. (2014). Uncovering customer service experiences with twitter: the case of airline industry. *Management Decision*.
- Guerrieri, V., Lorenzoni, G., Straub, L., & Werning, I. (2020). *Macroeconomic implications of covid-19: Can negative supply shocks cause demand shortages?* (Tech. Rep.). National Bureau of Economic Research.
- Huang, H., & Liu, S. Q. (2020). “donate to help combat covid-19!” how typeface affects the effectiveness of csr marketing? *International Journal of Contemporary Hospitality Management*.

- Ibrahim, N. F., & Wang, X. (2019). A text analytics approach for online retailing service improvement: Evidence from twitter. *Decision Support Systems, 121*, 37–50.
- Jia, X., Liao, S., Van der Heijden, B. I., & Guo, Z. (2019). The effect of socially responsible human resource management (srhrm) on frontline employees' knowledge sharing. *International Journal of Contemporary Hospitality Management*.
- Jiang, Y., & Wen, J. (2020). Effects of covid-19 on hotel marketing and management: a perspective article. *International Journal of Contemporary Hospitality Management*.
- Jones, P., & Comfort, D. (2020). The covid-19 crisis and sustainability in the hospitality industry. *International Journal of Contemporary Hospitality Management*.
- Kim, J., Kim, J., Lee, S. K., & Tang, L. R. (2020). Effects of epidemic disease outbreaks on financial performance of restaurants: Event study method approach. *Journal of Hospitality and Tourism Management, 43*, 32–41.
- Kirilenko, A. P., & Stepchenkova, S. O. (2017). Sochi 2014 olympics on twitter: Perspectives of hosts and guests. *Tourism Management, 63*, 54–65.
- 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).
- Kwon, W., Lee, M., & Back, K.-J. (2020). Exploring the underlying factors of customer value in restaurants: A machine learning approach. *International Journal of Hospitality Management, 91*, 102643.

- Le, H., Nguyen, T., Ngo, C., Pham, T., & Le, T. (2020). Policy related factors affecting the survival and development of smes in the context of covid 19 pandemic. *Management Science Letters*, *10*(15), 3683–3692.
- Lee, A., & Cho, J. (2016). The impact of epidemics on labor market: identifying victims of the middle east respiratory syndrome in the korean labor market. *International journal for equity in health*, *15*(1), 196.
- Lee, P. M. (1989). *Bayesian statistics*. Oxford University Press London:.
- Li, K., Zhang, Q., Kwong, S., Li, M., & Wang, R. (2013). Stable matching-based selection in evolutionary multiobjective optimization. *IEEE Transactions on Evolutionary Computation*, *18*(6), 909–923.
- Malhotra, A., Malhotra, C. K., & See, A. (2012). How to get your messages retweeted. *MIT Sloan Management Review*, *53*(2), 61–66.
- Marlowe, B., Brown, E. A., Schrier, T., & Zheng, T. (2017). Beverage bloggers: A developing relationship between wine blogger expertise and twitter followers. *International Journal of Hospitality Beverage Management*, *1*(1), 2.
- Martins, A., Riordan, T., & Dolnicar, S. (2020). A post-covid-19 model of tourism and hospitality workforce resilience.
- McKinsey, & Company. (2020). *The near-term impact of coronavirus on workers*. Retrieved April 6, 2020, from <https://www.mckinsey.com/industries/public-sector/our-insights/lives-and-livelihoods-assessing-the-near-term-impact-of-covid-19-on-us-workers?>

- Melvin, M., & Taylor, M. P. (2009). The global financial crisis: Causes, threats and opportunities. introduction and overview. *Journal of International Money and Finance*, 28(8), 1243–1245.
- Montenovo, L., Jiang, X., Rojas, F. L., Schmutte, I. M., Simon, K. I., Weinberg, B. A., & Wing, C. (2020). *Determinants of disparities in covid-19 job losses* (Tech. Rep.). National Bureau of Economic Research.
- Murray, W. C. (2020). Shifting motivations: A longitudinal study of preferred job rewards in the canadian lodging industry. *International Journal of Hospitality & Tourism Administration*, 1–25.
- Philander, K., & Zhong, Y. (2016). Twitter sentiment analysis: Capturing sentiment from integrated resort tweets. *International Journal of Hospitality Management*, 55(2016), 16–24.
- Reinsel, G. C. (2003). *Elements of multivariate time series analysis*. Springer Science & Business Media.
- Rivera, M. (2020). Hitting the reset button for hospitality research in times of crisis: Covid19 and beyond. *International journal of hospitality management*, 87.
- 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. *Journal of the American Statistical Association*, 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. *American Journal of Political Science*, *58*(4), 1064–1082.
- Rufai, S. R., & Bunce, C. (2020). World leaders' usage of twitter in response to the covid-19 pandemic: a content analysis. *Journal of Public Health*.
- Shankar, K. (2020). The impact of covid-19 on it services industry-expected transformations. *British Journal of Management*, *31*(3), 450.
- Stamolampros, P., Korfiatis, N., Chalvatzis, K., & Buhalis, D. (2019). Job satisfaction and employee turnover determinants in high contact services: Insights from employees' online reviews. *Tourism Management*, *75*, 130–147.
- Terpstra, T., Stronkman, R., de Vries, A., & Paradies, G. L. (2012). Towards a realtime twitter analysis during crises for operational crisis management. In *Iscram*.
- Timoshenko, A., & Hauser, J. R. (2019). Identifying customer needs from user-generated content. *Marketing Science*, *38*(1), 1–20.
- Tsionas, M. G. (2020). Covid-19 and gradual adjustment in the tourism, hospitality, and related industries. *Tourism Economics*, 1354816620933039.
- Vaughan, D., & Dancho, M. (2018). *furrrr: Apply mapping functions in parallel using futures*. (<https://cran.r-project.org/web/packages/furrr/index.html>)
- Verbeke, A. (2020). Will the covid-19 pandemic really change the governance of global value chains? *British Journal of Management*, *31*(3), 444.

- Vis, F. (2013). Twitter as a reporting tool for breaking news: Journalists tweeting the 2011 uk riots. *Digital journalism*, 1(1), 27–47.
- Yu, M., Li, Z., Yu, Z., He, J., & Zhou, J. (2020). Communication related health crisis on social media: a case of covid-19 outbreak. *Current Issues in Tourism*, 1–7.
- Zhang, J., Xie, C., Wang, J., Morrison, A. M., & Coca-Stefaniak, J. A. (2020). Responding to a major global crisis: the effects of hotel safety leadership on employee safety behavior during covid-19. *International Journal of Contemporary Hospitality Management*.

Table 1: Summary statistics of tweets sample

	# of tweets	Percentage
Whether a tweet has media		
Yes	46,792	49.60%
No	47,548	50.40%
How many likes a tweet had received		
0	52,894	56.07%
1	18,101	19.19%
2	8,294	8.79%
3	4,592	4.87%
4	2,887	3.06%
5	1,728	1.83%
>5	5,844	6.19%
How many times a tweet had been retweeted		
0	64,814	68.70%
[1, 10]	29,070	30.81%
[11, 50]	429	0.45%
[51, 100]	13	0.01%
>100	14	0.01%
The length of tweets (# of characters)		
(18, 100]	2,444	2.59%
(100, 200]	27,245	28.88%
(200, 300]	47,969	50.85%
(300, 400]	15,796	16.74%
>400	886	0.94%

Table 2: Topic summary: 94,340 tweets between October 1, 2019 and May 31, 2020

Topic No.	Topic Label	Topic Proportions	Top Words
Topic 9	Covid-19 impact	7.3%	covid, coronavirus, pandem, hospit, will, covid-, hoteli
Topic 15	Marketing news	5.8%	hospit, new, market, brand, podcast, blog, live
Topic 20	Advice and experience sharing for success	5.3%	can, custom, way, help, technolog, use, busi
Topic 24	New hotel opening	5.3%	hotel, resort, success, hostel, hoteli, entrepreneur, luxuri
Topic 13	Food delivery	5.2%	restaur, deliveri, hospit, food, pub, foodservic, cafe
Topic 18	Sector support	4.9%	hospit, busi, sector, retail, award, leisur, govern
Topic 29	Catering job ads	4.6%	career, job, cater, look, recruit, full, chef
Topic 5	Manager job ads	4.5%	job, manag, open, latest, hire, check, click
Topic 17	Guest experience	4.4%	hospit, guest, experi, welcom, stay, love, servic
Topic 12	Special day celebration	4.3%	thank, day, work, team, year, love, happi
Topic 22	Event hosting	4.0%	hospit, great, event, meet, follow, venu, forward
Topic 28	IIHM-related	3.7%	chef, hospit, food, iihm, iihmhotelschool, iihmbestyear, cater
Topic 6	Solution and training	3.5%	learn, manag, train, solut, student, career, join
Topic 21	Business running tips	3.4%	hospit, week, next, want, start, tip, give
Topic 14	Travel booking	3.3%	travel, holiday, tourism, data, vacat, book, insight
Topic 27	Interior design	3.2%	design, hospit, interiordesign, project, interior, light, architectur
Topic 19	Booking offers	3.1%	offer, today, book, contact, visit, profession, call
Topic 16	Coping with Covid-19	3.0%	need, support, pleas, help, let, save, local
Topic 8	Sector changes	2.6%	hospit, consum, will, chang, launch, health, continu
Topic 26	Drinks	2.6%	resourc, bar, free, drink, get, wine, cocktail
Topic 23	Hotel for sale	2.4%	hospit, sale, locat, just, clean, citi, nation
Topic 25	Sports event	2.1%	good, news, intern, season, sport, premier, beyond
Topic 4	Guest praise	1.8%	hospit, perfect, qualiti, suit, weekend, inn, enjoy
Topic 11	Technology trend	1.8%	industri, trend, impact, interest, transform, iot, smart
Topic 7	General job ads	1.7%	group, outdoor, ireland, compass, cateringjob, rvpark, campground
Topic 30	Generic industry discussion	1.6%	look, take, come, like, care, futur, readi
Topic 3	Restaurant survival guide	1.3%	restaurantbusi, staff, wait, guid, edinburgh, medic, cornwal
Topic 2	Event planning	1.2%	eventprof, realli, happen, what, tonight, eventplann, forget
Topic 1	Pakistan/Caribbean food	1.1%	total, pakistan, caribbean, gorgeous, die, accorhotel, infoappli
Topic 10	Furniture	0.8%	hospit, furnitur, cafe, horeca, decorcom, piec, customfurnitur

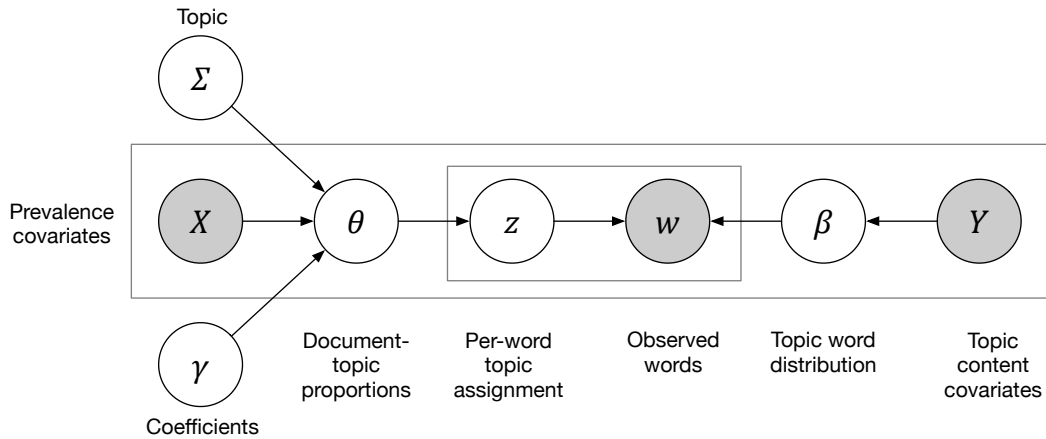


Figure 1: A graphical illustration of STM using plate notation (Adopted by M. E. Roberts et al., 2016)

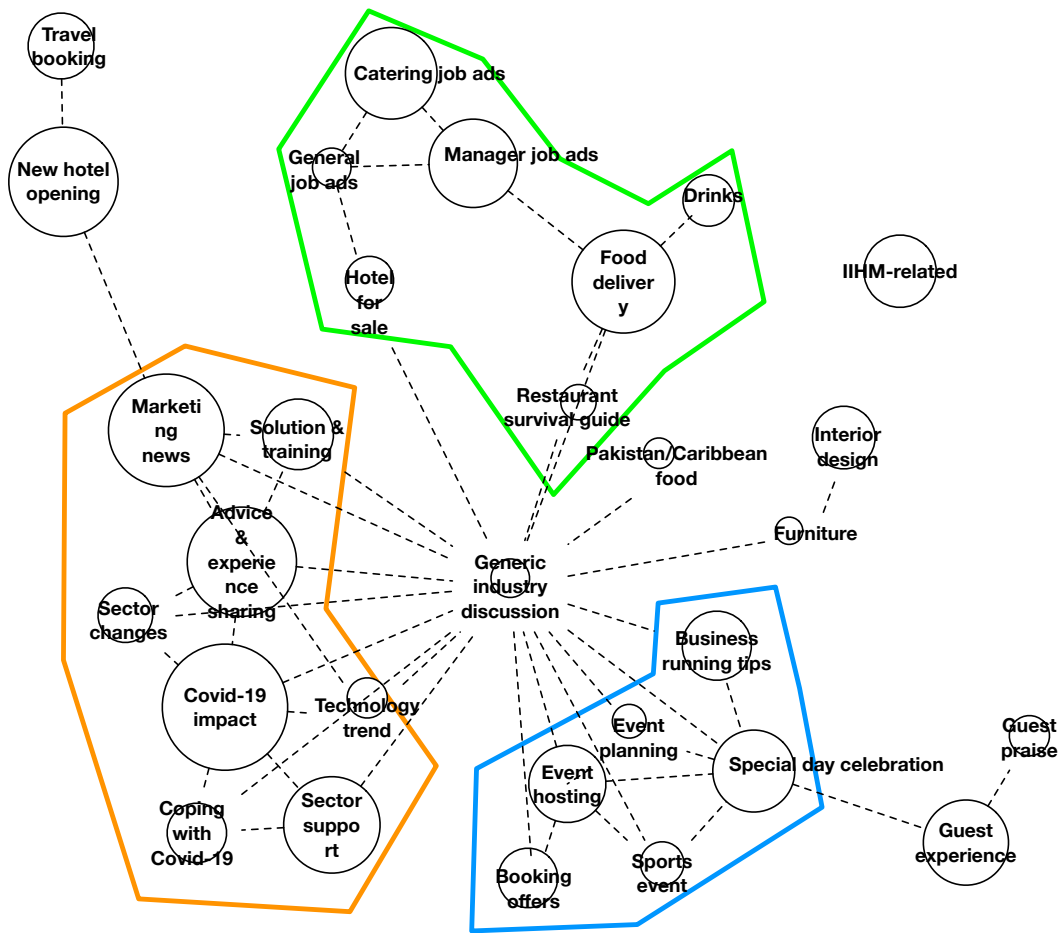


Figure 2: Network visualisation of topic connections based on the correlation between topics

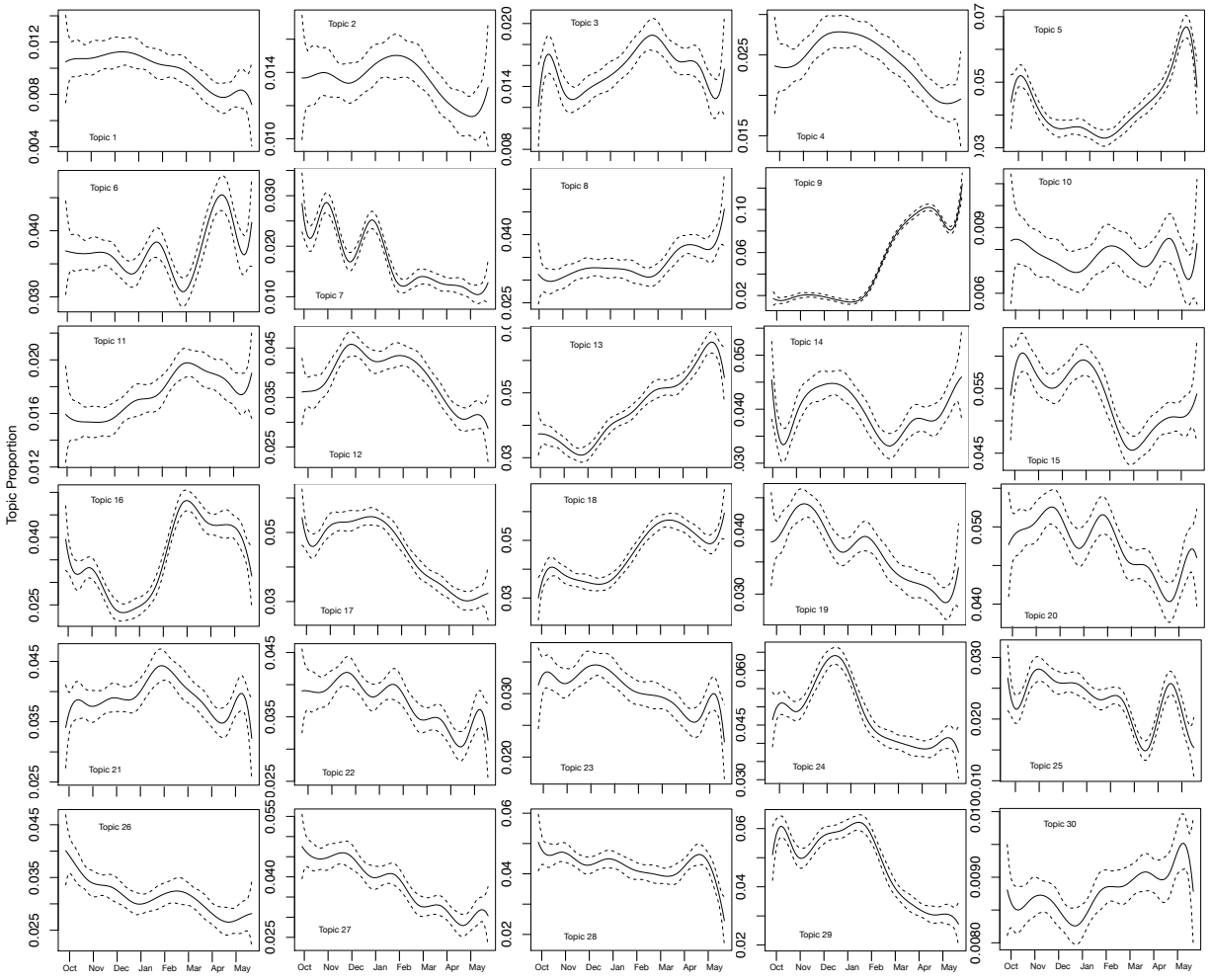


Figure 3: Topic prevalence over time