Information Systems Frontiers (2023) 25:1667–1682
https://doi.org/10.1007/s10796-022-10247-8

User Privacy, Surveillance and Public Health during COVID-19 – An Examination of Twitterverse

Paras Bhatt[1] · Naga Vemprala[2] · Rohit Valecha[1] · Govind Hariharan[3] · H. Raghav Rao[1]

Accepted: 4 January 2022 / Published online: 31 January 2022

Abstract

Online users frequently rely on social networking platforms to transmit public concerns and raise awareness about societal issues. With many government organizations actively employing social media data in recent times, the need for processing public concerns on social media has become a critical topic of interest across academic scholars and practitioners. However, the growing volume of social media data makes it difficult to process all the issues under a single umbrella, causing to overlook the main topic of interest within communication technologies, such as privacy. For example, during the COVID-19 pandemic, arguments on privacy and health issues exploded on Twitter, with several threads centered on contact tracking, health data gathering, and its usage by government agencies. To address the challenges of rising data volumes and to understand the importance of privacy concerns, particularly among users seeking greater privacy protection during this pandemic, we conduct a focused empirical analysis of user tweets about privacy. In this two-part research, our first study reveals three macro privacy issues of discussion distilled from the Twitter corpus, subsequently subdivided into 12 user privacy categories. The second study builds on the findings of the first study, focusing on the primary difficulties highlighted in the macro privacy subjects—contact tracing and digital surveillance. Using a document clustering approach, we present implications for the focal privacy topics that policymakers, agencies, and governments should consider for offering better privacy protections and help the community rebuild.

Keywords Coronavirus · COVID-19 · Privacy · Text mining · Hierarchical clustering · Social Media

1 Introduction

During a health crisis, people are often worried about their safety and healthcare facilities. In such circumstances, the number of people utilizing health services such as those covered by the Medicaid and Medicare programs increases ("Modern Healthcare," 2020, April 29). This increase has been associated with the generation of medical data including hospital records, personal data such as contact details and private data such as diagnosis and testing information. Unfortunately, there are numerous cases of negligence (Mercer, 2020, April 06) due to which data leakages happen thereby jeopardizing users' privacy during health crises.

Often during a health crisis, user privacy is not given much importance when planning response efforts (Palen et al., 2010). User data is generated at a rapid pace because of developments in telehealth/telemedicine, remote work, and unique supply chains. Online health counselling and therapy, virtual doctor visits and remote health monitoring require personal information from the users. This

✉ Paras Bhatt
paras.bhatt@utsa.edu

Naga Vemprala
naga.vemprala@gmail.com

Rohit Valecha
rohit.valecha@utsa.edu

Govind Hariharan
gharihar@kennesaw.edu

H. Raghav Rao
hr.rao@utsa.edu

[1] Department of Information Systems and Cyber Security, University of Texas at San Antonio, 1 UTSA Circle, San Antonio, TX 78249, USA

[2] Pamplin School Of Business, University of Portland, 5000 N Willamette Blvd, Portland, OR 97203, USA

[3] Department of Economics, Finance and Quantitative Analysis, Coles College of Business, Kennesaw State University, 560 Parliament Garden Way, MD 0403, Kennesaw, GA 30144, USA

information can be a potential target for adversaries to compromise user's privacy. There is a high probability of a lack of oversight regarding user privacy during emergencies. It comes as no surprise then that during health crises, incidents of breaches of user privacy have increased exponentially (The National Law Review, 2020, March 12).

As pointed out in the literature on the framework for formulating Bright Information Communication Technologies (ICT), during times such as the current pandemic there is an increase in public health concerns owing to an increased reliance on the internet and the need for privacy often gets overlooked at a time when it is even more critical to address not just cybersecurity but privacy as well (Lee et al., 2020). In accordance, researchers have called for developing solutions for global trust building so that users' privacy concerns are alleviated (Lee et al., 2018) with an emphasis on preventive cybersecurity and proactive privacy preservation (http://brightinternet.org/).

Social media is a rich source for investigating privacy discussions during health crises because it allows us to gather organic information pertaining to tracking and sharing of trending health topics as well as digital health data (Greenhalgh et al., 2020; Househ, 2012; Jain et al., 2016; Liu et al., 2021). The context of social media has seen increasing research (Kapoor et al., 2018) and is particularly well suited to examine conversations around privacy as it acts as a platform for individuals to share controlled information at their own will (Househ, 2012). Despite this, the extraction of privacy-related insights from social media in the aftermath of a health crisis have hitherto not been addressed in literature.

Along this backdrop, we ask the following research questions: 1) What aspects of health privacy do social media users discuss during the COVID-19 pandemic? and 2) what are the privacy discussions by the public around potentially the most invasive topics of contact tracing and surveillance? In addressing this question, this paper investigates privacy discussion around public health on a social media platform, Twitter, in the context of the ongoing COVID-19 pandemic. Using topic modeling approach, we conduct two studies to analyze social media messages on Twitter to ascertain what privacy topics related to their health information people are discussing during the COVID-19 pandemic. In study 1, we analyze user tweets and highlight the major categories of health privacy that social media users are tweeting about. In Study 2, we draw on the insights from Study 1 to create a taxonomy of the most salient privacy and cybersecurity discussions around the topics of contact tracing and digital surveillance that social media users on Twitter are engaged in. Our research adds to the growing body of knowledge about public interest in privacy concerns, contact tracing, and privacy violations in crisis situations like the current COVID-19 pandemic.

This investigation is important because the different aspects of privacy discussed on Twitter can help to guide policymaking by providing early insights into the concerns uppermost in people's minds. The insights from these discussions can help to strengthen frameworks to address and alleviate privacy concerns of people, with respect to technologies such as contact tracing and digital surveillance. The rest of the paper is organized as follows. In the next Sect. 2 we discuss the literature on user privacy in social media conversations. This is followed by the methodology Sect. 3 which describes the LDA topic modeling and hierarchical clustering approach to extract health privacy tweets of the users. In the following Sects. 4 and 5 we explain Study 1 and Study 2 respectively, and extract the various issues that users discuss on social media regarding health privacy. We also interpret the findings of our two studies, elaborate on privacy issues, and explain user privacy conversations that users discuss about the most on Twitter. The following discussion Sect. 6 explains the implications of the research, study contributions and future work. Finally, in the conclusion Sect. 7 we discuss the limitations of the study and outline our future work.

2 Literature Review

In this section, we discuss three related streams of literature on privacy issues from the context of healthcare data, contact tracing, and the role of social media in raising awareness during crisis events, which together lay the theoretical foundation of our study. The first stream focuses on privacy issues and addresses the concerns and benefits around the discussion of health information on social media. The second stream focuses on the core issues that are specific to health crisis such as COVID-19 where contact tracing is important to mitigate the risk of disease spread but may come at a cost of compromising the user privacy. The third and last stream focuses on the role of social media during crisis events.

2.1 Privacy Issues with Healthcare Data

The mode of accessing health information is changing as internet technologies, artificial intelligence and social media become more widely used (Trocin et al., 2021). On the internet, it is possible to discuss health information without alienating others or using an inaccessible language of healthcare (Cayton, 2006). With the spread of new infections and diseases, a free form of sharing health information is seen as having a positive impact on society. However, there are growing concerns about information privacy in relation to this type of free information sharing (Pershad, Hangge, Albadawi & Oklu, 2018). Recent studies on privacy have found that as internet adoption and technology integration into healthcare grows, there is a greater risk of user privacy being compromised (Greenhalgh et al., 2020). In the event of a healthcare crisis such as influenza, large-scale testing is required to halt the spread of the virus and identify any potential side effects from early-stage vaccinations (Signorini et al., 2011). During COVID-19, much of the industrialized world saw an increase in infections which forced governments to enforce lockdowns and individuals to remain indoors (Choudrie et al., 2021). To combat the pandemic, a variety of strategies have been implemented, including large-scale testing of patient samples. In such times, it is unavoidable to collect personal identifying information, which raises concerns about data loss. The number of reports of privacy breaches involving patient information has risen dramatically (HHS.gov, 2020, April 9).

2.2 Privacy Concerns, Contact Tracing, and Privacy Violations during Crisis

During the COVID-19 pandemic, people have relied on collaborative online video conferencing solutions to work and study remotely, schedule appointments with doctors using telehealth facilities. Scholars have also noted that a lack of understanding of privacy policies can influence whether or not people use telemedicine and virtual healthcare (Greenhalgh et al., 2020). Remote work has its fair share of privacy violations that can be counterproductive to the whole point of moving to online operations.

Another strategy for combating the virus is the increased use of contact tracing efforts by governments or other related agencies, which are tasked with monitoring the virus's spread. Contact tracing refers to efforts by public-health workers to track down potentially infected individuals (House & Keeling, 2010). Traditional contact tracing is used in situations of contagious outbreaks and relies on the knowledge of interpersonal network of physical interactions among people. However, due to privacy concerns of people and collection of noisy data, such tracing networks are difficult to reconstruct accurately (Farrahi et al., 2014). The authors note that communication traces obtained through mobile phones can be good proxies for physical interaction and may provide a valuable tool for contact tracing. Thus, during the crisis, several mobile applications were released by national governments to track the movement of its citizens—NHS COVID-19 (for the UK), Aarogya Setu (for India), COVIDSAFE (for Australia), Apple's Exposure notifications system. Recent contact tracing technology and methods are a matter of significant privacy concerns on the internet (Bhatt et al., 2020; Herold, 2006; Hiller & Russell, 2017). In summary, there are numerous concerns about privacy violations during crisis situations.

2.3 The Role of Social Media during Crisis

People search for information in times of crisis, especially in online environments. During this time, with a lack of consistent information from trusted agents there is a desire to gather as much information as possible, especially from social media. Social media, such as Twitter, has been effective in allowing public communication about the events of an emergency or crisis (Abedin & Babar, 2018). In health crises, it can provide early and valuable information about the situation and educate communities about preparedness measures, reduce the intensity of negative messages and promote positive messages (Rao et al., 2020). During health crises social media platforms can also play an important role in the lives of its users. Schillinger et al (2020) describes how they can serve as an outlet for disseminating information that promotes public health awareness. Zhou et al (2018) focus on how such platforms can help to improve the quality of care and communication. They can also serve as a cost effective and convenient venue for health intervention, patient health self-management, patient education, drug and healthcare service advertisements, etc. In addition, the social media context is particularly suited for examining conversations about privacy (Househ, 2012). Social media can serve as an outlet for people to express their views about the privacy of users during COVID-19. By enabling the exploration of social media content, we can extract intelligible information about people's privacy needs and concerns. While concerns about privacy on social media platforms are well recognized in the academic community, there is currently a lack of research in identifying user concerns in an automated manner (Saura et al., 2021). Organizations commonly use semi-structured interviews and qualitative research to identify user concerns specific to a domain or a specific event (Bandara et al., 2020). However, this is a time-consuming and laborious process, and in a global event like the COVID pandemic, identifying data privacy concerns in an automated manner is critical.

Despite this, the extraction of privacy-related insights in the aftermath of the health crisis has not received much attention in the literature. In accordance, this study examines privacy topics during the COVID-19 pandemic, outlining various issues discussed in the tweets. While some of these issues such as privacy policies, violations, and lack of transparency) have been previously studied by researchers (Janssen & van den Hoven, 2015; Sarathy & Robertson, 2003; Solove, 2008) from non-social media contexts, the resurgence of privacy concerns on Twitter during the pandemic underscores the importance of such discussions of health privacy specific to the context of social media networks for policymakers.

3 Methodology

Given the extensive use of Twitter during crisis communication management, researchers have adopted various techniques to extract meaningful information from tweets using text mining techniques like content analysis (Oh et al., 2011), social network analysis (Chatfield & Brajawidagda, 2012; Vemprala & Dietrich, 2019) and clustering (Aiello et al., 2013). Recently, sentiment analysis from subjective phrases has been combined with a machine learning algorithm to produce better disaster data classification accuracy (Ragini et al., 2018). We conduct two studies in this paper and follow a data driven exploration approach, similar to Bachura et al. (forthcoming 2022), to conduct exploratory research on data from Twitter conversation regarding users' health privacy discussions. A computational analysis of over 273 million tweets resulted in the extraction of 9 key health privacy topics in Study 1. Based on the public display of users' health privacy concerns shared on Twitter, we further distilled these topics into 3 main privacy issues in Study 2. Studying such public displays of privacy concerns are important for governments, organizations and policymakers. The social sharing of health related information concerns can provide clear guidance not only to organizations, but also to governments, allowing them to make informed decisions about what measures are working during a disaster and where they should focus their immediate attention. Researchers note that such social sharing is an important source of interpersonal interactions which people engage in to overcome from emotional experiences (Rodríguez-Hidalgo et al., 2017). Various topic classification methods are studied in combination with sentiment analysis to provide a generalized framework of how sentiment around a given topic changes over time in the study using disaster social media tweets during the Kerala floods (Mendon et al., 2021). For analyzing sentiments, the LDA and Hierarchical Clustering methods have been shown to work well. Information dissemination and the issue of credibility are another important stream of research related to crisis communication management, such as the COVID-19 pandemic. Understanding the underlying topics on a tweet-by-tweet basis is critical in addressing this issue. The authors were able to answer the question about user retweeting behavior by treating each tweet as if it represented only one topic discussion and by looking at the tweet characteristics (Son et al., 2020). Building on the prior literature related to the crisis response and topic classification, in this paper, we utilize hierarchical clustering, a text-mining approach, to obtain semantic patterns from tweets over a specific time period so that we can effectively model the various user privacy issues in the COVID-19 context.

Twitter data can be effectively used to detect subtle variations in public response (Tan et al., 2013) which are ever more visible during the COVID-19 pandemic. This study presents an analysis of tweets produced beginning from January 20, the day China officially confirmed the infection outside Hubei province until July 2020. We used Twitter streaming API to collect the tweets based on the keyword coronavirus and added COVID-19 to our list when the WHO officially named the disease.

3.1 Data Description

We developed a Python script that calls the streaming API and searches for keywords anywhere in the tweet's text and extracts the tweet and the tweet characteristics, including retweets count, the tweeting user's screen name, the tweet's time, and hashtags. By the end of July, we collected around 273 million tweets for 194 days. Figure 1 shows the records

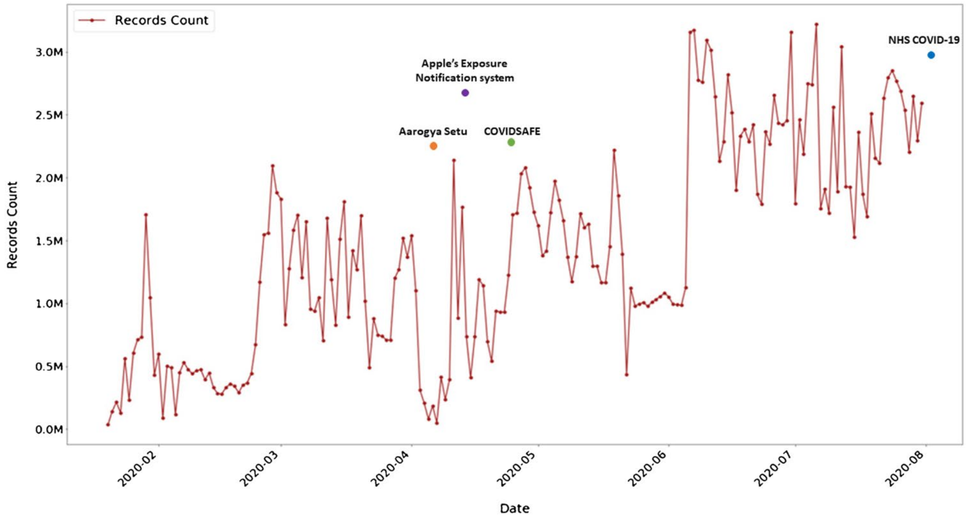

Fig. 1 Timeline of coronavirus tweets

count by days. There are considerable number of days where the daily tweet count crossed 2 million tweets a day. We utilized 6 months of tweets from January 20, 2020 to July 31, 2020 owing to a continuous stream of data during this period. The distribution of tweets across this period can be seen in Fig. 1. The 273 million tweets are representative of user discussions on the social media platform about COVID-19. Before we analyze the health privacy topics within users' conversations in these tweets, we have to extract the relevant tweets. This was done in two steps: First, we used a keyword-based search to extract privacy-related tweets by including keywords 'privacy' and 'confidentiality' in our search criteria. Second, we matched keywords specific to health using an LDA topic model created from the relevant web pages providing health updates on COVID-19 (Blei, 2012). To create topic models, we used the Python Gensim package's LDA model. Gensim includes several text processing methods, such as removing stop words (commonly used words in sentences that add no contextual meaning to the topics) and lemmatization (to return root words instead of inflectional ending of the words, e.g., mouse instead of mice and foot instead of foot). Additionally, we followed a heuristics approach to pre-process the text data to filter Twitter specific characters and filtered URL links, user mentions before passing it to the Gensim package LDA method. Gensim functions are optimized to converge in the fewest number of iterations and provide the topics as quickly as possible.

3.2 Extracting Health Privacy-Related Tweets

Topic model provides keywords that are closely related to each of the topic. We systematically analyzed the words under each topic, to represent unique keywords, and classified the tweets as health-related tweets if any of the respective keywords were present in the tweet text. Figure 2 shows the distribution of health privacy-related tweets. It shows the magnitude of discussions that relate to health privacy across our dataset with significant spikes in the months of March and June as the infection spread accelerated during these months.

We extract insights from the health privacy tweets in a sequence of two studies. In Study 1, we determine the major topics on the different aspects of health privacy being discussed by Twitter users. In Study 2, we delve further into the topics which were most salient in Study 1. For this deeper investigation into the main topics of cybersecurity and privacy, we also include additional tweets in the second study as the data for two additional months became available up to September 2020. Figure 3 presents the process for extraction and analysis for Study 1 and Study 2. These are further elaborated subsequently in the corresponding sections.

4 Study 1: Health Privacy Topic Discussion in Twitterverse

In this study, we begin our detailed analysis by identifying the critical health privacy topics that are at the forefront of discussions and concerns voiced in our data using the process characterized in the upper half of Fig. 3. Owing to the inherent nature of conversations on social media that continue for several days, we decided to use a topic modeling approach reinforced with hierarchical clustering. This enabled us to get new topics each week based on their popularity which was determined by the coherence score. We also utilized reverse boosting to ensure that popular topics in initial weeks did not affect the algorithm in the coming weeks by assigning additional feature weights to recent discussions. The model used the timestamps from tweets to infer the temporal progression of emerging topic discussions in the Twitterverse. The topic model generated the most popular weekly topics for 14 weeks. There were 549 unique topics within this period which directly mapped to

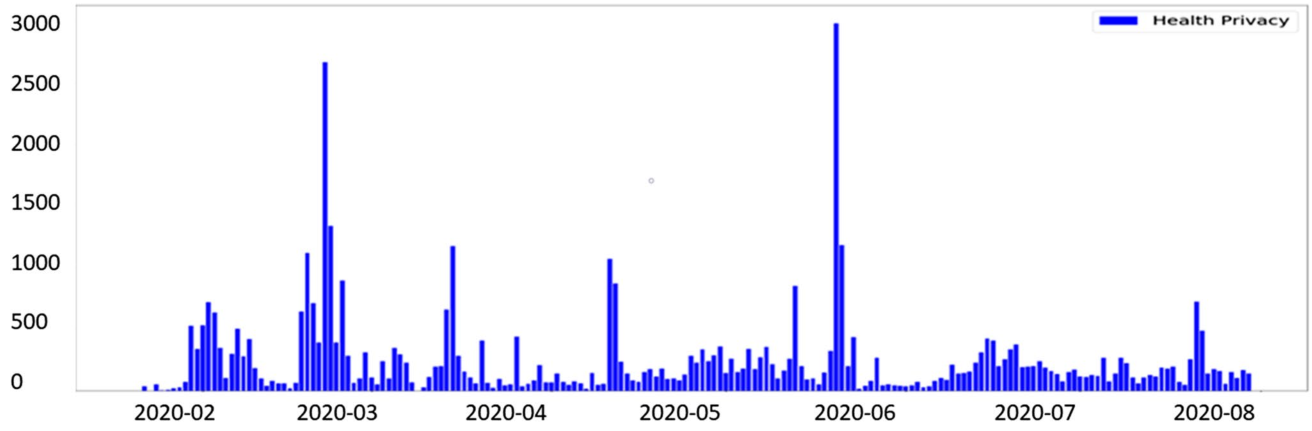

Fig. 2 Distribution of health privacy tweets

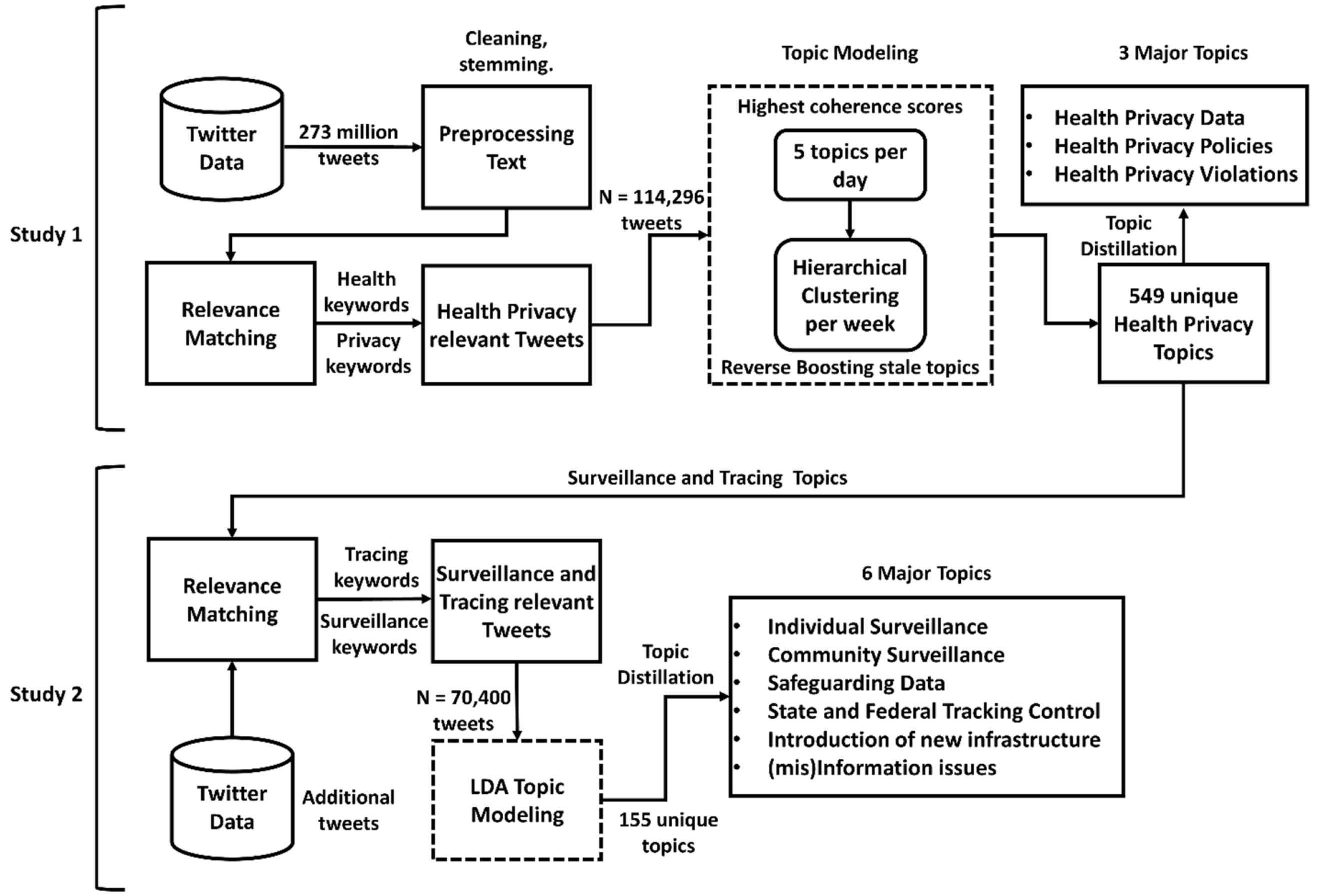

Fig. 3 Study 1 and 2 process model

114,296 individual tweets. We have also used the structural topic model (STM) to present the high-level topics within the overall privacy discussions (Roberts et al., 2019). From our analysis, we identified 9 distinct overarching health privacy topics. Consistent with the number of topics, we ran the STM to capture the 9 distinct topics in the overall tweets (Table 1).

To get the major topics of health privacy, we followed a topic distillation approach (Chakrabarti, 2001; Chakrabarti et al., 2001) based on the topic text and the frequency of topic occurrences. To this end we compared the text content of the topic with our set of keywords. For example, if a topic contained the words (bi-grams) ‘patient information’ and another topic contained ‘patient reports’, it was distilled under the bigger ‘patient data confidentiality’ topic. Using this topic distillation approach, the individual 549 health privacy topics were read by the first and second authors, and classified based on their content. After the initial round of distillation, there were 46 interrelated topics which were further distilled to 16 topics with significant differences within their content and minimal overlap. These topics were merged into 9 distinct overarching health privacy topics shown in Table 2 that provides a description of the distilled set of 9 privacy topics as well as a sample of tweets.

These 9 privacy topics were related to 3 major topics relating to digital privacy, lack of transparent privacy policies and privacy violations. These topics are grounded in prior research on privacy with articles addressing issues of digital privacy (Sarathy & Robertson, 2003), privacy violations (Solove, 2008), and lack of transparent privacy policies with respect to governments’ use of data (Janssen & van den Hoven, 2015).

To understand the underlying topics, we ran the hierarchical clustering algorithm. The df-idf (document frequency-inverse document frequency) metric was used to detect the discussion's central topic over a given period (Aiello et al., 2013). We used bi-grams (a combination of two words) to detect events by comparing bi-gram frequencies daily with those of the preceding days. The df-idf

Table 1 Topic keywords – generated using STM

Privacy topics	Keywords
Topic 1	privacy, need, health, time, data, safety, laws, company, good, policy, individual, online, big, tech, read, personal, crisis, digital, great, amp
Topic 2	contact, app, concern, track, tracing, use, mobile, personal, apple, government, google, track, data, spread, new, privacy, use, surveillance, use, communicate
Topic 3	important, protect, testing, source, work, legal, something, problem, think, home, lack, gov, answer, address, health, welfare, global, question, time, care
Topic 4	secrecy, spread, virus, ccp, trump, security, surveillance, concern, tracing, privacy, paranoia, epidemic, world, change, state, welfare, communist, health, journalist, contact
Topic 5	crisis, consent, report, education, video, settings, concerns, surveillance, system, information, inaccurate, error, warn, erode, require, see, develop, data, class, track
Topic 6	human, expose, test, respect, support, quarantine, user, available, contact, public, customer, report, case, confidential, help, privacy, positive, technology, individual, update
Topic 7	news, invasion, risk, report, keep, washington, political, election, apple, bios, care, digital, elevated, logs, response, suppression, need, around, cover, demand
Topic 8	report, use, one, city, cite, turnaround, leader, scare, legislative, simply, education, care, giving, fear, lead, exclusive, information, cover, welfare, public
Topic 9	medical, patient, attention, negligence, confidential, care, privacy, health, case, information, facility, died, social, part, want, human, life, right, company, using

approach clusters bi-grams rather than the regular document cluster approach that uses tf-idf technique (term frequency—inverse document frequency) to cluster documents within a given time interval. A regular clustering algorithm based on the tf-idf technique considers all these messages as a single large cluster and may ignore the emerging topics from new incoming messages (because the old, repetitive topics still carry a large weight), making it difficult to identify the development of new topics. As the primary focus of the df-idf approach is word co-occurrence, df-idf provides greater accuracy in extracting the central discussion tweets for each time interval. Bi-grams naturally cluster the co-occurring words together and provide an efficient cluster of sentences than the unigrams algorithm (Aiello et al., 2013). Once, the central tweets were extracted, we performed the content analysis to group the extracted focal tweets into different categories.

Through this approach, the 9 health privacy topics were related to user discussions about data privacy, privacy policies and privacy violations. (1) Data privacy has been a contentious issue ever since the emergence of the Big data era (Jain et al., 2016). The highly unstructured form of data as well as the continuous stream of data points makes it all the more difficult to isolate privacy incidents. (2) Privacy policies can safeguard people's privacy. It has been observed that in times of crises privacy policies are often not clear and sometimes there may even be a tradeoff that results in further privacy loss (Büscher et al., 2019). (3) As for privacy violations, crisis or not, they are commonplace as well as difficult to track. Palen and Dourish (2003) had observed that privacy is visceral in a networked world and violations do occur frequently, which rings true even in this contemporary time. We have visualized the trends of the central discussion of the topics over the timeline in Fig. 4.

Data Privacy Much of the work has shifted online during the pandemic including increase in Telemedicine (Smith et al., 2020). In addition, on account of the increased infection spread as well as testing of those infected, it has resulted in the generation of large health data. Therefore, conversations on Twitter rightly mimic this pattern as evidenced by the significant peaks seen in the health data graph. The peaks correspond to the general perception of health-related discussions during the month of May rising significantly with the total number of COVID-19 related deaths exceeding 100,000.

Privacy Policies The major topic of discussion revolves around the lack of transparency. People online are sharing their views on privacy and its implications for data regulation policies, guidelines, and rules. We see from our analysis that users regularly tweet about the absence of strong privacy policies that can govern data collection and storage. We see from the policy graphs a consistent discussion regarding privacy policies during the earlier days of transmission of the virus. As the information percolated to the masses, we see these discussions trailing off towards the end of our dataset as more robust health policies regarding data collection were put in place by state and federal authorities.

Privacy Violations We see from the graph that privacy violations are the most discussed topics during our dataset time period. They begin slowly, because of a tepid response to formulating contact tracing guidelines, but once the cases started increasing and as a result of an influx of contact

Table 2 Central discussions within the clusters of tweets over timeline

Privacy topics	Description
Digital Privacy Keywords: privacy settings; technology	Discussions about the potential of technology being misused to track and collect personal data Tweet Example: This government has no respect for privacy or citizen data. Bengaluru covid patient details with name, mobile number, full address sorted by ward wise, up online. […]
Patient Data Confidentiality Keywords: patient information; reports	Debate around issues relating to patients' personal information (diagnosis) being disclosed and the possible impact of such disclosure Tweet Examples: The …. government has quietly relaxed a confidentiality law that protects patient health data
Privacy in Education Keywords: K12; college	Privacy discussions regarding kids going back school and the privacy protections put in place by schools and colleges to prevent data leakages Tweet Examples: "The U.S. Department of Ed announces they may waive school testing requirements. #COVID—19 @ EricaLG
Lack of Transparency Keywords: underreporting; inaccuracies	Issues relating to COVID-19 response efforts taken by official agencies, state and local governments and the level of privacy protection observed during such efforts Tweet Example: #BREAKING COVID-19 cover-up continues: Confidential memos show Florida officials knew pandemic was likely, but didn't warn the public and refused to release data, even to local hospitals and state senators […]
Policy Misuse Keyword: privacy policies	Discussions around the privacy policies that will govern and be applicable to data collected during the COVID-19 pandemic Tweet Example: Privacy in the Time of Pandemic: COVID-19 Provides Opportunity to Revisit Regulation S-P Privacy Policies […]
Medical Negligence Keywords: healthcare, hospital administration	Possibilities of medical negligence affecting privacy of users and discussions about hospital staff potentially misusing patient data Tweet Example: Similar pattern of secrecy and negligence has been repeated in Iran, facilitating further spread of the Virus to Afghanistan and Kuwait […]
Privacy Violations Keyword: breach; erosion; rights	Right to privacy and the ability to keep personal information secure and protected from snooping Tweet Example: There is a risk that aspects of the battle against COVID-19 become a new "forever war," with privacy violations and censorship becoming semipermanent on the grounds that the virus, and its inevitable successors, can never be fully vanquished. […]
Tracking Issues Keywords: contact; trace; locate	Contact tracing and the ability of agencies, apps to digitally track and surveil users without their knowledge Tweet Example: RT @ZDNet: Coronavirus: They want to use your location data to fight pandemic. That's a big privacy issue […]
Ethical Issues Keywords: laws; ethics; consent	Issues relating to official agencies flouting current privacy laws while collecting data and not getting informed consent from users Tweet Example: "No thanks for these #SARSCoV2 Immunity Passports & Certification. I'll take the immunity, skip the passport. Until we get the serology tests, privacy, ethics issues on track. […]"

tracing mobile applications (Ahmed et al., 2020), more tracking efforts were made. We see consistent spikes in these issues being discussed on Twitter. Another important issue was how medical negligence has increased and users shared their views regarding health data violations, hospital administrations and leakage of health records.

5 Study 2: Discussions about Contact Tracing and Digital Surveillance Topics in Twittersphere

Within the 9 health privacy topics, tracking issues had the most frequency of occurrences and a majority of the discussion on Twitter revolved around these two topics – contact tracing and surveillance. Of the 549 topic models we had

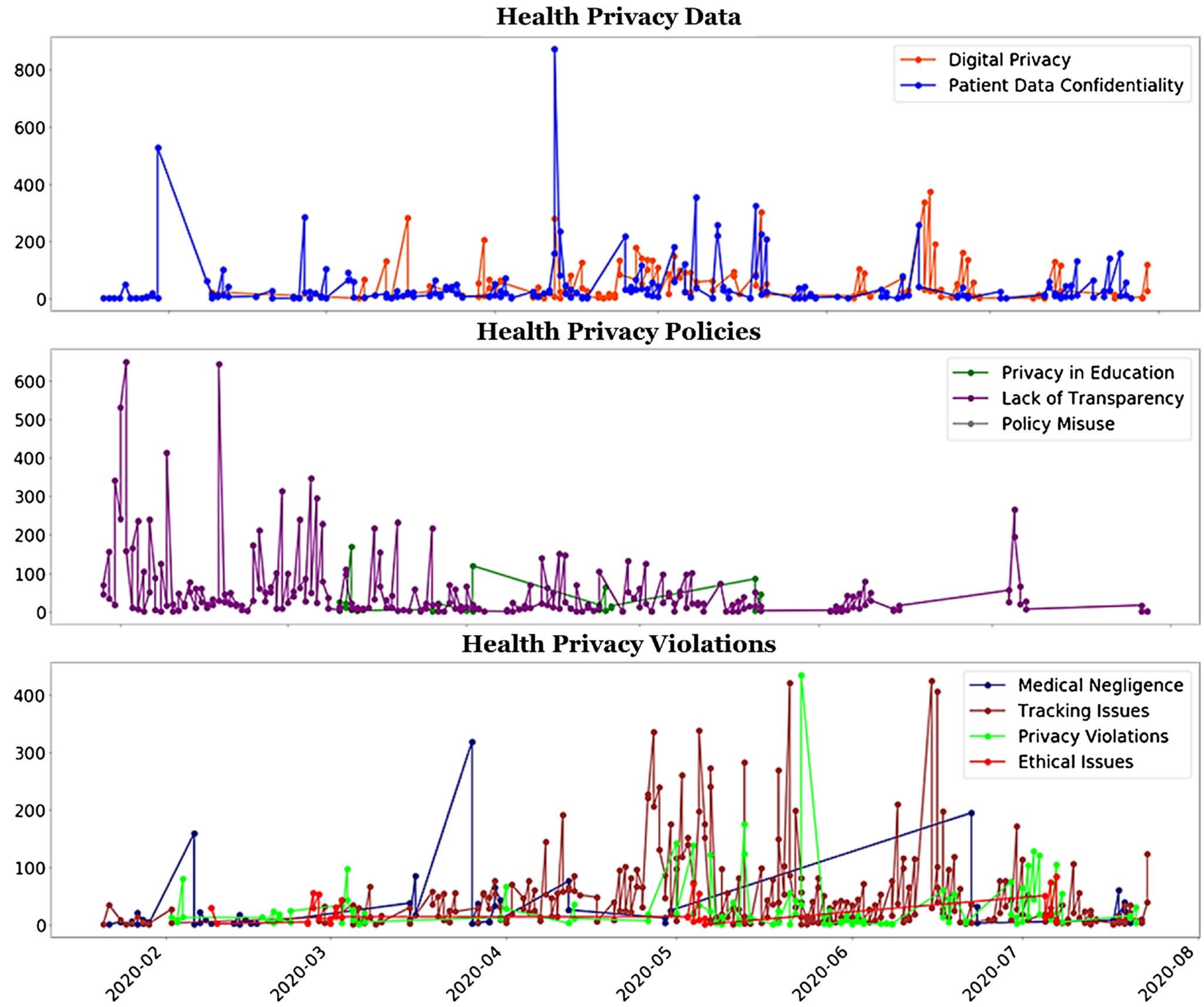

Fig. 4 Diffusion of privacy tweets by topics within health categories

initially obtained from our topic modeling approach, over a 100 applied to these two topics alone. Since COVID-19 first started spreading rapidly in March, these two topics have endured through our entire datasets with consistent discussions being generated across different weeks. It has ushered in new surveillance technologies that seek to parse people's digital footprints (Calvo et al., 2020). The level of information (including telephones, email addresses, etc.) collected in response to contact tracing efforts has been unprecedented (Madewell et al., 2020).

Since the two major health privacy topics and areas of discussions revolved around contact tracing and digital surveillance (which were consistently present topics of conversations across our time series data), we decided to conduct a follow up study using a second topic model that utilized n-gram word embeddings. These n-gram relates to using either one-word (unigrams) searches over the dataset or a combination of two subsequently occurring words (bigrams) or three subsequently occurring words (trigrams) to generate the topic of interest for each of the weeks in the dataset. Using this approach, we were able to generate topics from our list of keywords, which included a mix of n-gram words, that capture more relevant and coherent topics. The model focused exclusively on surveillance and contact tracing discussions with respect to health privacy concerns of people on social media on the initial dataset plus an additional two months of data up until September 30, 2020.

After running the model on this narrower search criteria, the number of topics found sharply decreased and showed more direct word associations with the tweet content. In total we extracted 155 unique topics that related to either surveillance or tracing or both which roughly mapped to around

70,400 individual tweets. These topics captured discussions around surveillance and contact tracing. For instance, users discussed the need for greater privacy protections in light of the introduction of new technology or systems that can track people and their contact with close acquaintances.

We again proceeded with the topic distillation approach of content analysis and frequency occurrences on the newer reduced set of 155 topics. The various topics discussed were highly correlated and after the first round of analysis the topics were reduced to 75 interrelated topics of discussion. For example, these interrelated topics included some that discussed 'increase in surveillance technology' and others that discussed a similar topic 'increase in surveillance programs/networks'. The grouping of similar topics helped to reduce the number of topics under consideration and laid the foundation to creating a smaller non overlapping set of privacy issues, that could be explained further. In this regard, the 75 topics were then distilled into 6 non-overlapping major privacy topics related to surveillance and contact tracing. Table 3 represents the 6 major privacy topics and key issues along with their representative tweets.

These topics relate to the major discussions on Twitter during the period of Study 2. In these discussions, new issues emerged that corresponded to users' discussions

Table 3 Central discussions within contact tracing and surveillance domain

Privacy topics	Description
Individual Surveillance Key issues: Phone/ Location/Map Tracking	Privacy discussions regarding individual privacy being compromised, personal freedom being curtailed, and civil liberties being relaxed Tweet Example: Your iPhone has COVID-19 tracking setting if you want to turn it on. I turn off location services on my phone. I know that they are capable of tracking you, but I don't want them to have my data. The data for a family of 5 is worth 20 k a year, and they get it for free. Sep 26, 2020
Community Surveillance Key issues: Government control over Citizens/People/Community	Discussions relating to the existence of mass surveillance by government in the community or by employers at the workplace Tweet Example: "NHS" COVID-19 app being trialed on the Isle of Wight doesn't just ask for the first part of your postcode… it's tracking your PRECISE location All the time Is this data you want to share with this Government? All the time? (NB. And it prevents your phone from sleeping). May 5, 2020
Safeguarding Data Key issues: Data Collection Procedures/Guidelines/Efficacy	Privacy discussions regarding the need for safeguarding data collected in the name of contact tracing and surveillance protocols Tweet Example: @……. great job at keeping kids safe from COVID-19 but giving out their personal information on live TV. Primary school location, name, child's name, age, parents name is very concerning and certainly not abiding by data protection and safeguarding laws. Jun 1, 2020
State and Federal Tracking Control Key issues: Government (local, state, federal) surveillance	Discussions about various local, state, and federal governments that seek to implement or control contact tracing and surveillance programs Tweet Example: Controversial tech firm pitches facial recognition for govt tracking of COVID-19: Clearview AI wooing state governments to use FR to "gather data on the spread of COVID-19. Privacy experts have sounded the alarm. Apr 30, 2020
Introduction of new Infrastructure Key issues: Increase in Surveillance Technology/Internet Tracking/Tracing Apps	Privacy discussions regarding the flurry of new technology and infrastructure being developed for contact tracing and surveilling people Tweet Example: This supposed Pandemic has become a cover for more surveillance technology to be introduced. COVID-19 could set a new norm for surveillance and privacy. May 11, 2020
(mis)Information issues Key issues: Spreading false information about tracing data being misused	Possibilities of misinformation about health privacy being circulated with respect to contact tracing and surveillance programs Tweet Example: Don't create, share misinformation or fake news. Don't breach privacy of patients by revealing their names: #Telangana Govt. Mar 30, 2020

regarding newer technology and infrastructure being developed during the pandemic that can have serious impact on privacy both in current times as well as in the future.

The Internet users' information privacy concerns (IUIPC) framework proposed by Malhotra et al. (2004) can help understand online consumers' concerns for information privacy. The first component that impacts users in any online setting is the collection of data. Data has been famously referred to as the price you pay for free services on the internet. Data collection is thus the first issue that might affect people's privacy decisions and their intention to engage or share their personal data. The second component is control which manifest into the perceived level of control that users feel they have over their data. If users believe that they have control over what is shared in the online space regarding their personal information, it manifests into lower privacy concerns. The final component is awareness of privacy practices. It translates to the level of disclosure and knowledge users are provided about the usage of their data. It is related to being transparent upfront about the intended use of people's data, and the purpose of data collection should be clearly communicated to users in order to inspire confidence in them to share personal details.

Through the two studies, we have ascertained the different privacy issues that users discuss online during the COVID-19 pandemic. We observe that our findings from Study 2 closely align with the existing privacy mechanisms of privacy issues discussed by Malhotra et al. (2004). Using our findings from Study 2 we explain how the major six topics revealed from our topic modeling approach relate to the privacy components.

Data Collection Two major privacy topics from our topic model that correspond to it are individual and community surveillance. The key issues within each of the surveillance topics were representative of the context that data collection for the purposes of contact tracing has. Within individual surveillance topic, the privacy issues raised by users included the access to phone data, maps data and location data. Apart from this, the encroachment of personal freedom and the dwindling civil liberties afforded to citizens were other major issues being discussed. Under community surveillance topic, the Twitter users discussed issues relating to the widespread tracking of travel and itinerary details, the use of mass surveillance programs to collect data and community tracking of affected populations. As noted by Lyon (2010), individual, institutional and community surveillance are all closely linked to personal data collection from users. It relates to users' choice of providing their personal data, which in some scenarios may not exist such as in the case of mandatory contact tracing. The existence of these two major topics shows how important data collection is to contact tracing. In order to inspire confidence among people to voluntarily participate in contact tracing programs, it is essential to clearly communicate to them the data collection objectives at the very outset.

Data Control Two major topics from our topic model related to control are safeguarding of data and state and federal tracking control. Within the safeguarding data topic, the privacy discussions mirrored data collection issues such that a majority of the users tweeted about data collection procedures and guidelines. The discussions also echoed a lack of data protections and security during the period of study. In times of crises, it is seen that data collection procedures as well as governments themselves are not fully transparent (De Araujo & Tejedo-Romero, 2016). A majority of the topics discussed online also pointed to the control of data as a major topic of interest, especially the level of control by local, state and federal institutions. The users prominently discussed the use of their personal data well after the pandemic ends. The emphasis of users on control sheds light on its importance for future surveillance and tracing programs. Since sharing personal information is always attached with a risk, people naturally like to have more control over what is done with their personal information (Prince, 2018). Therefore, it would be beneficial for policy makers to provide users with sufficient control over their data so that they can voluntarily and proactively participate in such programs in the future.

Awareness of Privacy Practices The primary issue is that people do not want to share information online unless they are aware of the intended use of that information (Hoffman & Novak, 1996). We see users' privacy issues on Twitter, extracted from our topic model, that discuss topics like the introduction of newer surveillance technology and systems. Such new applications are released and advertised as using contact tracing to limit the spread of the virus. However, in the case of such healthcare apps, it is often seen that the data collection process, the actual personal information requested and the policies governing data usage are rarely stated (Rosenfeld et al., 2017). In this manner the majority of contact tracing apps that are released are doomed from the start and suffer from low adoption rates (Kim et al., 2007). Therefore, for policy makers it is very important to advertise the intended use of information and create awareness about the privacy protections for adopters of contact tracing programs. Figure 5 shows a combination of our results from Study 2, with the 6 major privacy topics, and the IUIPC framework to generate a word cloud based on interrelated topics.

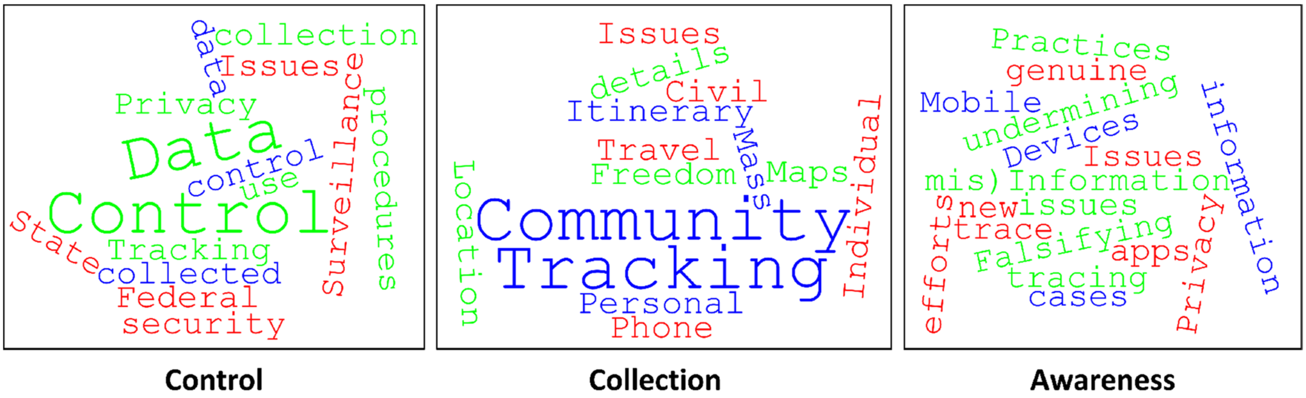

Fig. 5 WordCloud of 3 major privacy dimensions

6 Research Findings

In this paper, we have extracted privacy insights from social media discussions in the aftermath of a health crisis, COVID-19 pandemic. The findings from this paper conform with the Internet users' information privacy concerns (IUIPC) framework proposed by Malhotra et al. (2004). We segregated the major health privacy topics discussed by Twitter users into three issues. First, data privacy correlates with the first component of IUIPC, which is collection. Since the majority of discussions revolved around health data, its collection is naturally an issue that people discuss online. Second, privacy policies closely relate to control. As it has been shown by numerous studies, users like to understand the policies that control their data and govern its usage (Earp et al., 2005; Sadeh et al., 2009). Privacy policies exert effective control on data aggregators or collectors and prevent user data from being misused. Third, privacy violations align with the awareness of privacy practices' component.

6.1 Theoretical Implications

In terms of theoretical implications, we have outlined the need to consider user privacy and its sustained importance whence forming guidelines, designing protective measures, and establishing a response mechanism to counter the adverse privacy implications of pandemics. In this work, we explore the social sharing of health related information concerns through an analysis of over 270 million tweets. Using data driven exploration we uncover 9 different health privacy topics in Study 1 further distil them into 3 central privacy topics. Our data exploration approach contributes to enhancing privacy literacy and bridging theory and practice (Wissinger, 2017) by incorporating discussions on digital privacy shared on social media platforms during crises events like the COVID-19 pandemic. During crisis, privacy might not be a central focus for the authorities, but from our analysis of Twitter conversations, it most certainly is an emotional issue for the community. Prior research has noted the importance of social sharing of emotions on Twitter and other similar platforms as such sharing serves as a form of interpersonal interaction which helps people to overcome their emotions (Rodríguez-Hidalgo et al., 2017; Bachura et al., forthcoming 2022). During times of crisis, privacy might not be a central focus for the authorities, but from our analysis of Twitter users, it most certainly is an emotional issue for the community. The right to privacy is even enshrined in the constitution, and users do not want to compromise it for a sure pandemic or not. Therefore, privacy during times of crises such as pandemics is an essential avenue for future research in the domain of user privacy.

6.2 Practical Implications

In terms of practical implications, official agencies can utilize the insights to focus on alleviating the privacy issues of individuals in the society and design better information systems and enforce stronger privacy policies. For ensuring the success of any technologies, the participation of the people is very important. It is then sensible to install measures that satisfy the basic tenets of confidentiality set forth in the right to privacy so as to ensure users' privacy is protected, and people are assured of the sanctity of their information. This research can also help to drive official campaigns that target the need for community participation in handling crisis situations.

6.3 Policymaking Implications

The privacy issues discussed on Twitter have implications for governments and policy makers, and addressing them sufficiently can help revitalize tracing programs. With the increase in vaccination efforts across the developed world and the discovery of new variants of the virus, contact tracing and surveillance have become a contentious issue (John Hopkins School of Public Health, 2021, Feb 2). Also, the debate around the use of vaccination passports to enable safe travel across state and national boundaries has added to already existing privacy concerns of users (The New York Times, 2021, Feb 4). In this light, monitoring social media discussions are all the more important for deriving insights for policymaking to ensure success of such mechanisms. In addition, these privacy issues show the need for better of data protection policies and warrant the need to regulate the virtual space including online learning or remote work. There needs to be a stronger emphasis on regulating data collected during this crisis. Also, with respect to privacy violation issues it is imperative that the myths around contact tracing, surveillance and tracking are addressed. This would enable greater participation of the public in such tracing efforts provided their privacy issues are alleviated.

7 Conclusion

Using an LDA topic modeling approach we analyzed social media messages on Twitter and determined the most salient privacy topics that people are discussing during the COVID-19 pandemic. The topics of contact tracing and digital surveillance, which are related to user privacy and cybersecurity were also discussed. We also present the contribution from our work to the Bright ICT initiative and its importance in policymaking for ensuring user privacy preservation and cybersecurity. The vision of the Bright Internet can be used by governments and policy makers to develop solutions for global trust building so that users' privacy concerns are alleviated (Lee et al., 2018). Contact tracing and surveillance applications developed to reign in the pandemic can also be fortified with these Bright ICT initiatives so that there is an emphasis on preventive cybersecurity and proactive privacy preservation (http://brightinternet.org/).

During pandemic times, it is even more critical to ensure privacy is protected along with cybersecurity due to the increased reliance on the internet and virtual interactions. This paper utilizes Twitter data in a crisis in order to provide guidance on the cybersecurity and privacy concerns that are foremost in people's mind during a public health crisis. We plan to extend this work into a drawn-out investigation of the privacy needs of social media users. In our future efforts, we will also study the impact of privacy breaches on the population when their confidential data is used to drive a global response but without explicit user consent on data sharing. For other researchers in this domain, we present them with a few challenges regarding how to ensure privacy is maintained during times of a global emergency and how to sustain this privacy even when everything goes back to normal. With many news outlets, healthcare institutions and individual users on the Twitter, subtle variations in public sentiment can be observed on the platform (Tan et al., 2013). We aim to augment our data source in our future work and expand it to include other social media platforms as well.

Our research, like any other research into information management systems, has its limitations. First and foremost, our approach focuses on privacy concerns related to health wherein the severity of data privacy concerns, as well as their prevalence, varies from country to country. In this study, we did not separate tweets originating from a specific country or target the privacy concerns of a specific country when collecting the data. Nevertheless, governments are addressing the issue of data privacy and working toward regulating social media platforms so that individuals' privacy is not at stake. A tweet message's source can be determined by multiple approaches, including hashtags, geolocation tags, and the presence of location in the tweet text. Identifying the location allows public bodies and governments to strengthen data privacy laws on social media platforms by tagging tweets with the origin country of the tweet. We plan to extend our research at a later point in time, considering the location-specific attributes in combination with health-privacy tweets. Second, our data collection does not distinguish between authentic and fraudulent tweets. During a pandemic, many social media users around the world try to influence other users by spreading false or misleading information about a specific event or situation, as well as creating unnecessarily tense situations in the general public. In such a situation, the ability to comprehend the widely discussed topic in conjunction with the authenticity of information allows one to mitigate the risks associated with misinformation. We encourage researchers conducting misinformation-related studies to consider privacy concerns alongside misinformation tweets when examining topic comparisons by misinformation. Also, future researchers can consider Communication Privacy Theory to study the social sharing of information, especially in the context of health privacy which creates tension within individuals and increases or decreases their risk of privacy leakages (Petronio & Venetis, 2017). On the one hand, social sharing of private health information can help to garner support to cope with health problems, On the other, people's privacy concerns may prevent them from seeking such much needed support. Finally, the focus of our research is textual information shared on the internet. This study does not address the issue of sharing personal patient records as an image file or in any other format other than text format in order to avoid violating patient privacy.

Declarations

Conflicts of Interests All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

References

Abedin, B., & Babar, A. (2018). Institutional vs. non-institutional use of social media during emergency response: A case of twitter in 2014 Australian bush fire. *Information Systems Frontiers*, 20(4), 729–740.

Ahmed, N., Michelin, R. A., Xue, W., Ruj, S., Malaney, R., Kanhere, S. S., . . . Jha, S. K. (2020). A survey of covid-19 contact tracing apps. *IEEE Access, 8*, 134577-134601.

Aiello, L. M., Petkos, G., Martin, C., Corney, D., Papadopoulos, S., Skraba, R., . . . Jaimes, A. (2013). Sensing trending topics in Twitter. *IEEE Transactions on Multimedia, 15*(6), 1268-1282.

Choudrie, J., Patil, S., Kotecha, K., Matta, N., & Pappas, I. (2021). Applying and understanding an advanced, novel deep learning approach: A Covid 19, text based, emotions analysis study. *Information Systems Frontiers*, 1–35.

Bachura, E., Valecha, R., Chen, R., & Rao, H.R. (2022). The OPM data breach: An investigation of shared emotional reactions on Twitter. *MIS Quarterly, forthcoming*.

Bandara, R., Fernando, M., & Akter, S. (2020). Addressing privacy predicaments in the digital marketplace: A power-relations perspective. *International Journal of Consumer Studies, 44*(5), 423–434.

Bhatt, P., Vemprala, N., Valecha, R., & Rao, H. R. (2020) *Life vs. livelihood–user privacy during COVID-19 virus communication*. IFIP 8.11/11.13 Dewald Roode Information Security Research Workshop.

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM, 55*(4), 77–84.

Büscher, M., Perng, S.-Y., & Liegl, M. (2019). Privacy, security, and liberty: Ict in crises *Censorship, Surveillance, and Privacy: Concepts, Methodologies, Tools, and Applications* (pp. 199–217): IGI Global.

Calvo, R. A., Deterding, S., & Ryan, R. M. (2020). Health surveillance during covid-19 pandemic. *British Medical Journal, 369*. https://doi.org/10.1136/bmj.m1373.

Chakrabarti, S. (2001). *Integrating the document object model with hyperlinks for enhanced topic distillation and information extraction.* Paper presented at the Proceedings of the 10th international conference on World Wide Web.

Chakrabarti, S., Joshi, M., & Tawde, V. (2001). *Enhanced topic distillation using text, markup tags, and hyperlinks.* Paper presented at the Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval.

Chatfield, A., & Brajawidagda, U. (2012). *Twitter tsunami early warning network: a social network analysis of Twitter information flows*. Paper presented at the 23rd Australasian conference on information systems, Geelong, Australia.

De Araujo, J. F. F. E., & Tejedo-Romero, F. (2016). Local government transparency index: determinants of municipalities' rankings. *International Journal of Public Sector Management*.

Earp, J. B., Antón, A. I., Aiman-Smith, L., & Stufflebeam, W. H. (2005). Examining Internet privacy policies within the context of user privacy values. *IEEE Transactions on Engineering Management, 52*(2), 227–237.

Farrahi, K., Emonet, R., & Cebrian, M. (2014). Epidemic contact tracing via communication traces. PloS one, 9(5), e95133.

Greenhalgh, T., Wherton, J., Shaw, S., & Morrison, C. (2020). Video consultations for covid-19: *British Medical Journal Publishing Group*.

Herold, R. (2006). Addressing privacy issues during disaster recovery. *Information Security Journal, 14*(6), 16.

HHS.gov. (2020, April 9). OCR Announces Notification of Enforcement Discretion for Community-Based Testing Sites During the COVID-19 Nationwide Public Health Emergency. Retrieved from https://www.hhs.gov/about/news/2020/04/09/ocr-announces-notification-enforcement-discretion-community-based-testing-sites-during-covid-19.html Accessed on 2021, March 15

Hiller, J. S., & Russell, R. S. (2017). Privacy in crises: The NIST privacy framework. *Journal of Contingencies and Crisis Management, 25*(1), 31–38.

Hoffman, D. L., & Novak, T. P. (1996). Marketing in hypermedia computer-mediated environments: Conceptual foundations. *Journal of Marketing, 60*(3), 50–68.

House, T., & Keeling, M. J. (2010). The impact of contact tracing in clustered populations. *PLoS computational biology*, 6(3), e1000721.

Househ, M. (2012). *Re-examining Perceptions on Healthcare Privacy-Moving from a Punitive Model to an Awareness Model.* Paper presented at the International Conference on Health Informatics.

Jain, P., Gyanchandani, M., & Khare, N. (2016). Big data privacy: A technological perspective and review. *Journal of Big Data, 3*(1), 1–25.

Janssen, M., & van den Hoven, J. (2015). Big and Open Linked Data (BOLD) in government: A challenge to transparency and privacy? *Journal of Organizational Computing and Electronic Commerce, 26*(1–2), 3–13.

John Hopkins School of Public Health. (2021, Feb 2). Variants, Vaccines and What They Mean For COVID-19 Testing. Retrieved from https://www.jhsph.edu/covid-19/articles/variants-vaccines-and-what-they-mean-for-covid19-testing.html Accessed on 2021, March 15

Kapoor, K. K., Tamilmani, K., Rana, N. P., Patil, P., Dwivedi, Y. K., & Nerur, S. (2018). Advances in social media research: Past, present and future. *Information Systems Frontiers, 20*(3), 531–558.

Kim, H.-W., Chan, H. C., & Gupta, S. (2007). Value-based adoption of mobile internet: An empirical investigation. *Decision Support Systems, 43*(1), 111–126.

Lee, J. K., Chang, Y., Kwon, H. Y., & Kim, B. (2020). Reconciliation of privacy with preventive cybersecurity: The bright internet approach. *Information Systems Frontiers, 22*(1), 45–57.

Lee, J. K., Cho, D., & Lim, G. G. (2018). Design and validation of the bright internet. *Journal of the Association for Information Systems, 19*(2), 3.

Liu, R., Gupta, S., & Patel, P. (2021). The Application of the Principles of Responsible AI on Social Media Marketing for Digital Health. *Information Systems Frontiers*, 1–25.

Lyon, D. (2010). Surveillance, power and everyday life *Emerging digital spaces in contemporary society* (pp. 107–120): Springer.

Madewell, Z. J., Yang, Y., Longini, I. M., Halloran, M. E., & Dean, N. E. (2020). Household Transmission of SARS-CoV-2: A Systematic Review and Meta-analysis. *JAMA Network Open, 3*(12), e2031756–e2031756.

Malhotra, N. K., Kim, S. S., & Agarwal, J. (2004). Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information Systems Research, 15*(4), 336–355.

Medicare applications raise anxiety for seniors in pandemic. (2020, April 29). Retrieved from https://www.modernhealthcare.com/medicare/medicare-applications-raise-anxiety-seniors-pandemic Accessed on 2021, March 15

Mendon, S., Dutta, P., Behl, A., & Lessmann, S. (2021). A Hybrid approach of machine learning and lexicons to sentiment analysis: enhanced insights from twitter data of natural disasters. *Information Systems Frontiers*, 1–24.

Mercer. (2020, April 06). COVID-19 raises HIPAA privacy, security issues. Retrieved from https://www.mercer.com/our-thinking/law-and-policy-group/COVID-19-raises-hipaa-privacy-security-issues.html Accessed on 2021, March 15

Oh, O., Agrawal, M., & Rao, H. R. (2011). Information control and terrorism: Tracking the Mumbai terrorist attack through twitter. *Information Systems Frontiers, 13*(1), 33–43.

Petronio, S., & Venetis, M. K. (2017). Communication privacy management theory and health and risk messaging. In *Oxford Research Encyclopedia of Communication*.

Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). Stm: An R package for structural topic models. *Journal of Statistical Software, 91*(1), 1–40.

Palen, L., Anderson, K. M., Mark, G., Martin, J., Sicker, D., Palmer, M., & Grunwald, D. (2010). A vision for technology-mediated support for public participation & assistance in mass emergencies & disasters. *ACM-BCS Visions of Computer Science, 2010*, 1–12.

Palen, L., & Dourish, P. (2003). *Unpacking" privacy" for a networked world*. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.

Prince, C. (2018). Do consumers want to control their personal data? Empirical evidence. *International Journal of Human-Computer Studies, 110*, 21–32.

Ragini, J. R., Anand, P. R., & Bhaskar, V. (2018). Big data analytics for disaster response and recovery through sentiment analysis. *International Journal of Information Management, 42*, 13–24.

Rao, H. R., Vemprala, N., Akello, P., & Valecha, R. (2020). Retweets of officials' alarming vs reassuring messages during the COVID-19 pandemic: Implications for crisis management. *International Journal of Information Management, 55*, 102187.

Rosenfeld, L., Torous, J., & Vahia, I. V. (2017). Data security and privacy in apps for dementia: An analysis of existing privacy policies. *The American Journal of Geriatric Psychiatry, 25*(8), 873–877.

Sadeh, N., Hong, J., Cranor, L., Fette, I., Kelley, P., Prabaker, M., & Rao, J. (2009). Understanding and capturing people's privacy policies in a mobile social networking application. *Personal and Ubiquitous Computing, 13*(6), 401–412.

Sarathy, R., & Robertson, C. J. (2003). Strategic and ethical considerations in managing digital privacy. *Journal of Business Ethics, 46*(2), 111–126.

Saura, J. R., Ribeiro-Soriano, D., & Palacios-Marqués, D. (2021). From user-generated data to data-driven innovation: A research agenda to understand user privacy in digital markets. *International Journal of Information Management*, 102331.

Schillinger, D., Chittamuru, D., & Ramírez, A. S. (2020). From "infodemics" to health promotion: A novel framework for the role of social media in public health. *American Journal of Public Health, 110*(9), 1393–1396.

Signorini, A., Segre, A. M., & Polgreen, P. M. (2011). The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic. *PloS one, 6*(5), e19467.

Smith, A. C., Thomas, E., Snoswell, C. L., Haydon, H., Mehrotra, A., Clemensen, J., & Caffery, L. J. (2020). Telehealth for global emergencies: Implications for coronavirus disease 2019 (COVID-19). *Journal of Telemedicine and Telecare, 26*(5), 309–313.

Solove, D. J. (2008). *Understanding privacy*. Harvard University Press.

Son, J., Lee, J., Oh, O., Lee, H. K., & Woo, J. (2020). Using a Heuristic-Systematic Model to assess the Twitter user profile's impact on disaster tweet credibility. *International Journal of Information Management*, 54, 102176.

Tan, S., Li, Y., Sun, H., Guan, Z., Yan, X., Bu, J., . . . He, X. (2013). Interpreting the public sentiment variations on twitter. *IEEE transactions on knowledge and data engineering, 26*(5), 1158-1170.

The New York Times. (2021, Feb 4). Coming Soon: The 'Vaccine Passport'. Retrieved from https://www.nytimes.com/2021/02/04/travel/coronavirus-vaccine-passports.html Accessed on 2021, March 15

Trocin, C., Mikalef, P., Papamitsiou, Z., & Conboy, K. (2021). Responsible AI for digital health: a synthesis and a research agenda. *Information Systems Frontiers*, 1–19.

Vemprala, N., & Dietrich, G. (2019). *A social network analysis (SNA) study on data breach concerns over social media*. Paper presented at the Proceedings of the 52nd hawaii international conference on system sciences.

Wissinger, C. L. (2017). Privacy Literacy: From Theory to Practice. *Communications in Information Literacy, 11*(2), 378–389.

Zhou, L., Zhang, D., Yang, C. C., & Wang, Y. (2018). Harnessing social media for health information management. *Electronic Commerce Research and Applications, 27*, 139–151.

Paras Bhatt is a PhD student and research assistant in the Department of Information Systems and Cyber Security at the University of Texas at San Antonio. He holds a master's degree in Computer Information Systems from Prairie View A & M University. Prior to his master's he has worked in the Banking industry for Bank of America as a finance and legal associate to monitor and detect fraud in global financial transactions. His research interests include social media, machine learning, natural language processing, mobile healthcare and IoT access control. His work has been presented at various international conferences, workshops, including Bright Internet Global Summit (BIGS), DRW privacy workshop, AMCIS, HICSS, ACM-SACMAT.

Naga Vemprala is an assistant professor in the Pamplin School Of Business at University of Portland, Oregon. He received his PhD in Information Technology from the Department of Information Systems and Cyber Security at the University of Texas at San Antonio. His total professional experience includes nearly 10 years working in IT service industry for leading multinational companies as a software developer and data analyst. His research interests include social media, data analytics, machine learning, and natural language processing. His work has been presented at various international conferences, workshops, including Bright Internet Global Summit (BIGS), DRW privacy workshop, AMCIS, ECIS, HICSS, and published in International Journal of Information Management (IJIM), Decision Sciences Journal (DSJ), and American Behavioral Scientist (ABS).

Rohit Valecha is an associate professor in the Department of Information Systems and Cyber Security at The University of Texas at San Antonio. His Ph.D. is from the University at Buffalo in management science and systems. He has research interests in crisis response, social media and information security and privacy. He has published in several journals, including Management Information Systems Quarterly (MISQ), Journal of the Association for Information Systems (JAIS), and other outlets. His work has received best paper and other awards at the American Conference on Information Systems (AMCIS), Design Science Research on Information Systems and Technology (DESRIST), Conference on Information Systems and Technology (CIST), Workshop on Information Technology and Systems (WITS) and others.

Govind Hariharan is a Professor of Economics and Health Management and Informatics (HMI) Fellow at Kennesaw State University whose research specializes in the application of economics in regulation and public policy. He received his MA in economics from the Delhi School of Economics (1984) and his PhD in economics from the State University of New York at Buffalo (1991). His experience includes positions at the State University of New York at Buffalo and West Virginia University, and consulting for numerous organizations such as the World Bank. His research interests are in the areas of health, finance and technology especially of older adults. His scholarly publications include papers in the Journal of Health Economics, Journal of Risk and Uncertainty, Annals of Operations Research, Journal of Homeland Security and Emergency Management, Finance Research Letters (winner of Stephen Ross Award), Financial Services Review, and the Review of International Economics. His work has received funding from the US Department of State and the National Science Foundation. Dr. Hariharan has been a recipient of numerous teaching awards and has taught in many countries in Asia. He was the recipient of the KSU Community Engagement Award in 2018 and has served on many advisory boards and councils including the Atlanta Regional Commission, Georgia Department of Revenue, Wellstar Institute for Better Health, American Diabetes Association and various health care task forces.

H. Raghav Rao H. Raghav Rao is AT&T Chair Professor of Information Systems and Cybersecurity, Carlos AlvarezCollege of Business and a professor of computer science (courtesy appointment) at the University ofTexas, San Antonio. He graduated from Purdue University. He was a distinguished visiting faculty atSwansea University in the summer of 2020 and 2021 and will be a Fulbright Nehru scholar at IIMBangalore in the summer of 2022 and 2023.