Journal of Biomedical Informatics 111 (2020) 103601

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Defining facets of social distancing during the COVID-19 pandemic: Twitter analysis

Jiye Kwon [a,1], Connor Grady [b,1], Josemari T. Feliciano [c], Samah J. Fodeh [c,d,*]

[a] Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, CT, USA
[b] Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, CT, USA
[c] Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA
[d] Department of Emergency Medicine, Yale School of Medicine, New Haven, CT, USA

ARTICLE INFO

Keywords:
Social distancing
COVID-19
Social media
Infodemiology
Infectious disease outbreak

ABSTRACT

Objectives: Using Twitter, we aim to (1) define and quantify the prevalence and evolution of facets of social distancing during the COVID-19 pandemic in the US in a spatiotemporal context and (2) examine amplified tweets among social distancing facets.

Materials and methods: We analyzed English and US-based tweets containing "coronavirus" between January 23-March 24, 2020 using the Twitter API. Tweets containing keywords were grouped into six social distancing facets: implementation, purpose, social disruption, adaptation, positive emotions, and negative emotions.

Results: A total of 259,529 unique tweets were included in the analyses. Social distancing tweets became more prevalent from late January to March but were not geographically uniform. Early facets of social distancing appeared in Los Angeles, San Francisco, and Seattle: the first cities impacted by the COVID-19 outbreak. Tweets related to the "implementation" and "negative emotions" facets largely dominated in combination with topics of "social disruption" and "adaptation", albeit to lesser degree. Social disruptiveness tweets were most retweeted, and implementation tweets were most favorited.

Discussion: Social distancing can be defined by facets that respond to and represent certain events in a pandemic, including travel restrictions and rising case counts. For example, Miami had a low volume of social distancing tweets but grew in March corresponding with the rise of COVID-19 cases.

Conclusion: The evolution of social distancing facets on Twitter reflects actual events and may signal potential disease hotspots. Our facets can also be used to understand public discourse on social distancing which may inform future public health measures.

1. Introduction

The emergence of novel coronavirus disease (COVID-19) and its etiologic cause, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has prompted an international effort to limit its morbidity and mortality. Social distancing has been one of the primary mitigation strategies in the United States, which compels individuals to stay at home as much as possible and avoid close contact with others to reduce transmission and intensity of the pandemic [1–3]. The realization of social distancing and its intersection with many aspects of our social, educational, professional, and emotional lives is unprecedented in its degree of magnitude. Social distancing can be viewed as a multi-faceted public health measure, involving many stakeholders, practices, and consequences. Development and evaluation of a measure of communication on a public health intervention—here, social distancing—may be beneficial for public health and public policy officials [4].

Twitter is a microblogging platform by which users (tweeters) socialize and tweet through the network. Twitter users contribute original content through tweeting thoughts, opinions, and news and engage with existing content by retweeting, favoriting, and replying to others' tweets. Nearly 65 million people in the U.S. have Twitter accounts with a median user age of 40 years, compared to the U.S. median age of 47 years; the distribution of gender and race is nearly equivalent to the general public according to the Pew Research Center [5]. Capturing data

* Corresponding author at: 300 George Street, PO Box 208009, New Haven, CT 06520, USA.
E-mail address: samah.fodeh@yale.edu (S.J. Fodeh).

[1] First authors who contributed equally to this work.

https://doi.org/10.1016/j.jbi.2020.103601
Received 19 May 2020; Received in revised form 8 September 2020; Accepted 9 October 2020
Available online 14 October 2020

from this younger population is important because younger, working-age individuals are likely to be a more mobile population thereby being the targets of social distancing efforts. Non-protected tweets are publicly available and approximately 3% of tweets are voluntarily geotagged, making them available for research purposes [6]. While other traditional sources, such as news media, have been used to model public health crises [7], Twitter has the advantage of real-time content availability and has already been harnessed during past infectious disease outbreaks. Twitter data has been used for Lyme disease surveillance across the UK and Ireland [8], and was leveraged during and after the 2016 Zika epidemic to understand how public health information was amplified and by which groups of individuals [9,10]. Influenza forecasting has been improved by incorporating Twitter data, reducing forecast error by 17–30%, an even stronger predictor than Google Flu Trends [11]. During the 2014 Ebola outbreak, researchers showed that health information on Twitter was largely spread by broadcasting, allowing public health officials to identify influential users to spread health information [12]. These examples demonstrate the emergence of Twitter as a modern health science tool that can substitute other more time-intensive methods, such as large national surveys, and provide rapid access and large data volume. In recent studies, we have utilized Twitter to better understand critical social and behavioral outcomes such as suicide risk [13,14]. Our findings have shown that certain patterns of word use on Twitter can define many social behaviors associated with suicide risk.

Given that social distancing is the primary behavioral measure taken to limit the spread of COVID-19, this study aims to define the multidimensional measures of social distancing by quantifying its perception, implementation, and impact through Twitter. We achieve our goal by collecting tweets related to "coronavirus". We analyze the contents of the tweets and map them into categories that correspond to the relevant social distancing facets. We then quantify the prevalence of these facets across different states and over time.

1.1. Objectives

In this study, we adopt a supply-based infodemiology approach [4] to analyze facets of social distancing on Twitter with the overarching goal of informing public health policy and practice. Our newly developed information supply measure is the prevalence of tweets concerning the different social distancing facets. The objective of this study is two-fold: (1) define and quantify the prevalence and evolution of social distancing facets in a US spatiotemporal context and (2) examine the most amplified tweets among social distancing facets.

2. Material and methods

2.1. Data collection

We downloaded tweets that contained the term "coronavirus" between January 23rd and March 24th, 2020 using the Twitter API. For each tweet we had the following variables: the users' handle, time of tweet, number of retweets and favorited at the time of collection, geographic coordinates, and text of tweet. Only unique tweets written in English and geotagged to the United States were included in the analyses. Our data is missing tweets from February 29th through March 5th and March 10th-March 13th, 2020. Despite missingness in the data, we were able to capture the overall trends of social distancing facets, thus we skipped imputation. All analyses were performed using R (version 3.6.1) and tidytext package (version 0.2.2).

2.2. Identifying and grouping tweets into social distancing facets

We define social distancing as a multi-faceted intervention where these facets or stages unfold as we are living through the COVID-19 experience. These facets may change in our subsequent studies as COVID-19 events evolve. The facets of social distancing are: (1) *purpose and justification* of imposing this disruptive nation-wide behavioral measure. The purpose is to slow the spread of COVID-19 to levels manageable by healthcare systems. (2) *Implementation* of social distancing to not only avoid mass gatherings but also maintain a 6-feet distance amongst individuals. The advisories imposing this facet translated to closing non-essential businesses, restaurants, schools, and colleges. (3) *Social activity disruptions* which impose travel restrictions and emphasize less human face-to-face interactions; an abrupt change in individuals' and communities' highly interconnected networks. (4) *Adaptation* to social distancing by accepting a new way of life and virtually conducting daily life activities. Examples include online schooling, working remotely through teleconferencing, online food shopping, telehealth-based visits as well as online entertaining through platforms such as Netflix. (5) *Positive emotions* and (6*) negative emotions* facets associated with the emotional response to social distancing. These facets could potentially measure the levels of distress culminating over time as a result of disrupting social behaviors and activities that are usually associated with mental and emotional wellness.

Following the methodology of Yoon and Bakken [15], keywords were first selected based on the authors' understanding of the topic and grouped into 6 themes, which we call facets, following a top-down approach, as is common practice in Twitter-based studies [9,13,14,16,17] (Appendix A). In Twitter studies, automated coding of topics based on keywords has been shown to strongly correlate with manual coding [16] supporting our methodology.

Tweets including terms like "flatten the curve" that tell the motivation for social distancing as to limit viral transmission and protect vulnerable populations were included in the "purpose" facet. The "implementation" facet comprised of tweets capturing content related to institutional closures and public health advisories as to limit exposure to others. The "social disruption" group included tweets concerning the cancellation of social events such as parties, mass gatherings, and other disruptions to daily life. The "adaptation" facet is captured by how individuals adapted their livelihood into virtual settings in the form of online social activity, working remotely, and studying remotely. Tweets that contain words like Zoom, teleconferencing, and Netflix were included. Finally, the last two groups, "positive emotions" and "negative emotions" were designed to capture tweets which provided insights into users' feelings and attitudes related to and coinciding with the COVID-19 pandemic. Facets were not mutually exclusive as a tweet could be assigned to more than one topic.

We examined frequent word pairs (bigrams) for each facet to qualitatively validate that tweets were related to the facet they were assigned to as has been done in other Twitter-based studies as they offer more insight into the sentiment of tweets than examining unigrams on their own [18–20]. A representative sample is shown in Appendix B. For example, among tweets assigned to our social disruption facet, "travel ban" frequently appeared despite only "travel" being a keyword used to define this facet.

2.3. U.S. trends of social distancing facets

The primary analysis concerned a description of the trends of social distancing facets for the entire dataset of U.S. tweets on a daily and weekly basis. The proportion of a given social distancing facet per day is calculated by dividing the number of tweets belonging to that facet by the total number of tweets that day. These proportions were used for relative comparison over the entire study period and compared to events on the ground. The proportion of a given social distancing facet per week is calculated by dividing the number of tweets belonging to that facet by the total number of tweets that week. Trends of these facets were followed, and χ^2 tests were used to evaluate weekly change in tweet proportions.

2.4. Spatiotemporal analysis

Geographic coordinates of the tweets from the different social distancing facets were plotted as pie graphs on a map of the US for January, February, and March. We also aggregated tweets into locations by rounding latitudes to the fourth decimal place and longitudes to the third decimal place. The diameter of the pie graphs corresponded to the volume of tweets in a given location. For simplified visual interpretation, we introduced a month-specific threshold on the number of tweets to be displayed on the map to eliminate noise that does not rise to a meaningful pattern. Hence, only facets that meet the threshold were plotted at corresponding locations.

2.5. Amplified tweets in social distancing facets

To estimate the relative amplification of social distancing facets on Twitter, we calculated two measures of tweet amplification for each facet: average number of retweets per tweet and average number of favorites per tweet. For retweets in a given facet, we divided the number of retweets of all tweets in a facet by the total number of tweets in that facet. Similarly, we divided the number of all favorites of tweets in a facet by the total number of tweets in that facet to yield a score of average favorites per tweet.

3. Results

3.1. Descriptive analysis

Our final tweet dataset of 259,529 unique tweets from 115,485 unique users demonstrated an increasing level of coronavirus-related content over the study period. The total tweet count grew from 11,240 (4.3%) tweets in the final nine days of January to 33,713 (13.0%) tweets in February, and to 214,576 (82.7%) tweets in March. The number of unique users increased in a similar fashion, from 8232 in January tweets, and 18,591 in February, to 100,979 in March. The cities with the highest number of tweets changed by month, but in March alone, 8,045 (3.7%) tweets originated from Los Angeles, CA, 6325 (2.9%) tweets from Manhattan, NY, and 4145 (1.9%) tweets from Washington, D.C.

Themes among the most frequently used hashtags exhibited heterogeneity by month with January tweets using hashtags focused on the early, localized impact of the virus in China, such as #china, #coronavirusoutbreak, and #wuhan. Similar hashtags were observed in February with more pronounced presence; for example, #coronavirusoutbreak increased by 180%. March hashtags overall focused on the U.S. and social distancing with hashtags such as #socialdistancing, #quarantine, and #stayhome. A full list of the top 10 hashtags per month are included in Appendix C.

3.2. U.S. trends of social distancing facets

The daily proportions of tweets belonging to the six social distancing facets are shown in Fig. 1 with representative tweets shown in Appendix D. The proportion of social distancing facet tweets grew from 22.2% in January to 23.4% in February and to 33.3% in March. In January, during the earlier phase of the COVID-19 outbreak, tweets among the facet "negative emotions" predominated with all other facets relatively less pronounced. "Implementation" grew in February, accounting for about 10% of coronavirus-related tweets, and then exhibited a strong upward trend in March increasing 245% from February 28th to March 23rd. "Implementation" tweets comprised 24% of coronavirus-related tweets at their peak on March 23rd. Content relevant to "adaptation" such as working from home and studying online in addition to tweets related to the "purpose" of social distancing also increased over the study period, albeit to a lesser degree. "Adaptation" tweets peaked on March 17th, accounting for 4.3% of tweets and "purpose" topic tweets peaked on March 24th at 3.1%. Alternatively, tweets concerning "social disruption" peaked on February 3rd with an 88% increase compared to the previous day making up 8.7% of the total tweets. This facet declined steadily before peaking again on February 23rd comprising 6.3% of the total tweets, which is a 122% increase in comparison to the previous day. After March 18th, "social disruption" accounted for the lowest number of tweets among the facets. We analyzed the relationship between time (in weeks) and each facet of the social distancing using chi-squared (χ^2) tests. These tests showed that the increase in the proportion of all facets, besides "social disruption", were statistically significant (*P*

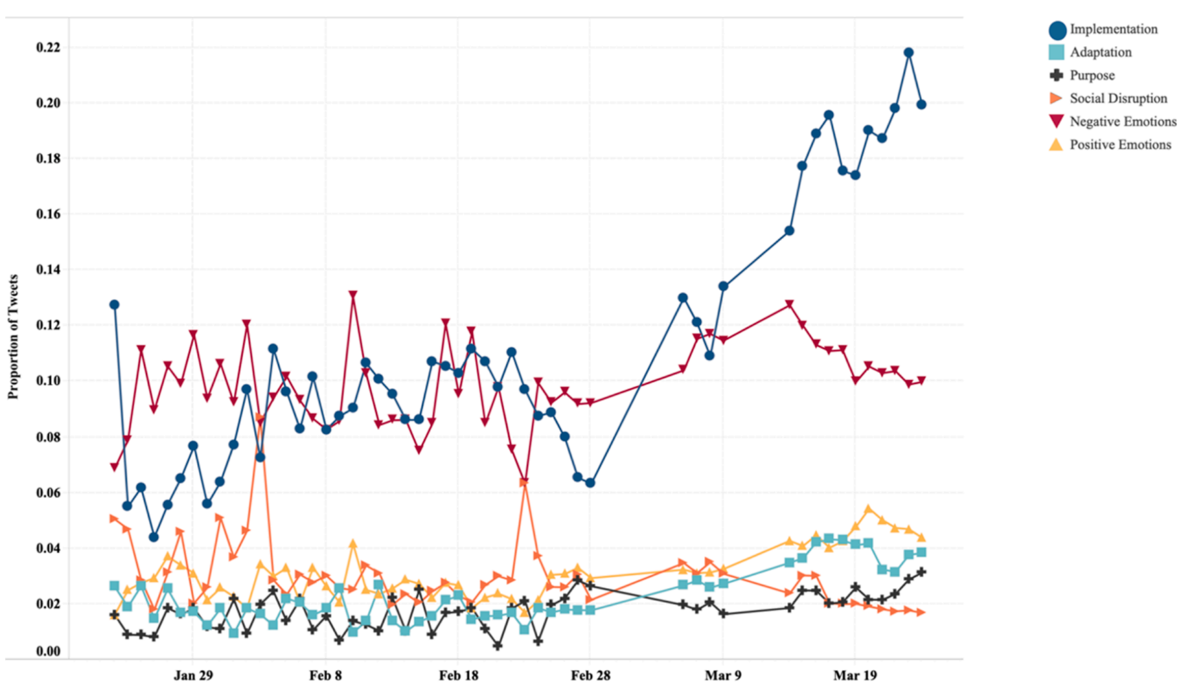

Fig. 1. Trends in daily tweet proportions of social distancing facets from January 23rd to March 24th, 2020.

< .001), and the decrease in social disruption tweets was statistically significant ($P < .001$).

3.3. Spatiotemporal analysis

Social distancing facets were analyzed spatiotemporally using tweets' geotagged location for January, February, and March (Figs. 2–4). Each pie graph corresponds to a cluster of tweets such as in a large city or metropolitan area with the diameter being a function of tweet volume, and overlapping graphs simply indicate nearby cities (e.g. Manhattan and Long Island). Over this period, tweets increased in both volume and locations. Areas generating a high number of social distancing tweets included Los Angeles and San Francisco, CA, Manhattan, NY, Washington, D.C., Chicago, IL, Houston, TX, and Tampa, FL.

The dominant facet per location demonstrates a progression of topic usage. In January and February (Figs. 2 and 3), social distancing tweets were sparsely distributed across the states and mostly generated from highly populated areas on the east and west coasts. "Negative emotions" was the most prevalent facet across the states overall, followed by "implementation" and "social disruption"; however, some heterogeneity was observed. Notably, in January, "implementation" was more prevalent in west coast cities such as Los Angeles and San Francisco compared to other metropolitan areas (e.g. New York City) and some cities including Washington, D.C. had a larger relative number of "social disruption" tweets. In February, as the COVID-19 outbreak began to spread internationally, tweets of social distancing facets were generated from more cities (e.g. New York City, Chicago, and Houston) including "implementation", "social disruption", "negative emotions", and "adaptation".

Fig. 4 depicts the growth of the social distancing facets in March 2020 as seen in the large and more numerous pie graphs. Overall, the predominant facet was "implementation", followed by "negative emotions". "Adaptation" came next in California, New York, Washington, D. C., Texas, Florida, and Illinois. These observations reflect the onset and enforcement of social distancing on the ground. "Positive emotions" was also pronounced in March mirroring discussions about expectations of social distancing to slow the spread of COVID-19. Tweets relating to social distancing's "purpose" became more apparent in March, notably in Kansas, Los Angeles, and New York. The large cluster of tweets seen off the coast of western Florida is believed to be a minor fault of the mapping software which plotted tweets from Tampa, FL off the Florida coast. The other tweet coordinates and their plotted locations are taken in confidence as they appear in cities largely impacted by the pandemic.

3.4. Amplified tweets in social distancing facets

Examining the average number of retweets and favorites of social distancing facets tweets, we noticed that "social disruption" tweets were most amplified through retweeting with 3.74 retweets per tweet on average. This was then followed by "implementation" (3.36 retweets per tweet), "purpose" (3.30 retweets per tweet), "negative emotions" (3.03 retweets per tweet), "positive emotions" (2.22 retweets per tweet), and "adaptation" (2.16 retweets per tweet). On the other hand, "implementation" exhibited the highest number of favorites, an average of 14.84 favorites per tweet, followed by "negative emotions" (12.03 favorites per tweet), "social disruption" (9.7 favorites per tweet), "purpose" (9.4 favorites per tweet), "positive emotions" (9.33 favorites per tweet), and finally, "adaptation" (8.33 favorites per tweet).

4. Discussion

In this Twitter analysis of social distancing-related tweets during the COVID-19 pandemic, several observations emerged. During the early phases of the COVID-19 pandemic in January and February, outbreaks were confined to China and nearby countries; consequently, tweets during these early months were thought to be confined and in reference to the situations in these countries. The U.S. saw a dramatic increase in COVID-19 cases in March, prompting intense national social distancing efforts; accordingly, tweets were regarded as referring to U.S. events, attitudes, and reactions.

In January and February, it was shown in Figs. 2 and 3 that the prevalence of tweets captured locations that started voicing out through the defined facets. Interestingly, locations captured during the early phases of the outbreak are states that were thought to experience the COVID-19 earlier than other states, including Washington, Illinois, and California. In these two maps, the higher prevalence of the "implementation" and "social disruption" facets in the west coast cities (Los Angeles, San Francisco, and Seattle) compared to other major cities could be attributed to their status as major hubs for international flights, many of which originate from East Asia. The "social disruption" facet

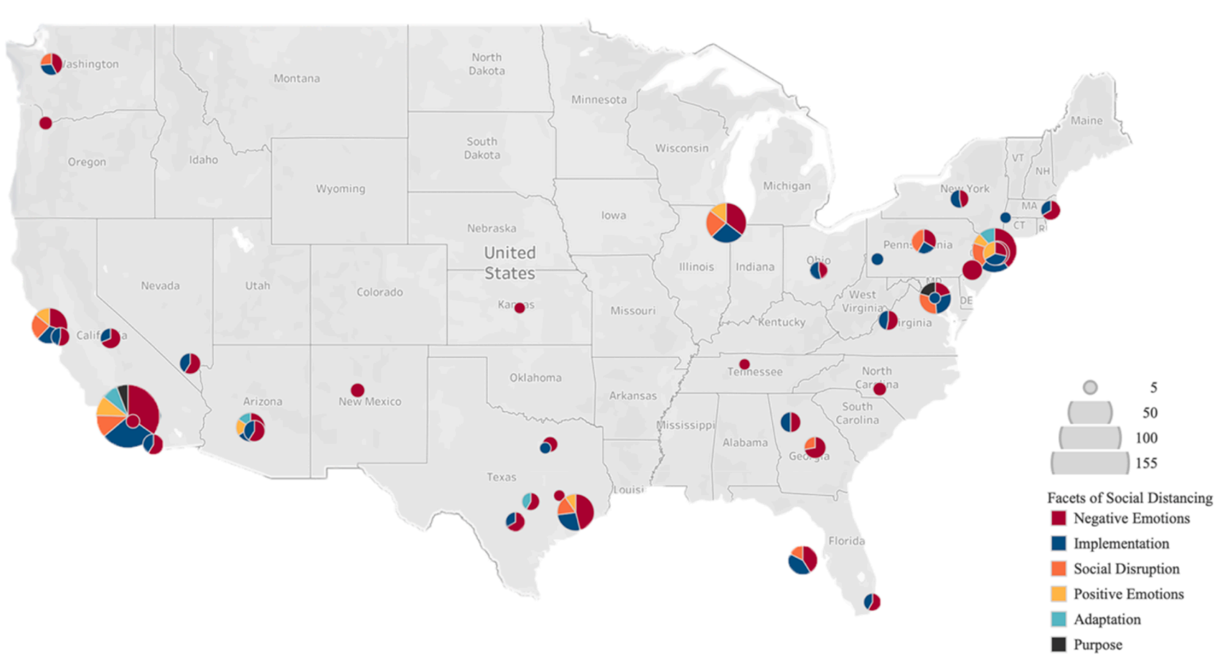

Fig. 2. All facets of social distancing mapped for January 2020. Larger diameter denotes higher volume of tweets. Threshold = 5.

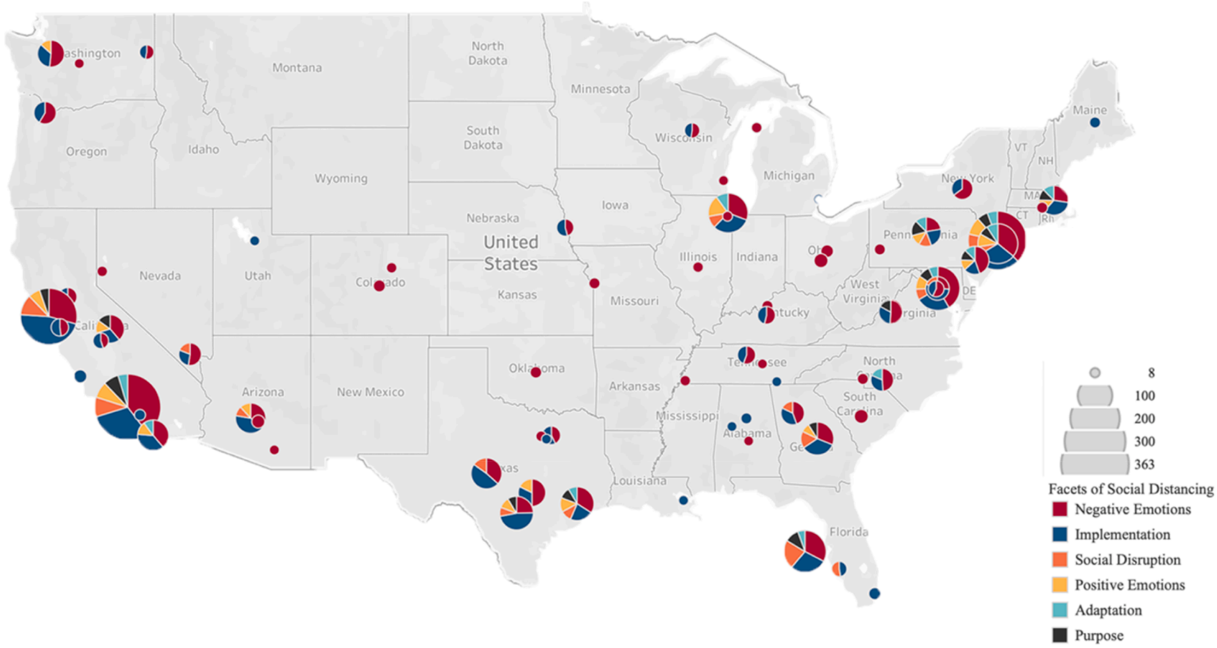

Fig. 3. All facets of social distancing mapped for February 2020. Larger diameter denotes higher volume of tweets. Threshold = 8.

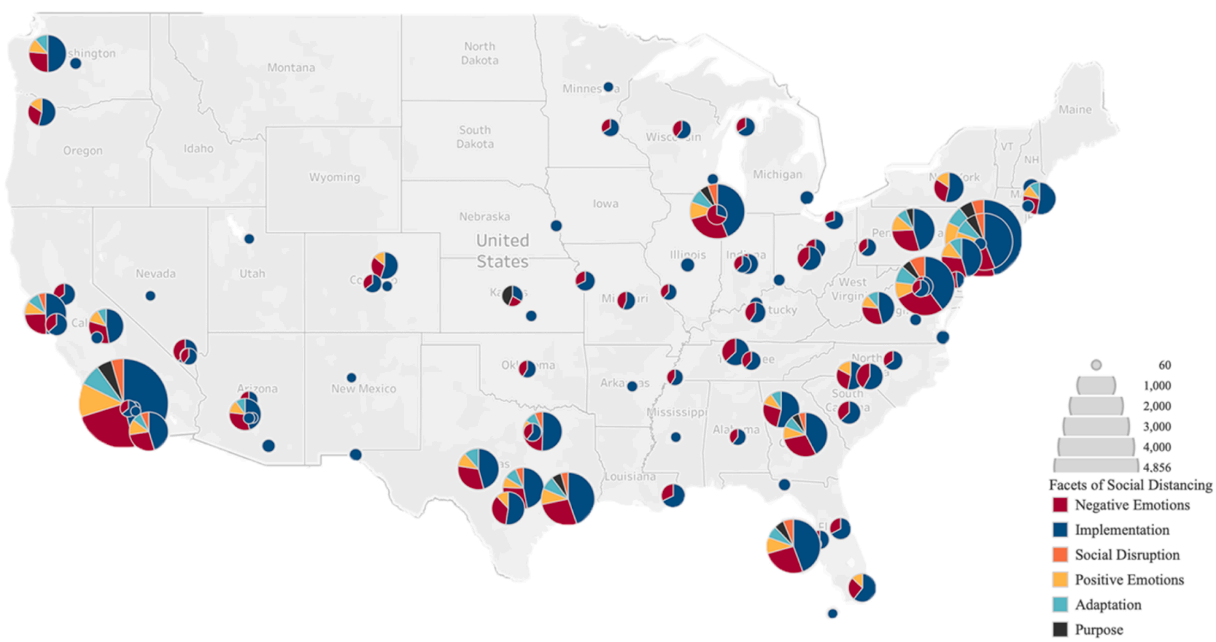

Fig. 4. All facets of social distancing mapped for March 2020. Larger diameter denotes higher volume of tweets. Threshold = 60.

was particularly pronounced during the same time in these cities and others including Manhattan, NY and Washington, D.C., albeit more so in January. In February, we observed two notable peaks of "social disruption" (Fig. 1) on February 3rd and 23rd that correspond and may be reactionary to travel restrictions taken into effect on the evening of February 2nd and Italy going on a nationwide lockdown on the 23rd [21,22]. This facet captured reactions to the imposed advisories and restrictions as it was formed based on words like "travel". Negative emotions were also highly prevalent in many locations, suggesting an expected reaction to a novel infectious disease of concern, uncertainty, and fear [23–25] among U.S. users while the outbreak was largely confined to China.

Turning to the month of March where social distancing efforts were realized in the U.S. and the number of COVID-19 cases increased dramatically, there were certainly more locations across the country that voiced out about these facets of social distancing. Trends in these facets (Fig. 1) demonstrate that tweets relating to the implementation of social distancing increased markedly in March and expanded to many locations across the U.S.. This was likely reactionary to and coincided with nationwide social distancing initiatives such as the Centers for Disease Control and Prevention's recommendation to cancel events with more than 50 people on March 15th [26] and events with more than 10 people for higher risk populations [27]. The peak of implementation in our study was on March 23rd with 24% of all tweets in our dataset related to social distancing implementation. This was around the same time when New York City was declared the U.S. COVID-19 epicenter [28], and taken together with the large volume of tweets from the New York area in March (Fig. 4), perhaps the situation in New York was in part responsible for this peak. Along with rise of implementation, negative emotions increased in early March but subsequentially decreased until March 24th possibly as a result of increased emergence of other facets. Less represented facets, such as “positive emotions”, “purpose”, and “adaptation” were still thought to play key roles in users’ reactions and reflection on social distancing especially across locations such as Los Angeles, TX, FL, CO, and NY. In these locations, pie graphs with larger diameters and higher representations of social distancing facets were seen. The facets “positive emotions” and “purpose” became more pronounced in March, and adaptation reached and maintained its peak across all weekdays on the third week of March as people returned to work and school but on online platforms. This increase in “adaptation” could also explain the decrease in “social disruption”, observed in Fig. 1, perhaps as individuals acclimated to the new routines and practices.

Social distancing tweets as a whole were predominantly generated from the Northeast, South, and West. Relating this to the observed case counts on the ground [29], among the list of states with the highest recorded COVID-19 case count included areas in the Northeast, South region and West coast, specifically California. These figures not only reveal locations with high numbers of facets of social distancing tweets but also reveal locations with relatively low tweet volume. For example, in February, Miami, FL has a low volume of social distancing tweets but grew in March which corresponded with the rise of COVID-19 cases in the city [29]. This suggests that overall volume of social distancing tweets can reflect the relative case count in respective locations.

Our secondary research objective examined amplified tweets to further understand the drivers of the public’s perception during the COVID-19 pandemic and a period of intense social distancing. We defined amplified facets as facets which had high relative engagement, measured by an average number of retweets and favorites. As described in the Risk Amplification through Media Spread model, amplified tweets play a key role in the public’s perception and response [30] and signal which content is meaningful to users [25]. Our study showed that the most amplified facet of social distancing tweets measured by the retweet count was “social disruption” and least amplified was the “adaptation” group. In terms of favorites, “implementation” had the highest number of mean favorites per tweet and the least amplified was “adaptation”. These results suggest that not only do users find these topics meaningful and worth engaging with, but also demonstrate that the topics of “implementation” and “social disruption” were highly broadcasted among social networks. As such, these facets can be leveraged to promote public health actions by echoing the wavelength that the public shares.

Our study responds to the growing interest in the application of infodemiology in public health [4]. We define facets of social distancing with the advantages of real-time, publicly available Twitter data. Through infoveillance and infodemiology, previous studies have shown that Twitter may have the potential to serve as an aid for infectious diseases surveillance tool [31]. A longitudinal study like this one is especially useful during an outbreak [4] for informing intervention efforts by providing a closer look at the prevalence of multiple facets of social distancing. Observing the change in social distancing facets mapped through time provides insight into location-specific content, including possibly when certain localities experience cases. Spatiotemporal analysis of tweets may be of higher importance than just temporal analysis which is often performed. Some have argued that temporal analysis coupled with a spatial dimension tend to match the actual infectious disease epidemiology and have potential to detect possible outbreaks or early signals of a potential outbreak [8].

5. Limitations

There are several limitations to the current study. Twitter users are not entirely representative of the U.S. population as they tend to be younger, more educated, and more likely to identify as Democrats [5]. However, younger individuals likely play a large role in transmitting SARS-CoV-2, so understanding the practices of this group is very insightful [32]. And while a 2018 survey showed that the top 10% of tweeters generated 80% of content on the platform [5], our high user to tweet ratio suggests this is not a limitation in our study. At the time of initiating data collection (January 23, 2020), there was no cohesive nomenclature nor official name of the disease, so we chose a term widely known and used: “coronavirus”. Only tweets including the word “coronavirus” were downloaded from the Twitter API and included in the analysis. Over the course of the pandemic, terminology has shifted toward other nomenclature such as COVID-19, SARS-CoV-2, or referred to colloquially as “corona”, and in some circles as the “Wuhan virus” or “China virus”. However, these other names frequently appeared in our data set and were captured to an extent. We demonstrated that “coronavirus” was highly used as we collected over 250,000 unique tweets representing approximately 3% of all related tweets as we limited tweets to those geotagged to the U.S.. Additionally, the number a tweet has been retweeted is dependent on when the data is collected. Our data collection practices were not consistent in regard to time of day. Nevertheless, given the long period of data collection, this should not be concerning. Finally, tweets belonging to positive and negative emotion facets were classified in a way that did not necessitate they be in regard to social distancing topics (as did other facets) but only to coronavirus. Still, these tweets are useful as they coincide with intensive social distancing efforts and thus offer important insight into how individuals reacted emotionally during this period.

6. Conclusion

We conclude that social distancing can be defined in terms of facets which may respond to certain moments and events in a pandemic. Social distancing efforts during the COVID-19 pandemic are unprecedent and measurement of these practices is challenging, but Twitter can be applied to understanding the public’s practice of and response to social distancing. The spatiotemporal analysis of multiple facets of social distancing in this study helps evaluate the penetration of information and has the potential to provide insights in evaluating public health measures.

CRediT authorship contribution statement

Jiye Kwon: Conceptualization, Formal analysis, Investigation, Writing - original draft, Writing - review & editing. **Connor Grady:** Conceptualization, Formal analysis, Investigation, Writing - original draft, Writing - review & editing. **Josemari T. Feliciano:** Investigation, Data curation. **Samah J. Fodeh:** Conceptualization, Formal analysis, Writing - review & editing, Supervision, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank Dr. Daniel Weinberger for his contributions to the early stages of this project and Dr. Glen Nowak for his inputs on Twitter methodology and guidance on health communication. We would also like to acknowledge the use of GitHub as a platform for publishing our dashboard titled “Twitter-based Social Distancing” under https://samahfodehlab.github.io/. We plan on uploading all source code and Tweet IDs pertaining to this study on GitHub as well.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jbi.2020.103601.

References

[1] Coronavirus disease 2019 (COVID-19): Prevent Getting Sick, Centers for Disease Control and Prevention, 2020.
[2] A. Wilder-Smith, D.O. Freedman, Isolation, quarantine, social distancing and community containment: pivotal role for old-style public health measures in the novel coronavirus (2019-nCoV) outbreak, J. Travel Med. 27 (2020).
[3] R.J. Hatchett, C.E. Mecher, M. Lipsitch, Public health interventions and epidemic intensity during the 1918 influenza pandemic, Proc. Natl. Acad. Sci. USA 104 (2007) 7582–7587.
[4] G. Eysenbach, Infodemiology and Infoveillance: Framework for an Emerging Set of Public Health Informatics Methods to Analyze Search, Communication and Publication Behavior on the Internet, J. Med. Internet Res. 11 (2009), e11.
[5] P.R. Center, Sizing Up Twitter Users, 2019.
[6] L. Sloan, J. Morgan, Who Tweets with Their Location? Understanding the Relationship between Demographic Characteristics and the Use of Geoservices and Geotagging on Twitter, PLoS ONE 10 (2015), e0142209.
[7] S. Collinson, K. Khan, J.M. Heffernan, The Effects of Media Reports on Disease Spread and Important Public Health Measurements, PloS one 10 (2015) e0141423-e.
[8] J.S.P. Tulloch, R. Vivancos, R.M. Christley, A.D. Radford, J.C. Warner, Mapping tweets to a known disease epidemiology; a case study of Lyme disease in the United Kingdom and Republic of Ireland, J. Biomed. Inform. X 4 (2019), 100060.
[9] S. Vijaykumar, G. Nowak, I. Himelboim, Y. Jin, Virtual Zika transmission after the first U.S. case: who said what and how it spread on Twitter, Am. J. Infect. Control 46 (2018) 549–557.
[10] K.W. Fu, H. Liang, N. Saroha, Z.T.H. Tse, P. Ip, I.C. Fung, How people react to Zika virus outbreaks on Twitter? A computational content analysis, Am. J. Infect. Control 44 (2016) 1700–1702.
[11] M.J. Paul, M. Dredze, D. Broniatowski, Twitter improves influenza forecasting, PLoS Curr. 6 (2014).
[12] H. Liang, I.C. Fung, Z.T.H. Tse, J. Yin, C.H. Chan, L.E. Pechta, et al., How did Ebola information spread on twitter: broadcasting or viral spreading? BMC Public Health. 19 (2019) 438.
[13] S.J. Fodeh, J.L. Goulet, C. Brandt, A.-T. Hamada, Leveraging Twitter to better identify suicide risk, 2017, p. 1–7.
[14] F. Samah Jamal, D.B. Edwin, W. Rixin, S. Dennis, B. Robert, G. Joseph Lucien, et al., Suicide Risk on Twitter, Int. J. Knowl. Discov. Bioinform. (IJKDB) 8 (2018) 1–17.
[15] S. Yoon, S. Bakken, Methods of knowledge discovery in tweets, NI 2012 (2012) 463.
[16] C. Chew, G. Eysenbach, Pandemics in the Age of Twitter: Content Analysis of Tweets during the 2009 H1N1 Outbreak, PLoS ONE 5 (2010), e14118.
[17] R. Gaspar, S. Gorjão, B. Seibt, L. Lima, J. Barnett, A. Moss, et al., Tweeting during food crises: A psychosocial analysis of threat coping expressions in Spain, during the 2011 European EHEC outbreak, Int. J. Hum. Comput. Stud. 72 (2014) 239–254.
[18] N. Hassan, W. Gomaa, G. Khoriba, M. Haggag, Credibility Detection in Twitter Using Word N-gram Analysis and Supervised Machine Learning Techniques, Int. J. Intell. Eng. Syst. 13 (2020) 291–300.
[19] M. Conway, S. Doan, A. Kawazoe, N. Collier, Classifying disease outbreak reports using n-grams and semantic features, Int. J. Med. Inform. 78 (2009) e47–e58.
[20] Y. Wan, Q. Gao, An Ensemble Sentiment Classification System of Twitter Data for Airline Services Analysis, in: 2015 IEEE International Conference on Data Mining Workshop (ICDMW), 2015, pp. 1318–1325.
[21] H. Kretchmer, Key milestones in the spread of the coronavirus pandemic, World Economic Forum, 2020.
[22] Press Briefing by Members of the President's Coronavirus Task Force. January 31, 2020 ed, The White house, Washington DC.
[23] H. Jeong, H.W. Yim, Y.J. Song, M. Ki, J.A. Min, J. Cho, et al., Mental health status of people isolated due to Middle East Respiratory Syndrome, Epidemiol. Health. 38 (2016), e2016048.
[24] L. Hawryluck, W.L. Gold, S. Robinson, S. Pogorski, S. Galea, R. Styra, SARS control and psychological effects of quarantine, Toronto, Canada, Emerg. Infect. Dis. 10 (2004) 1206–1212.
[25] S. Li, Y. Wang, J. Xue, N. Zhao, T. Zhu, The Impact of COVID-19 Epidemic Declaration on Psychological Consequences: A Study on Active Weibo Users, Int. J. Environ. Res. Public Health. 17 (2020).
[26] D. Kopecki, CDC recommends canceling events with 50 or more people for the next eight weeks throughout US, CNBC2020.
[27] Get Your Mass Gatherings or Large Community Events Ready, Coronavirus Disease 2019 (COVID-19), CDC, 2020.
[28] H. Secon, A. Woodward, D. Mosher, A comprehensive timeline of the new coronavirus pandemic, from China's first COVID-19 case to the present, Business Insider, 2020.
[29] E. Dong, H. Du, L. Gardner, An interactive web-based dashboard to track COVID-19 in real time, Lancet Infect. Dis. (2020). S1473-3099(20)30120-1.
[30] S. Vijaykumar, Y. Jin, G. Nowak, Social Media and the Virality of Risk: The Risk Amplification through Media Spread (RAMS) Model 12 (2015) 653.
[31] S. Masri, J. Jia, C. Li, G. Zhou, M.-C. Lee, G. Yan, et al., Use of Twitter data to improve Zika virus surveillance in the United States during the 2016 epidemic, BMC Public Health. 19 (2019) 761.
[32] N. Bai, Coronavirus Is Sickening Young Adults and Spreading Through Them, Experts Say, 2020.