

Mapping the temporally evolved geographic disparities of public perceptions of COVID-19 through social media

Binbin Lin*

Department of Geography, Texas A&M University, Binbin Lin - bb2020@tamu.edu

* Corresponding author

Keywords: COVID-19, social media, perception, awareness, sentiment

Abstract:

Since the emergency and global transmission of COVID-19, the U.S. government implemented several measures requiring or suggesting residents wear masks, keep social distancing, or stay at home to control the spread of the coronavirus. However, residents in different communities showed diverse, changing public perceptions toward COVID-19 and its management strategies, and suffered uneven health impacts and recoveries. Whether the spatial temporal disparities in public perceptions affect the pandemic's health impacts is unknown and needs to be investigated to combat the current and future infectious diseases.

Social media platforms have been awash with messages related to COVID-19, proving immense valuable data for sensing the pulse of pandemic. Twitter is one of the most populous social media worldwide, allowing users to post messages up to 280 characters tagged with the posting time, as well as the coordinates of posting based on the user preference. As a result, the extensive Twitter data embedded with time and location information during the pandemic offers an innovative approach to observe the public perceptions to COVID-19 at multiple spatial-temporal scales in near real-time. However, studying human behaviors during the pandemic and drawing scientific conclusions from social media data are challenging. First, social media data contain a sheer amount of noisy information irrelevant to Covid-19. Second, data bias issue is a persistent challenge in social media data mining research. Existing research has found that Twitter data are biased toward younger, well-educated, and wealthier population living in urban communities. Public awareness and sentiment analysis based on Twitter data without considering the underlying demographic biases may overlook the perception of certain social groups and lead to unfair estimations.

This study collected Twitter data in the U.S. from January 2020 to December 2021. The overarching research question is: What are the spatial-temporal disparities of public perceptions to COVID-19, including awareness and sentiment, reflected on social media? The objectives are three-fold: (1) to develop a novel social media data mining framework for tracking demographic-unbiased public awareness and sentiment; (2) to quantify the multiscale disparities of public awareness and sentiment to COVID-19; (3) to examine the effects of public perceptions on COVID-19 health impacts. This study will yield insights into the applications of social media data in public health and inform the government to mitigate risks from current and future epidemics.

Figure 1 shows the framework of Twitter data mining for demographic-unbiased public awareness and sentiment toward Covid-19. There were five main steps. First, we collected Twitter data through the Twitter Academic Application Programming Interface (API), and geolocated tweets with geotags. Second, we set a list of COVID-19 related keywords to identify COVID-19 related tweets and separated the whole Twitter data into COVID-19 Twitter data and background Twitter data. Third, we calculated the percentage of COVID-19 related Twitter data over the whole Twitter data as the Ratio index to represent public awareness to COVID-19, and estimated sentiment (negative, neutral, or positive) of each Twitter user based on all COVID-19 related tweets through VADER method. Fourth, we employed multimodal, multilingual, and multi-attribute (M3) model to detect the age and gender of Twitter users, and adjusted the demographic bias of Twitter data based on the difference between demographic structure of Twitter users and population. Finally, Sentiment Adjusted by Demographics (SAD) index and awareness (Ratio) index were yielded in this framework. The SAD index represented the demographic-unbiased percentage of negative Twitter users toward COVID-19.

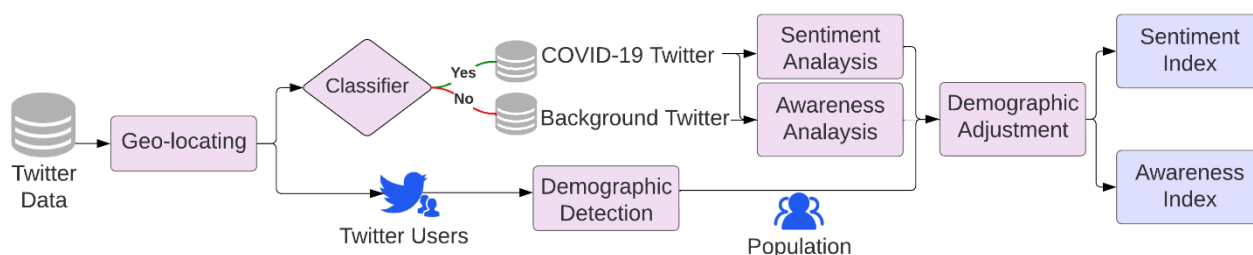


Figure 1. Framework of Twitter data mining for demographic-unbiased public awareness and sentiment toward COVID-19.

Figure 2 shows the temporal patterns of the daily Ratio (per million population) and SAD indexes in the U.S. in 2020, which reflect the general national trends of public awareness and sentiment toward COVID-19. The daily Ratio index showed one peak in the outbreak phase from March to May in 2020 and stayed a low value throughout the remaining period. SAD index values showed three peaks. The highest peak was observed in January and February in 2020, and the widest peak was from June 2020 to January 2021. The third peak was from June 2021 to the end of 2021.

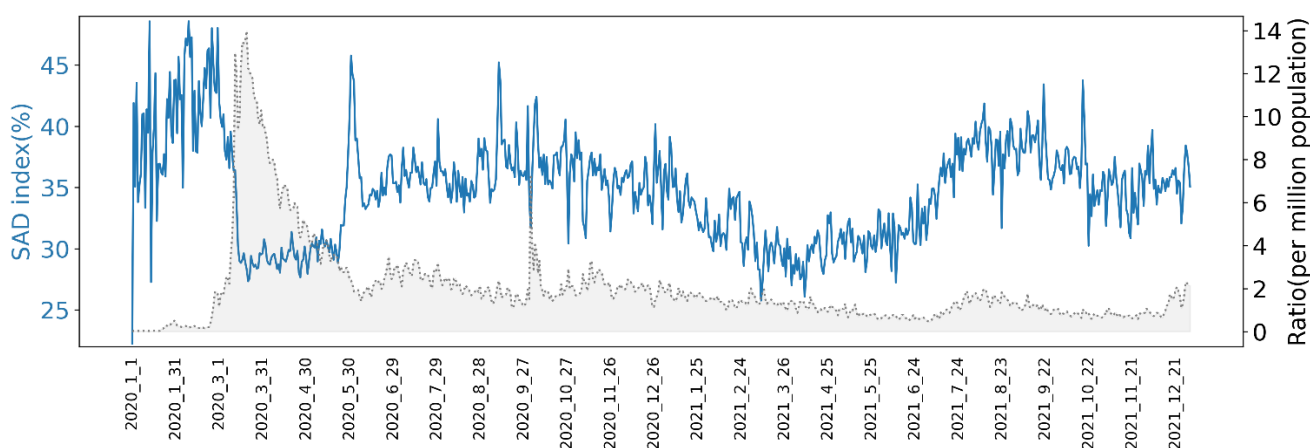


Figure 2. Time series of daily public awareness (Ratio index) and sentiment (SAD index) in the U.S. in 2020 and 2021

The overall Ratio (per thousand population) and SAD indexes at the state level in 2020 and 2021 were computed to uncover the spatial disparities in COVID-19 perceptions (Figure 3). Wyoming, Montana, and Alaska showed highest percentage of negative Twitter users, while COVID-19 discussions on Twitter in Vermont and Hawaii are more optimistic than in other states. As for the public awareness, Nevada presented the highest Ratio index value, followed by the New York, California, Massachusetts, and Maryland.

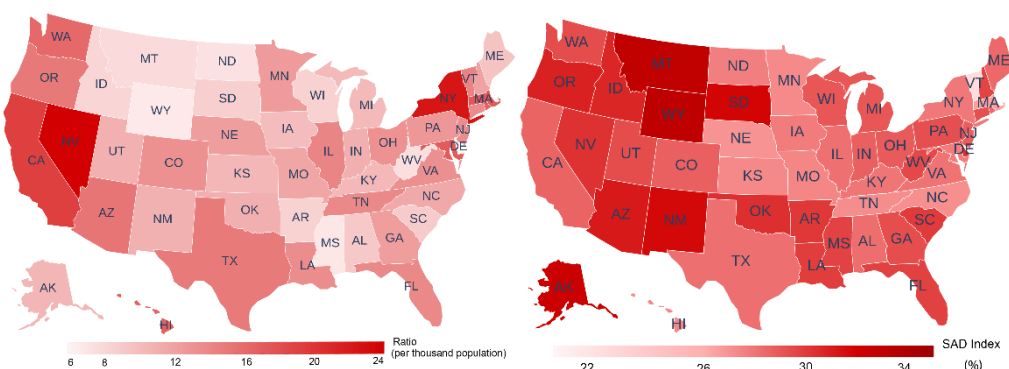


Figure 3. Spatial distribution of yearly public awareness (Ratio) in 2020 and overall sentiment (SAD index, representing the demographic-unbiased percentage of negative Twitter users) in 2020 and 2021

The proposed framework can be used to monitor demographic-unbiased human behaviors in social media analytics. The results yielded from this study can inform policymakers to develop targeted measures for public perception toward COVID-19, and thereby better control the pandemic spread.

Note: Due to the time limitation, Ratio index in this abstract was calculated by the number of COVID-19 related tweets over population. It will be updated to the percentage of COVID-19 related Twitter data over the whole Twitter data in the future.