Multi-Granular Geovisualizations of Public Perceptions of COVID-19 through the Lens of Twitter

Binbin Lin[a]*, Lei Zou[a], Mingzheng Yang[a] and Bing Zhou[a]

[a] Department of Geography, Texas A&M University, College Station, USA
* bb2020@tamu.edu

Keywords: COVID-19, social media, situational awareness, sentiment analysis, demographic bias

Introduction

The COVID-19 pandemic has led to a global crisis. In the U.S., for instance, the COVID-19 has caused over one million fatalities and profound societal impacts over the past two years (Onyeaka et al. 2021). Several measures were proposed to control the spread of the coronavirus, e.g., wearing masks, keeping social distance, sheltering in place, or taking vaccinations. However, residents in different communities showed diverse, changing awareness and sentiment toward COVID-19 and its management strategies, affecting their compliance with pandemic control measures. Therefore, it is necessary to monitor spatial-temporal variations of public perceptions to unravel COVID-19 and its relevant policies' societal responses and inform effective pandemic management.

Social media platforms have been awash with messages regarding public concerns and reactions to COVID-19, providing immense valuable data for sensing the pulse of the pandemic. Preceding studies have put efforts into tracking public awareness and sentiment toward COVID-19 through social media (Lin et al. 2022; Manguri, Ramadhan, and Amin 2020). However, previous work assumes social media users as an impartial representation of the whole population and overlooks the bias in users' age and gender distribution.

This study aims to address the demographic bias in social media by granular analysis and visualizations of public perceptions of COVID-19 through the lens of Twitter. The objectives are two-fold: (1) to develop a framework for measuring public awareness and sentiment toward COVID-19 by demographics; (2) to reveal the multiscale spatial-temporal disparities of Covid-19 perceptions by different social groups. We collected Twitter data (referred to as tweets) in the U.S. in 2020 and identified users' demographics, i.e., age and gender. Public awareness and sentiment by age, gender, and COVID-19 phases were computed at three spatial scales, national, state, and county. The knowledge gained from this study will shed light on COVID-19 perceptions among diverse social groups in each phase and inform governments to customize pandemic mitigation strategies for different communities. The proposed framework can be used to remove the demographic bias in social media analytics.

Method

As shown in Figure 1, this study consists of two steps. The first step is Twitter data collection, cleaning, and information extraction. Geotagged Twitter data in the U.S. were obtained through the Twitter Academic Application Programming Interface (API). Based on each tweet's source, we removed tweets posted by bots and identified COVID-19-related tweets by keywords filtration.
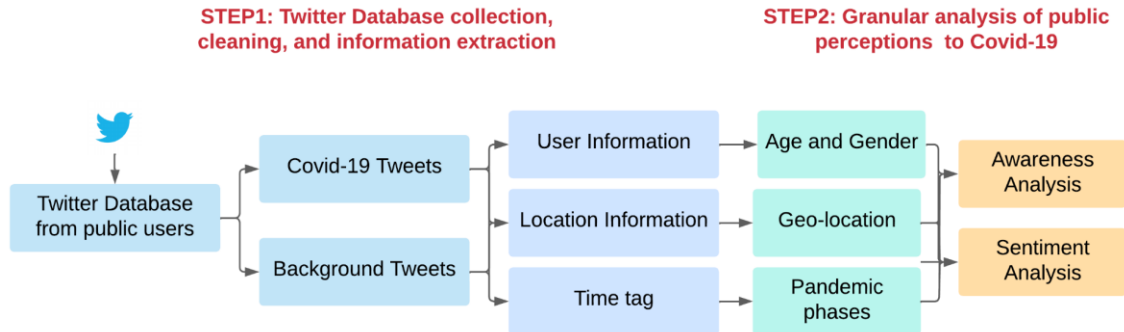

Figure 1: The analysis workflow of Twitter-based Covid-19 perception assessment

Next, we determined the age and gender, COVID-19 phases, and locations of Twitter data for subsequent analysis. First, this study employed an open-source package, multimodal, multilingual, and multi-attribute (M3) model, for classifications of Twitter users' ages and genders (Wang et al. 2019). Twitter users were categorized as below 18, 19-29, 30-39, and above 40 years old, and female and male based on users' names, usernames, profile images, and biographical information. Second, tweets were geo-located at the state and county levels based on the geotags. Finally, we divided the pandemic in the U.S. in 2020 into three stages, outbreak, ferment, and full-blown, according to the major peaks of seven-day averages of confirmed cases.

Two types of public perceptions, awareness and sentiment, were quantified and visualized by age, gender, and pandemic phases. Public awareness of COVID-19 was evaluated by the Ratio index (Lin et al. 2022), the percentage of COVID-19 related tweets among all tweets. The public sentiment was calculated using the VADER model (Hutto and Gilbert 2014), a rule/lexicon-based sentiment analysis toolset that can automatically assign sentiment scores (-1 to 1) to texts.

Results

Temporal trends in the U.S.

Figure 2 shows the temporal patterns of the daily Ratio and Sentiment 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 shows two peaks. The highest and widest climax was sustained in the outbreak phase from March to May, and another small peak was in early October. Sentiment index values were negative from January 20th to March 11th and stayed primarily positive throughout the remaining year, with a few short negative periods in the full-blown phase.
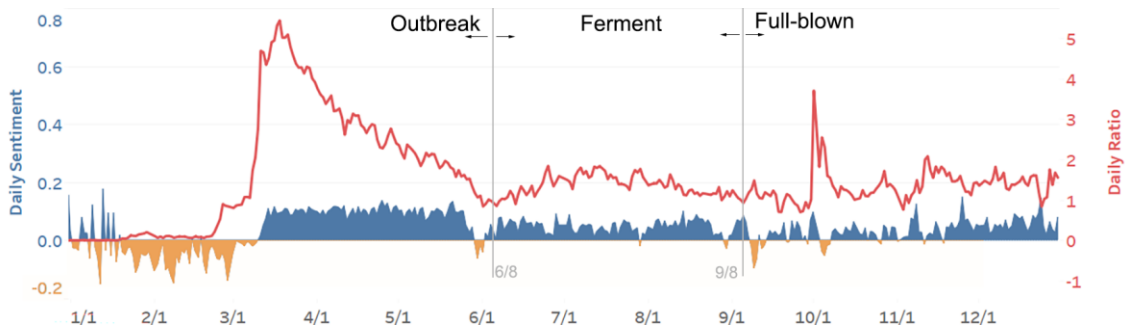
Figure 2: Daily trends of Ratio and Sentiment indexes in the U.S.

Spatial-temporal patterns at the state level

The yearly Ratio and Sentiment indexes at the state level in 2020 were computed to uncover the coarse-grained spatial disparities in COVID-19 perceptions (Figure 3). Nevada was the only state showing an overall negative sentiment, while COVID-19 discussions on Twitter in Vermont and Hawaii are more optimistic than in other states. As for the public awareness, Vermont presented the highest Ratio index value, followed by the District of Columbia, New Hampshire, Montana, and Maine, the majority of which are located in the Northeastern U.S.
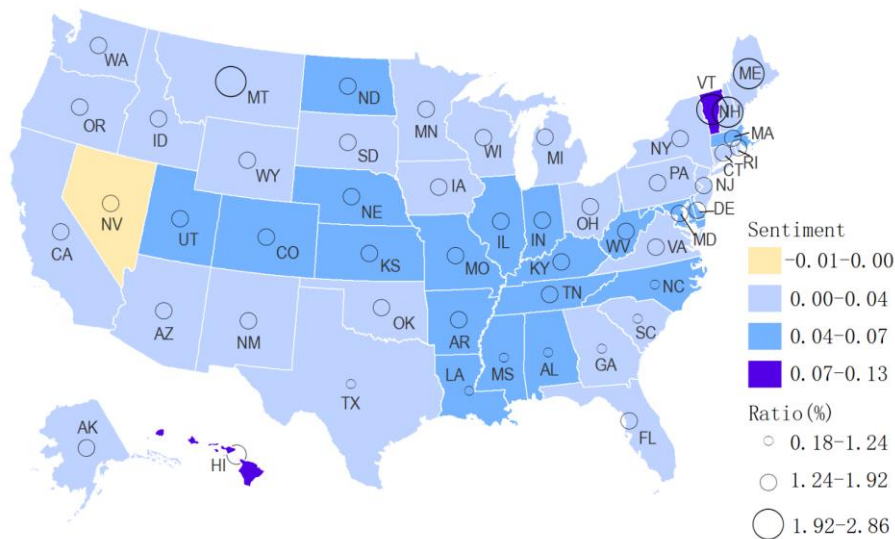

Figure 3: State map of sentiment and Ratio indexes in the U.S.

The monthly state-level sentiment reveals the evolving human emotions toward COVID-19 in 2020 (Figure 4). Most states showed similar temporal patterns in that the Sentiment index was negative until around March and turned positive from April to December 2020. The Sentiment index in North Dakota, Utah, Colorado, Arkansas, Tennessee, Louisiana, Mississippi, and Hawaii was positive for most of 2020. It is worth mentioning that Twitter users from Nevada and Wyoming had negative sentiment toward COVID-19 in two periods of 2020. One was the pre-outbreak phase, and the other was the full-blown stage.
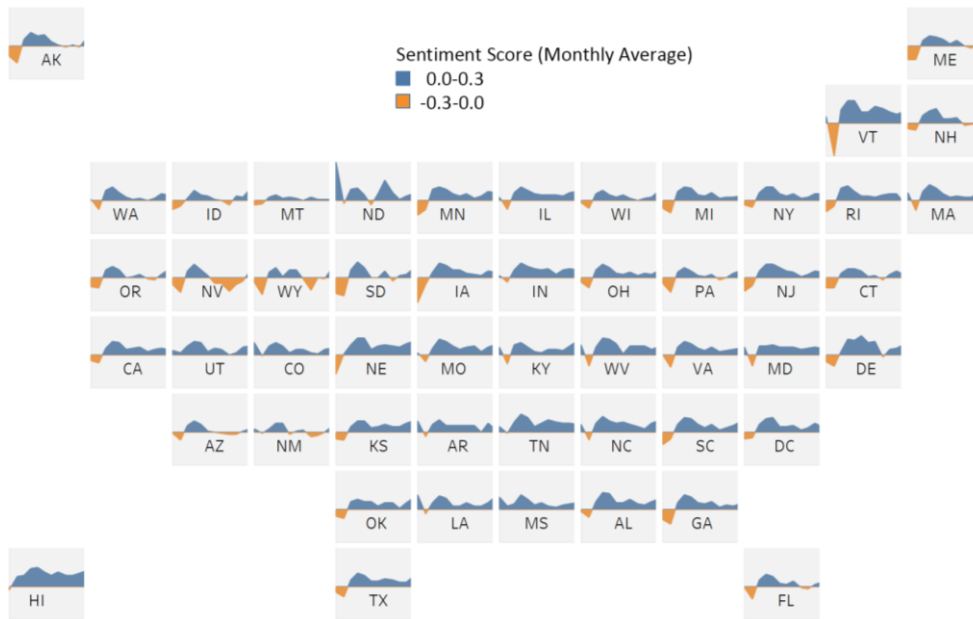
Figure 4: Monthly trends of sentiment score in the U.S. at the state level.

Figure 5 depicts the daily temporal trends (seven-day moving average) of the Ratio index at the state level. Twitter users in all states paid scarce attention to COVID-19 in January and February but expressed a sharp uptick in awareness in March. The Ratio index began to decline in April and remained relatively stable from June to December, with only two weak peaks in July and November among most states.
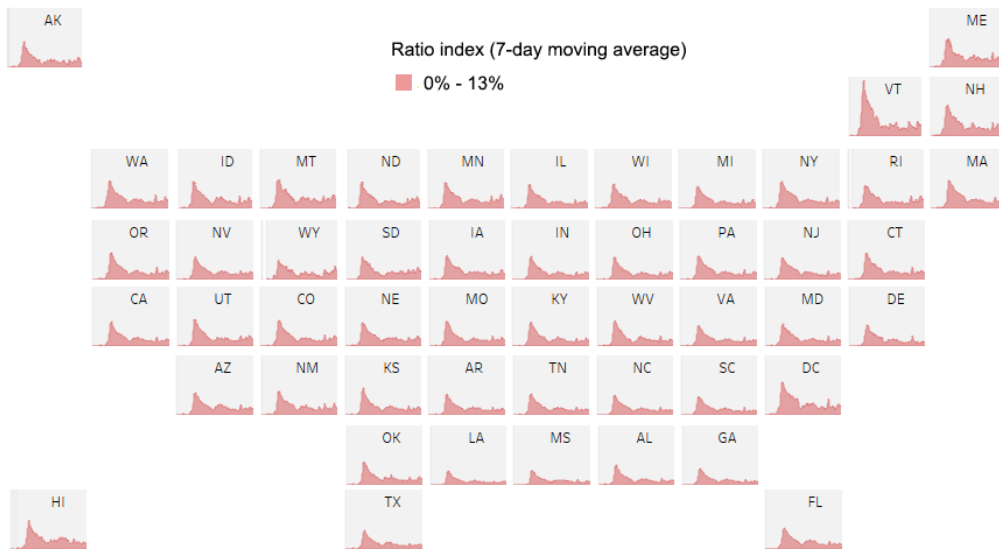

Figure 5: Ratio index (7-day moving average) in the U.S. at the state level.

Demographical and spatial disparities at the county level

Figure 6 displays the spatial disparities of Ratio indexes in the U.S. at the county level. In general, counties in Western U.S. and North-eastern U.S. demonstrated relatively higher Ratio index values.
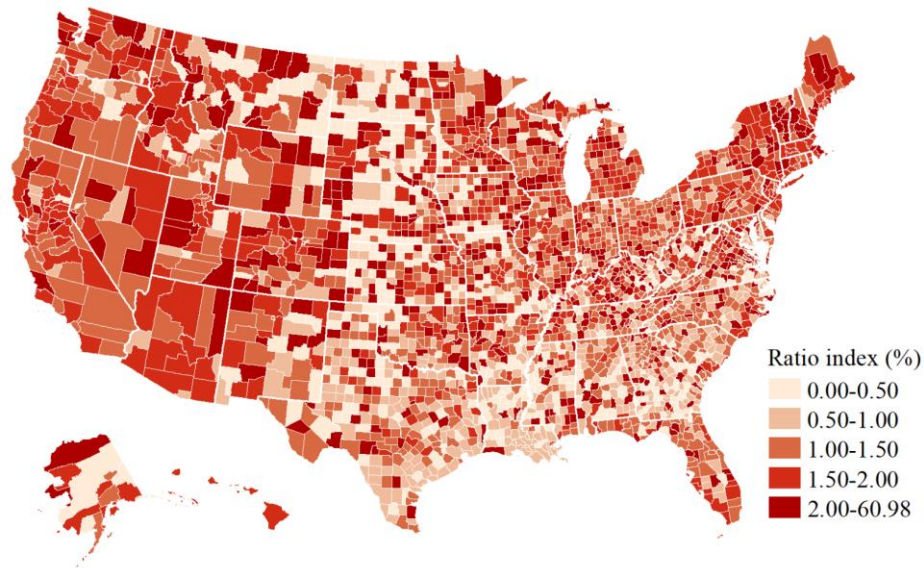
Figure 6: Ratio index in the U.S. at the county level.

As shown in Figure 7, counties that suffered from negative COVID-19-related emotions were scattered all over the U.S. and inconsistent among different gender-age residents. It confirms that COVID-19 triggered severe and uneven mental stress in various social groups. Specifically, male and female Twitter users under 18-year-old in 36.63% and 34.57 % of counties showed negative yearly sentiment toward Covid-19, higher than any other social group. Female Twitter users under 18-year-old in 1.70% of counties suffered extremely negative emotions (Sentiment index <-0.75). It reveals that children and teenagers, especially girls, are mentally more vulnerable to the pandemic.
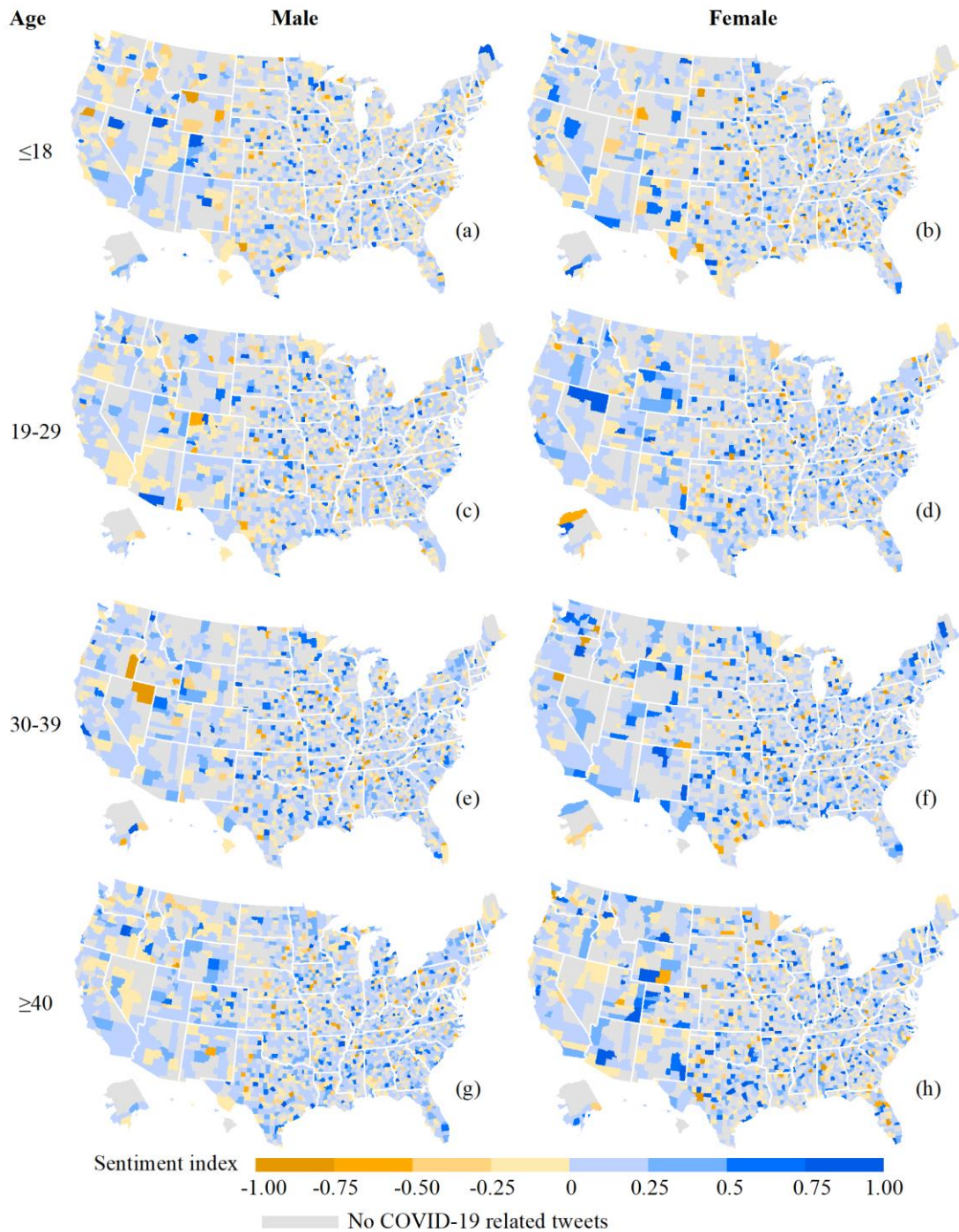
Figure 7: The COVID-19 Sentiment index by age and gender in the U.S. at the county level. (a), (c), (e), and (g) are sentiment maps among male residents under 18 ages, 19-29, 30-39, and older than or equal to 40, respectively. (b), (d), (f), and (h) are sentiment maps among female residents of the four age groups.

Discussion and Conclusion

This research visualized and revealed the disparities in public awareness and sentiment toward COVID-19 by age and gender and COVID-19 phases in 2020 at three spatial scales in the U.S. The visualizations and analysis yield valuable findings. First, significant

spatial-temporal disparities exist in public perceptions of COVID-19. Twitter users in the U.S. were most concerned about this pandemic in March. Users in Northeastern U.S showed the highest awareness to epidemic at both the state and county levels. Nevada is the only state showing a negative yearly sentiment toward COVID-19 on Twitter, with two depressed periods in the pre-outbreak and full-blown phases. Second, different social groups show distinct levels of COVID-19 perceptions on Twitter. Female residents under 18-year-old showed the lowest sentiment toward the pandemic, indicating that they are mentally more vulnerable to the disturbance and impact caused by COVID-19 compared to other social groups. These results pave the way for granular social sensing of the pulse of the pandemic among diverse social groups and inform policymakers to develop targeted measures for distinct communities to diminish the COVID-19 negativity.

Future research will focus on two directions. First, the COVID-19 discussions on social media are evolving through time, and discussion topics varied from different population groups. Therefore, we will explore the changing COVID-19-related themes on social media in various social groups. Second, we will establish an algorithm considering Twitter user demographics for unbiased COVID-19 awareness and sentiment geovisualizations.

References

Hutto, C., and Eric Gilbert. 2014. "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text." *Proceedings of the International AAAI Conference on Web and Social Media* 8(1):216–25.

Lin, Binbin, Lei Zou, Nick Duffield, Ali Mostafavi, Heng Cai, Bing Zhou, Jian Tao, Mingzheng Yang, Debayan Mandal, and Joynal Abedin. 2022. "Revealing the Linguistic and Geographical Disparities of Public Awareness to Covid-19 Outbreak through Social Media." *International Journal of Digital Earth* 15(1):868–89. doi: 10.1080/17538947.2022.2070677.

Manguri, Kamaran H., Rebaz N. Ramadhan, and Pshko R. Mohammed Amin. 2020. "Twitter Sentiment Analysis on Worldwide COVID-19 Outbreaks." *Kurdistan Journal of Applied Research* 54–65. doi: 10.24017/covid.8.

Onyeaka, Helen, Christian K. Anumudu, Zainab T. Al-Sharify, Esther Egele-Godswill, and Paul Mbaegbu. 2021. "COVID-19 Pandemic: A Review of the Global Lockdown and Its Far-Reaching Effects." *Science Progress* 104(2):00368504211019854. doi: 10.1177/00368504211019854.

Wang, Zijian, Scott Hale, David Ifeoluwa Adelani, Przemyslaw Grabowicz, Timo Hartman, Fabian Flöck, and David Jurgens. 2019. "Demographic Inference and Representative Population Estimates from Multilingual Social Media Data." Pp.

2056–67 in *The World Wide Web Conference*, *WWW '19*. New York, NY, USA: Association for Computing Machinery.