THE COVID-19 PANDEMIC:

What Can We Learn from Social Media at Scale and in Real-time?

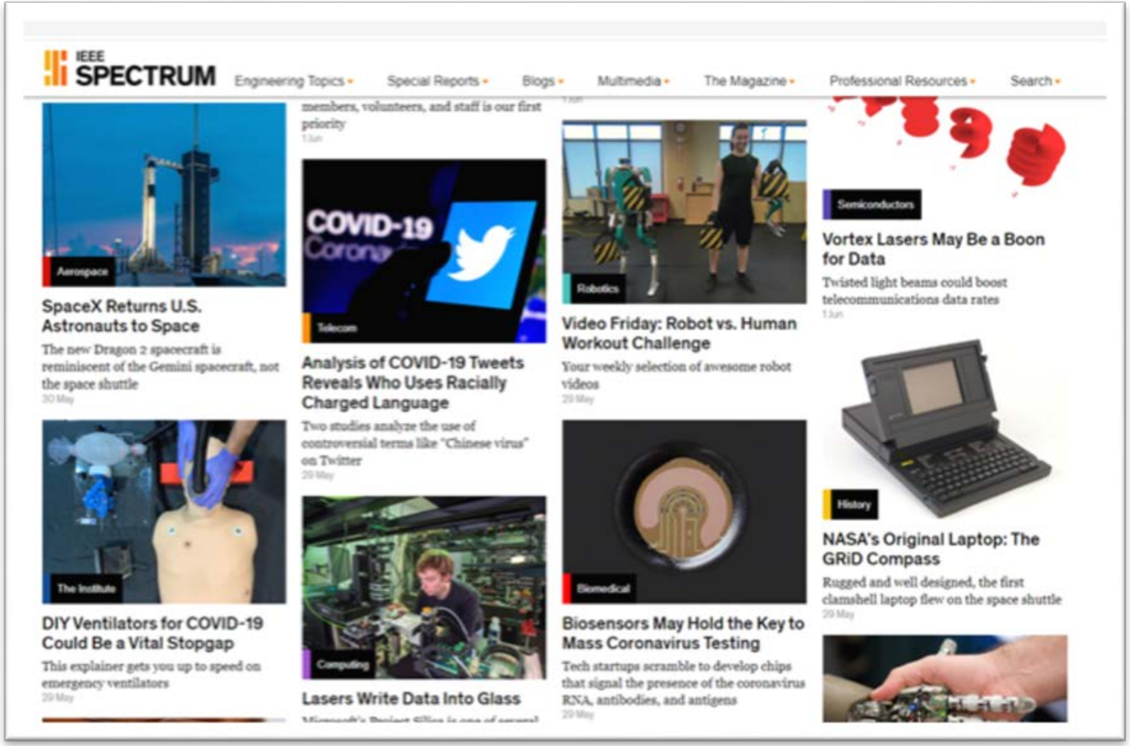
Jiebo Luo, Hanjia Lyu, Yu Wang, Long Chen, Yipeng Zhang,
Viet Duong, Xupin Zhang, Yubao Liu, Xiyang Zhang, Tongyu Yang

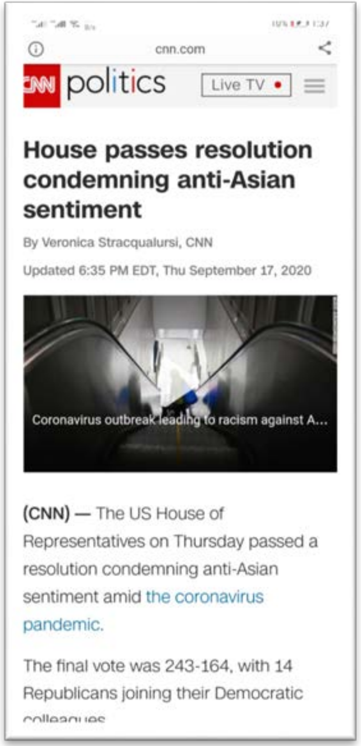
September 24, 2020

BACKGROUND

- The COVID-19 pandemic has severely affected people's daily lives and caused tremendous economic losses worldwide.
- Its influence on public opinions and people's mental health conditions has not received as much attention.
- The related literature in these fields has primarily relied on interviews or surveys, largely limited to small-scale observations.
- In contrast, the rise of social media provides an opportunity to study many aspects of a pandemic at scale and in real-time. Meanwhile, the recent advances in machine learning and data mining allow us to perform automated data processing and analysis.

PRESS COVERAGE





OUR WORK

- Characterizing Twitter users and topics regarding the use of controversial terms for COVID-19.

- Understanding how college students respond differently than the general public to the pandemic.

- Monitoring depression trends throughout COVID-19.

- Studying consumer hoarding behaviors during the pandemic.

01

Characterizing Twitter users and topics regarding the use of controversial terms for COVID-19

Hanjia Lyu, Long Chen, Tongyu Yang, Yu Wang, Jiebo Luo

CHARACTERIZING TWITTER USERS AND TOPICS REGARDING THE USE OF CONTROVERSIAL TERMS FOR COVID-19

- With the world-wide development of 2019 novel coronavirus, although WHO has officially announced the disease as COVID-19, one controversial term - "Chinese Virus" is still being used by a great number of people. When they refer to COVID-19, there are mainly two ways: using controversial terms like "Chinese Virus" or "Wuhan Virus", or using non-controversial terms like "Coronavirus"
- We find significant differences between these two groups of Twitter users across their demographics, user-level features like the number of followers, political following status, as well as their geo-locations.
- Tweets using controversial terms contain a higher percentage of anger as well as negative emotions. They also point to China more frequently.

CHARACTERIZING TWITTER USERS AND TOPICS REGARDING THE USE OF CONTROVERSIAL TERMS FOR COVID-19

- Young people tend to use non-controversial terms to refer to COVID-19.
- Male users constitute a higher proportion, but the proportion of female users in the ND group is higher than that in the CD group.

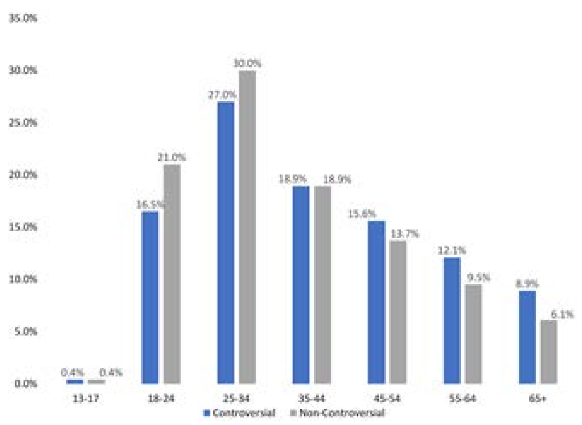


Fig. Age Distribution among users of controversial terms and users of non-controversial terms

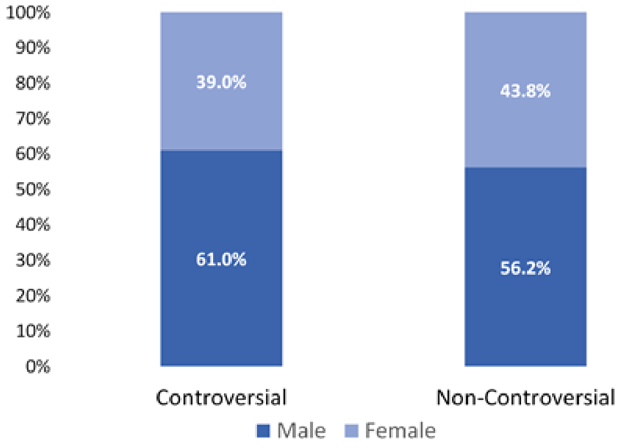


Fig. Gender Distribution among users of controversial terms and users of non-controversial terms

CHARACTERIZING TWITTER USERS AND TOPICS REGARDING THE USE OF CONTROVERSIAL TERMS FOR COVID-19

- Users in the ND group have been using Twitter for a longer time, and have a larger social capital which means they have more followers, friends, and statuses.
- The proportion of verified users in the ND group (2.0%) is higher than that of the CD group (0.6%).

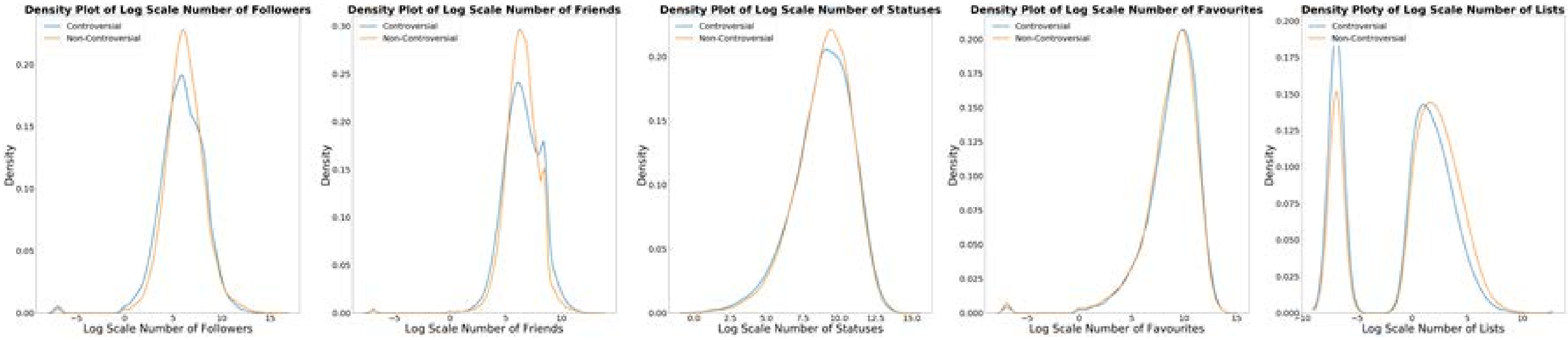


Fig. Density plots (log-scale) of the normalized numbers of followers, friends, statuses, favourites, and listed memberships

CHARACTERIZING TWITTER USERS AND TOPICS REGARDING THE USE OF CONTROVERSIAL TERMS FOR COVID-19

- There are more users following Donald Trump in the CD group than in the ND groups.
- The proportion of users in the ND group following the members of the Democratic Party is higher.

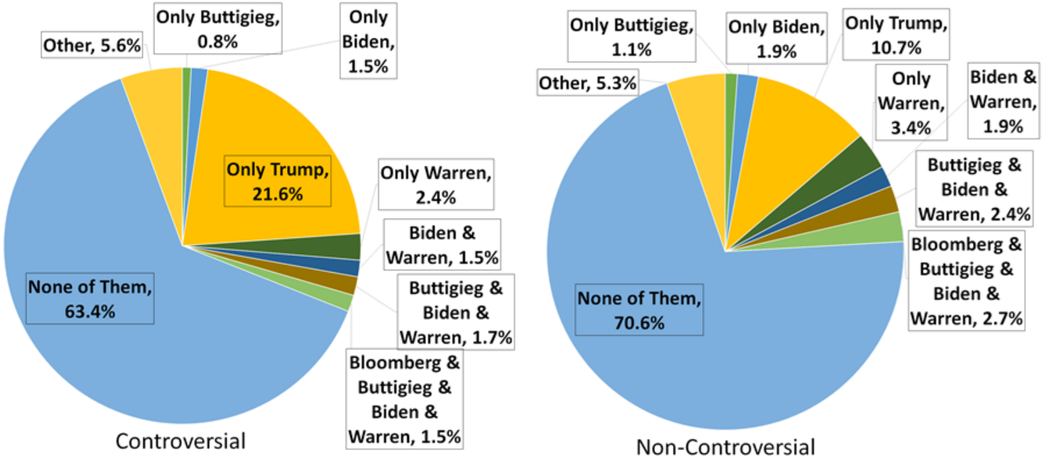
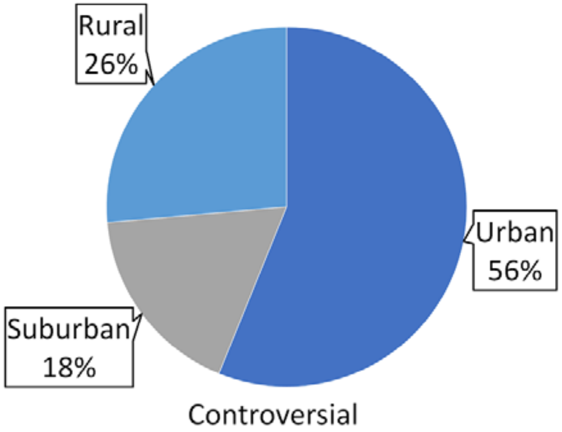


Fig. Proportion of Political Following Status

CHARACTERIZING TWITTER USERS AND TOPICS REGARDING THE USE OF CONTROVERSIAL TERMS FOR COVID-19

- Users living in rural or suburban areas are more likely to use the controversial terms than users living in urban areas



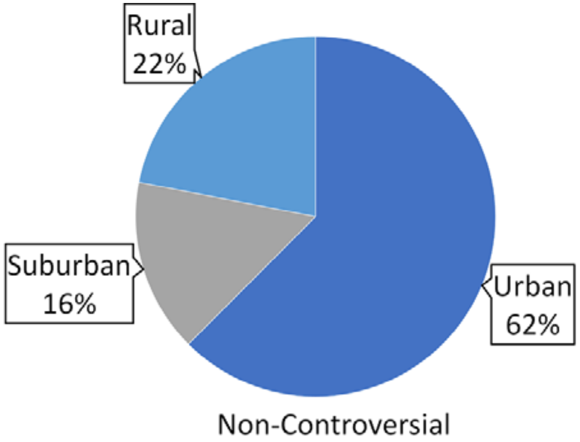


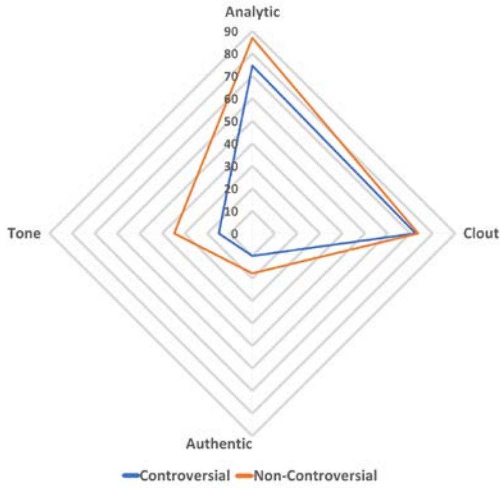
Fig. Tweet percentages in urban, suburban and rural areas

CHARACTERIZING TWITTER USERS AND TOPICS REGARDING THE USE OF CONTROVERSIAL TERMS FOR COVID-19

- Topics in the controversial posts are more related to China, even after the keywords related to “Chinese virus” were removed before the analysis.
- Discussions in non-controversial posts are more related to fighting the pandemic in the US.

CHARACTERIZING TWITTER USERS AND TOPICS REGARDING THE USE OF CONTROVERSIAL TERMS FOR COVID-19

- There are also differences across the sentiment of the tweets posted by the users using controversial terms and the users using non-controversial terms.



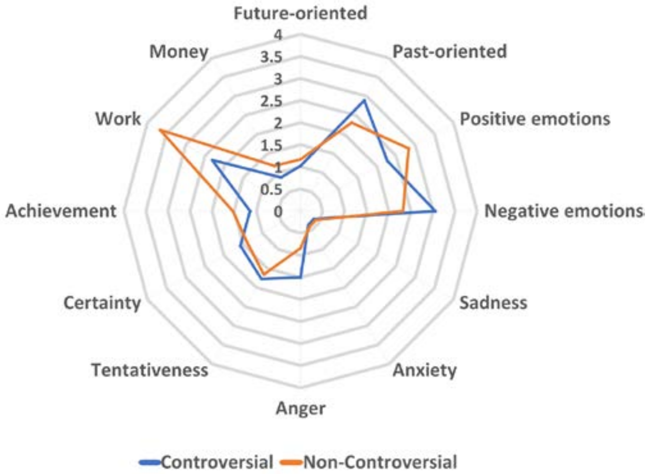


Fig. Linguistic profiles for the tweets of CD/ND

02

Understanding how college students respond differently than the general public to the pandemic

Viet Duong, Phu Pham, Tongyu Yang, Yu Wang, Jiebo Luo

UNDERSTANDING HOW COLLEGE STUDENTS RESPOND DIFFERENTLY THAN THE GENERAL PUBLIC TO THE PANDEMIC

- Following the closure of the University of Washington on March 7th, more than a thousand colleges and universities in the United States have cancelled in-person classes and campus activities, impacting millions of students.
- This paper aims to discover the social implications of this unprecedented disruption in our interactive society regarding both the general public and higher education populations by mining people's opinions on social media.
- We discover several topics embedded in a large number of COVID-19 tweets that represent the most central issues related to the pandemic, which are of great concerns for both college students and the general public.
- We find significant differences between these two groups of Twitter users with respect to the sentiments they expressed towards the COVID-19 issues.

UNDERSTANDING HOW COLLEGE STUDENTS RESPOND DIFFERENTLY THAN THE GENERAL PUBLIC TO THE PANDEMIC

- College students tend to focus their discussions on topics closely surrounding their living environment, such as school closure and local news.

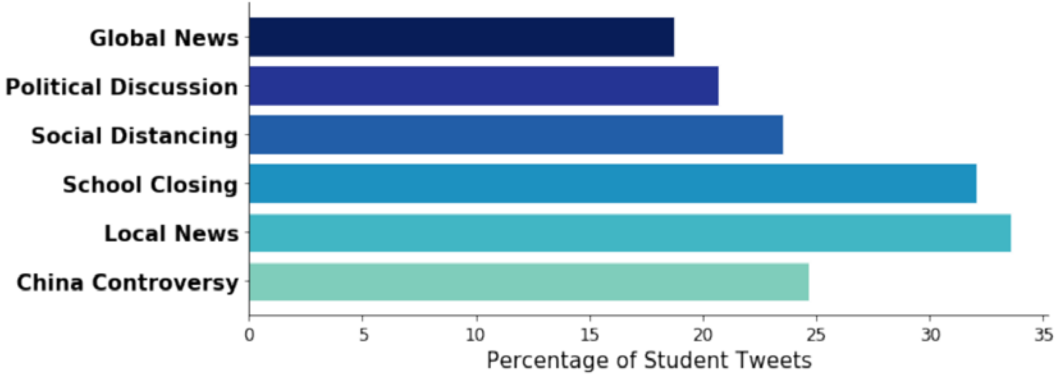


Fig. Student Tweets Contribution towards the Top 6 Topics

UNDERSTANDING HOW COLLEGE STUDENTS RESPOND DIFFERENTLY THAN THE GENERAL PUBLIC TO THE PANDEMIC

- Overall, a very small percentage of positive sentiments are expressed among the COVID-19 tweets. College students are shown to be significantly more negative.

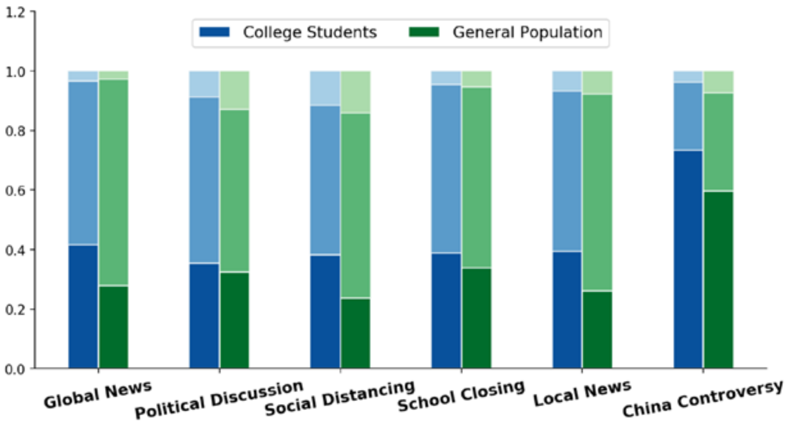


Fig. Sentiment Distributions (%) towards the 6 Most Frequent COVID-19 Topics. Percentage Blocks from Bottom to Top: Negative, Neutral, Positive

UNDERSTANDING HOW COLLEGE STUDENTS RESPOND DIFFERENTLY THAN THE GENERAL PUBLIC TO THE PANDEMIC

- Non-neutral tweets on the Social Distancing and School Closing topics express worrying emotions towards COVID-19.
- All the tweets revealing concerns on school closure are negative.
- Many students exhibited aggression to the foreign community, blaming them for the current disruptions in their lives as a result of social distancing.
- College students also disclosed details of their online learning experience, and mostly showed dislikes for remote learning (81.3%).

Subtopic Label	Subtopic Keywords	Negative Tweets
Showing aggression	asia, people, stay, home, worse_european, everyone, piss, fucking, work, fight	81.5%
Detailing precautions	worry, safe, world, tour, stay, people, cancel, take, wash_hand, precaution	65.1%
Expressing concerns	sick, know, go, work, see, watch, really, think, grocery_store, family	85.5%

Tab. Subtopics of Social Distancing

Subtopic Label	Subtopic Keywords	Negative Tweets
Detailing current situations	knock, follow_government_instruction, feel, survival_rate_whole, country_panicking, get, people, fuck, right, week, campus	98.5%
Detailing remote study	school, close, shut, student, find, get, campus, live_streaming_instead, email, anymore	81.3%
Expressing concerns	people, due, cancel, go, concern, tell, week, imagine, nasty, break	100%

Tab. Subtopics of School Closing

UNDERSTANDING HOW COLLEGE STUDENTS RESPOND DIFFERENTLY THAN THE GENERAL PUBLIC TO THE PANDEMIC

- It is encouraging that our college community remains aware and vocal on the racism problem related to the "Chinese virus" controversy, which sends a powerful message on the public's intolerance of racist behaviors on social media for the betterment of our society.

Subtopic Label	Subtopic Keywords	Negative Tweets
Calling out racism	chinese, president, call, racist, refer, flu, reply, people, fuck, trump	95.1%
Addressing attitudes	people, take, perspective, sick, seriously, asian, friend, know, ass, time	97.3%
Detailing public response	call, guy, get, keep, chinese, think, remember, wuhan, response, piss	91.7%

Tab. Subtopics of China Controversy

03

Monitoring depression trends throughout COVID-19

Yipeng Zhang, Hanjia Lyu*, Yubao Liu*, Xiyang Zhang, Yu Wang, Jiebo Luo

MONITORING DEPRESSION TRENDS THROUGHOUT COVID-19

- The influence of COVID-19 on people's mental health conditions has not received as much attention.
- To study this subject, we choose social media as our main data resource and create by far the largest English Twitter depression dataset containing 2,575 distinct identified depression users with their past tweets.
- We train three transformer-based depression classification models on the dataset, evaluate their performance with progressively increased training sizes, and compare the model's “tweet chunk”-level and user-level performances.
- Inspired by psychological studies, we create a fusion classifier that combines deep learning model scores with psychological text features and users' demographic information and investigate these features' relations to depression signals.
- We demonstrate our model's capability of monitoring both group-level and population-level depression trends by presenting two of its applications during the COVID-19 pandemic.

MONITORING DEPRESSION TRENDS THROUGHOUT COVID-19

- Previous studies have used **n-gram language models**, **topic models**, and deep learning models such as **1-dimensional convolutional neural networks (CNN)** and **bidirectional long short-term memory (BiLSTM)** to classify depression at the user level using Twitter data.
- All these works use small samples of **fewer than 500 users.**
- Shen et al. (2017) extended previous studies by expanding the dataset to contain **1,402** depression users and using a multimodal dictionary learning approach to learn the latent features of the data.
- In this study, we create a dataset of **5,150** Twitter users, including half identified depression users and half control users, along with their tweets within past three months and their Twitter activity data.
- We investigate the performance of some of these models, including BERT, RoBERTa, and XLNet on our dataset.

MONITORING DEPRESSION TRENDS THROUGHOUT COVID-19

- We progressively add data to our training set and notice a clear performance growth on all models, which validates the importance of our dataset.

Model	Train-Val Set	Accuracy	F1	AUC	Precision	Recall
Attention BiLSTM	1k users	70.7	69.0	76.5	70.9	67.3
	2k users	70.3	68.3	77.4	70.7	66.1
	4.65k users	72.7	71.6	79.3	72.1	71.1
CNN	1k users	71.8	72.6	77.4	72.7	72.6
	2k users	72.8	74.5	80.3	72.2	76.9
	4.65k users	74.0	70.9	81.0	77.4	68.9
BERT	1k users	72.7	74.4	79.8	72.0	76.9
	2k users	75.7	76.3	82.9	76.1	75.7
	4.65k users	76.5	77.5	83.9	76.3	78.8
RoBERTa	1k users	74.4	75.7	82.0	74.2	77.3
	2k users	75.9	77.9	83.2	73.8	**82.5**
	4.65k users	76.2	**78.0**	84.1	74.4	81.9
XLNet	1k users	73.7	75.1	80.7	73.2	77.2
	2k users	74.6	76.8	82.6	72.6	81.5
	4.65k users	**77.1**	77.9	**84.4**	**77.5**	78.3

Tab. Chunk-level performance (%) of all 5 different models using training-validation sets of different sizes

MONITORING DEPRESSION TRENDS THROUGHOUT COVID-19

- We build a more accurate classification model upon the deep learning models along with linguistic analysis of dimensions including personality, Linguistic Inquiry and Word Count (LIWC), sentiment features and demographic information.

Features	Accuracy	F1	AUC
VADER	54.9	61.7	54.6
Demographics	58.7	56.0	61.4
Engagement	58.7	62.3	61.7
Personality	64.8	67.8	72.4
LIWC	70.6	70.8	76.0
V+D+E+P+L	71.5	72.0	78.3
XLNet	78.1	77.9	84.9
All (Rand. Forest)	78.4	78.1	84.9
All (Log. Reg.)	78.3	78.5	**86.4**
All (SVM)	**78.9**	**79.2**	86.1

Tab. User-level performance (%) using different features. We use SVM for classifying individual features

MONITORING DEPRESSION TRENDS THROUGHOUT COVID-19

- The depression level trends are different between DP and ND groups.
- The distributions of the topics of DP and ND are different.

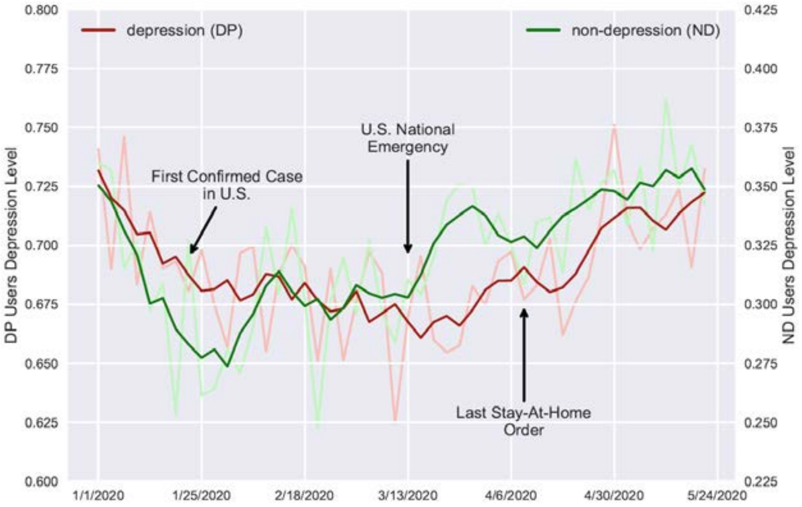


Fig. DP-ND trends

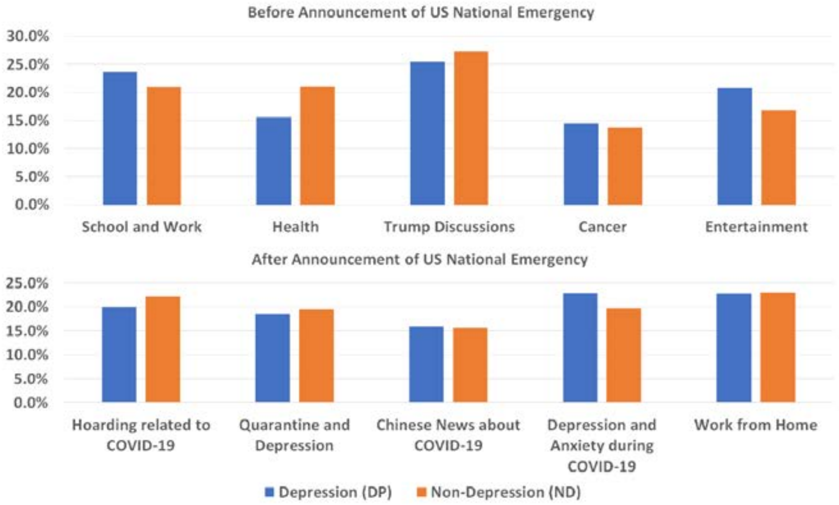


Fig. Percentage of DP-ND topics

MONITORING DEPRESSION TRENDS THROUGHOUT COVID-19

- The depression level trends are different.
- The distributions of the topics are different.

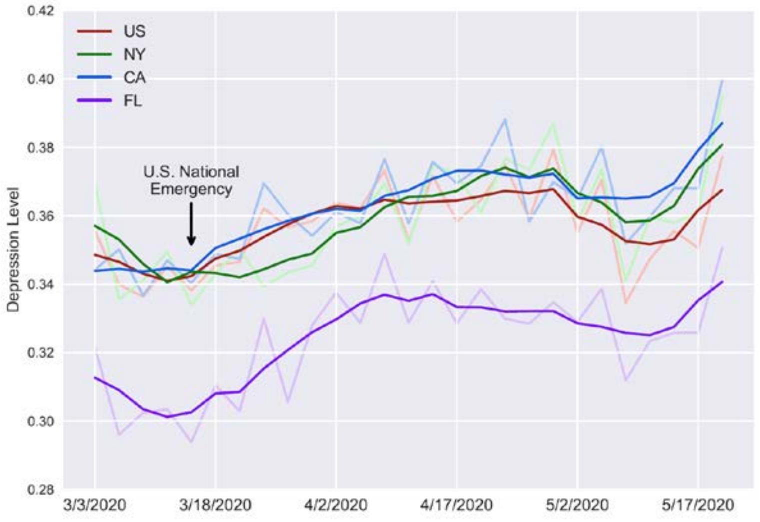


Fig. State-level trends

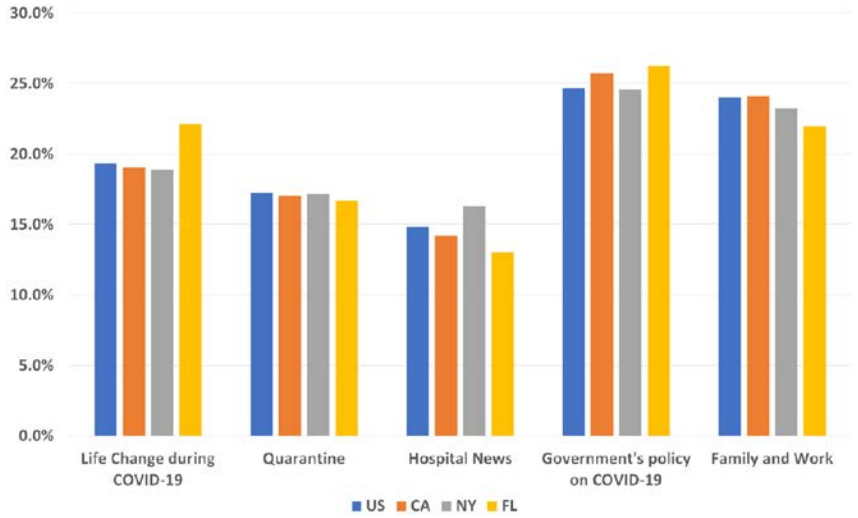


Fig. Percentage of State-level topics

04

Studying consumer hoarding behaviors during the pandemic

Xupin Zhang, Hanjia Lyu, Ravi Dhar, Jiebo Luo

STUDYING CONSUMER HOARDING BEHAVIORS DURING THE PANDEMIC

- During the COVID-19 pandemic, social media has acted as a double-edged sword: while it is a rich source for obtaining useful information concerning the pandemic, it also shapes the fears. For instance, when posts of panic-buying (e.g., toilet papers, hand sanitizers) proliferate on social media platforms, people might make panic purchases after they see such posts.

- We analyze the hoarding patterns of 43,102 Twitter users in the United States over the past three months, and particularly compared the hoarding related tweets across age, gender, family status, and geographic locations.

- We find significantly higher anxiety scores for the hoarding related tweets than the general tweet contents.

STUDYING CONSUMER HOARDING BEHAVIORS DURING THE PANDEMIC

- We apply a rule-based method to separate the tweets and their authors into two groups on the basis of whether the tweets indicate hoarding behaviors (Hoarding Group, HG) or express the idea of stopping hoarding (Not Hoarding Group, NHG).

- There are differences between the consumers of HG and NHG groups across age, gender, the population density of their locations, whether they live in the coastal states, but no significant differences over the family status - whether they have kids.

STUDYING CONSUMER HOARDING BEHAVIORS DURING THE PANDEMIC

- Younger adults (18-35) tend to post tweets that ask people to stop hoarding.
- The proportion of females of NHG is significantly higher than that of HG.
- The proportion of the users living the urban areas of HG is relatively larger.
- The consumers living the coastal states tend to post tweets that tell people to stop hoarding or panic buying.

STUDYING CONSUMER HOARDING BEHAVIORS DURING THE PANDEMIC

- We apply the LDA model to investigate the differences of the topics between HG and NHG groups.
- The tweets of the HG group focus on food, toilet papers and medical stuff, while the tweets of the NHG group focus on toilet papers, public health and shortage.

Classification	Topics	Top 10 Topic Words
HG	Toilet paper	panic paper toilet buy else everyone people know right go
	Food	panic buy people go food get hoard need grocery COVID
	Medical	people hoard need hand world keep medical sell leave sell
NHG	Public health	hoard stop leave wealth health public sick greed danger massive
	Toilet paper	stop panic paper toilet entire country bake people go hoard
	Shortage	buy stop need panic shortage stock essential hoard low people

Fig. Topics generated by the LDA model for both HG and NHG tweets

STUDYING CONSUMER HOARDING BEHAVIORS DURING THE PANDEMIC

- Food was always the major topic of HG during this time period. Even there were fluctuations, the overall proportion of the food topic was increasing. The proportions of the other 2 topics decreased as the pandemic developed. The discussions about the toilet was once heated around March 11.

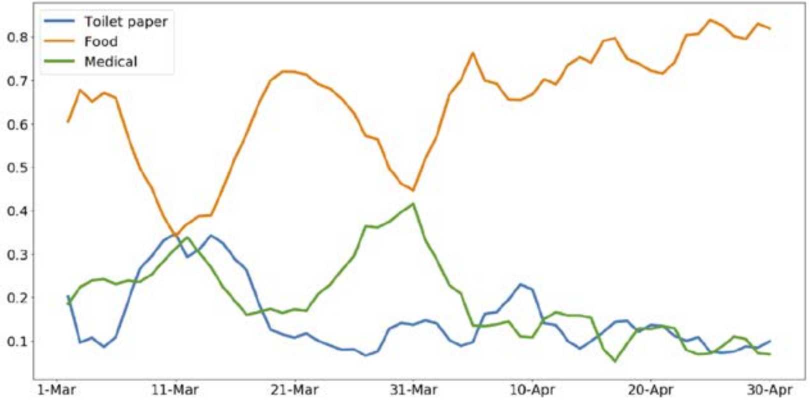


Fig. Topics generated by the LDA model for both HG and NHG tweets

STUDYING CONSUMER HOARDING BEHAVIORS DURING THE PANDEMIC

- The anxiety scores for the hoarding related tweets are calculated using LIWC.
- The LIWC anxiety mean for the hoarding related tweets is significantly higher than that for the general tweet contents.

Fig. Topics generated by the LDA model for both HG and NHG tweets

THANK YOU!