Early outbreak detection and proactive localized management using twitter data: case study of coronavirus pandemic

During a disease outbreak, timely non-medical interventions are critical in preventing the disease from growing into an epidemic and ultimately a pandemic. However, taking quick measures demands the capability to detect the early warning signs of the outbreak. This work collects Twitter posts surrounding the 2020 coronavirus pandemic, mentioning most common symptoms of COVID-19 including cough and fever geolocated to United States. Through examining the variation in Twitter activities at state level, we found a temporal window between the rises in online reported symptoms and officially reported cases which varies between X1 day X2 day.

*“Starting the new year off right with a cough and fever!”*

*“Starting the new year off right, sick as a dog with a high fever and a nasty cough. Craptastic.”*

*“Starting 2020 with a fever and flu like symptoms is not how I pictured this decade starting”*

*“my ribs hurt when I cough so I don’t want to cough but I have to cough I hate it here”*

# Introduction:

These are only a few examples of many twitter messages (known as *tweets*) that people have posted in early 2020 in the United States, complaining about intense flu-like symptoms such as dry cough and fever later on recognized as most common symptoms of COVID-19.

SARS-CoV-2, the virus that causes COVID-19, is thought to have first transmitted from an animal host to humans in Wuhan, China in late 2019. On March 11, after rapid increasing of the cases outside the China, World Health Organization eventually declared, the novel coronavirus as a pandemic [2]. As of April 25, it is officially reported that around three million people are infected by this virus in 210 countries and territories around the world and 2 international conveyances [3].

During a pandemic with high infection rate, prompt mitigatory actions play a crucial role in decelerating the spread and preventing new hotspot of the disease. Though, taking immediate actions requires the capability to detect the early warning signs of the outbreak and to characterize the dynamic of spread in a near real-time fashion.

In case of COVID-19 pandemic, delay in developing the test kits, limited number of kits, complicated bureaucratic health care systems and lack of transparency in data collection procedures are the major origins of postponement in effective preventive interventions and mitigatory responses [4, 5].

To fill this gap, Epidemic Intelligence (EI) is being used to explore alternative mostly informal sources of data to gather information regarding disease activity, early warning and infectious disease outbreak [6]. Human activities and interactions on the web are one of these informal sources. For instance, Google Flu Trends exploits web search queries to estimate flu activity [7].

Social media content is another powerful tool which provides invaluable crowd-sourced near real-time data for sensing health trends.

Twitter is a micro-blogging service with around 330 million monthly active users that let users communicate through short messages (tweets) [8]. Twitter permits third parties to explore tweets and collect data about posters and their locations. It provides the opportunity to harness tweets data to detect to early signs of outbreaks which can ultimately support decision makers in taking more informed actions [9].

In this paper we explore twitter data during COVID-19 pandemic across the United States at the state level, for the most common symptoms of COVID-19 including Cough and fever. The results are compared to the formal case dataset [John Hopkins git hub] to offer a framework for outbreak early detection.

The rest of this paper is organized as follows: Section 2 describes the related work that harness the twitter data to provide early warning. In Section 3, we present our data collection methodology for extracting relevant information from Twitter. Detailed data analysis are performed in Section 4 to establish correlation with John Hopkins data. In Section 5 we present …. Finally we conclude in Section 6.

# Related Work:

A number of studies have been conducted on twitter data to explore, model and predict disease outbreaks. Achrekar et al. 2011 presents an framework which monitors messages posted on Twitter with a mention of flu indicators to track and predict the emergence and spread of an influenza epidemic in a population [5]. Chen et al. 2016, propose an approach to aggregate users’ states in a geographical region for better estimation of trends [10]. Smith et al. 2016, offer a method to distinguish between personal infection tweets versus more general awareness tweets (being concerned about the epidemic) [11]. Masri et al. 2019, utilize tweet data in United States to develop a prediction model which predicts number of ZIKA cases one week in advance [12]. Similarly, Maurice et al. 2019, offer a method to improve Malaria surveillance system in Nigeria [13].

different forms of social networks like Del.icio.us, Facebook, Flickr, Linkedln, Wikipedia and Youtube etc. Sitaram et al. demonstrated how social media content like chatter from Twitter can be used to predict real-world outcomes of forecasting box-office revenues for movies [7]. Sakaki et al. used a probabilistic spatiotemporal model to build an autonomous earthquake reporting system in Japan using twitter users as sensors and applying Kalman filtering and particle filtering for location estimation [8]. Meme Tracking in news cycles as explained by Leskovec et al. was an attempt to model information diffusion in social media like blogs and tracking handoff from professional news media to social networks [9]. Twitter has been used for real-time notifications such as large-scale fire emergencies, downtime on services provided by content providers [10] and live traffic updates. There have been efforts in utilizing twitter data for predicting national mood [11], currency tracing and performing market and risk analysis. Tweetminster, a media utility tool design to make UK politics open and social, analyses political tweets, to establish the correlations between buzz on Twitter and election results. Ginsberg et al. in his paper discussing his approach for estimating Flu trends proposed that the relative frequency of certain search terms are good indicators of the percentage of physician visits and established a linear correlation to weekly published ILI percentages between 2003 and 2007 for all nine regions identified by CDC [6]. In June 2010, we introduced SNEFT architecture as a continuous data collection engine which combines the detection and prediction capability on social networks in discovering real world flu trends [12].

Conclusion:

## Reference:

[1] <https://www.nytimes.com/2020/01/06/world/asia/china-SARS-pneumonialike.html>

[2] <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen>

[3] <https://www.worldometers.info/coronavirus/>

[4] <https://www.washingtonpost.com/investigations/2020/04/18/timeline-coronavirus-testing/?arc404=true>

[5] Achrekar, Harshavardhan, et al. "Predicting flu trends using twitter data." *2011 IEEE conference on computer communications workshops (INFOCOM WKSHPS)*. IEEE, 2011.

[6] De Quincey, Ed, and Patty Kostkova. "Early warning and outbreak detection using social networking websites: The potential of twitter." International Conference on Electronic Healthcare. Springer, Berlin, Heidelberg, 2009.

[7] <https://www.google.org/flutrends/about/>

[8] <https://www.omnicoreagency.com/twitter-statistics/>

[9] Grover, Sangeeta, and Gagangeet Singh Aujla. "Prediction model for influenza epidemic based on Twitter data." *International Journal of Advanced Research in Computer and Communication Engineering* 3.7 (2014): 7541-7545.

[10] Chen, Liangzhe, et al. "Syndromic surveillance of Flu on Twitter using weakly supervised temporal topic models." *Data mining and knowledge discovery* 30.3 (2016): 681-710.

[11] Smith, Michael, et al. "Towards real-time measurement of public epidemic awareness: Monitoring influenza awareness through Twitter." *AAAI Spring Symposium on Observational Studies through Social Media and Other Human-Generated Content*. 2016

[12] Masri, Shahir, et al. "Use of Twitter data to improve Zika virus surveillance in the United States during the 2016 epidemic." *BMC public health* 19.1 (2019): 761.

[13] Maurice, Nduwayezu, et al. "Malaria Epidemic Prediction Model by Using Twitter Data and Precipitation Volume in Nigeria." 멀티미디어학회논문지 22.5 (2019): 588-600.