Risk Communication in Asian Countries: COVID-19 Discourse on Twitter

Sungkyu Park\textsuperscript{1,2}\textsuperscript{*}, Sungwon Han\textsuperscript{1,2}\textsuperscript{*}, Jeongwook Kim\textsuperscript{1,2}\textsuperscript{*}, Mir Majid Molaie\textsuperscript{1}, Hoang Dieu Vu\textsuperscript{3}, Karandeep Singh\textsuperscript{2}, Jiyoung Han\textsuperscript{1}, Wonjae Lee\textsuperscript{1}, Meeyoung Cha\textsuperscript{2,1}

\textsuperscript{1} Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea
\textsuperscript{2} Institute for Basic Science (IBS), Daejeon, South Korea
\textsuperscript{3} Hanoi University of Science and Technology (HUST), Hanoi, Vietnam

\section*{Abstract}

COVID-19 has become one of the most widely talked about topics on social media. This research characterizes risk communication patterns by analyzing the public discourse on the novel coronavirus from four Asian countries: South Korea, Iran, Vietnam, and India, which suffered the outbreak to different degrees. The temporal analysis shows that the official epidemic phases issued by governments do not match well with the online attention on COVID-19. This finding calls for a need to analyze the public discourse by new measures, such as topical dynamics. Here, we propose an automatic method to detect topical phase transitions and compare similarities in major topics across these countries over time. We examine the time lag difference between social media attention and confirmed patient counts. For dynamics, we find an inverse relationship between the tweet count and topical diversity.

\section{Introduction}

The novel coronavirus pandemic (COVID-19) has affected global health and the economy, and it has become a crucial topic on online platforms. Upon this crisis, many people participate in risk communication. Yet we do not know well about the virus, it can lead to much of the misinformation. What troubles is the fact-checking speed is not as quick as misinformation. In fact, when it comes to COVID-19, there had been damage due to false claims. The rhetoric has shifted from health preventive measures to the anti-vaccination movement.

In this research, we gather data from online to understand public discourse. Understanding the public concern will help determine which misinformation to debunk first, which also contributes to fighting the disease.

\textsuperscript{*}Equal contribution. Correspondence to: Meeyoung Cha <meeyoungcha@gmail.com>.

Meanwhile, given the current advanced information and communication technologies, people have interacted via social media and instant messengers and vastly share news, information, and thoughts associated with various topics (Lazer et al., 2018). The problem is that the speed of propagation is much faster than that of fact-checking (Kim et al., 2018). People also tend to share misinformation much faster and deeper than real information (Vosoughi et al., 2018; Kwon et al., 2013), and therefore, a vast amount of misinformation and/or a mixture of right and wrong confuses people to acknowledge what to follow and to take actions when they need it. To illustrate the current information crisis with an overflow of information, a new term named Infodemic (information + pandemic) has been newly introduced\textsuperscript{1}.

We aim to discern what people say in the wild. For instance, if we could identify a particular type of misinformation that is prevalent in only a handful of countries first, then we could inform people in other countries before the misinformation becomes a dominant topic and poses a crucial issue on public health of those countries. In this light, we have set up the following research questions.

\begin{itemize}
  \item Can official epidemic phases issued by governments reflect the online interaction patterns?
  \item How to automatically divide topical phases based on a bottom-up approach?
  \item What are the major topics corresponding to each topical phase?
  \item What are the unique traits of the topical trends by country, and are there any notable online communicative characteristics that can be shared among those countries?
\end{itemize}

\textsuperscript{1}Coronavirus Disease 2019 (COVID-19) Situation Report. https://bit.ly/2SKC18X.
2 Related Research

Issue Attention Cycle. A theoretical model named Issue Attention Cycle could provide a pertinent theoretical framework for our analyses (Downs, 1972). The model conceptualizes how an issue rises into and fades away from the center of public attention. In the first stage, labeled the pre-problem stage, an undesirable social condition (e.g., the appearance of COVID-19) emerges but has not yet captured much public attention while some experts or related groups of people may be already warned. The second stage, alarmed discovery and euphoric enthusiasm, occurs when a triggering event (e.g., the national spike of newly confirmed cases of COVID-19 or WHO’s statement on COVID-19) heightens public awareness of the issue. In the third stage, realizing the cost of significant progress, people begin to recognize the hardship that requires a major restructuring of society and significant sacrifices of some groups in the population to solve the problem. This inevitably causes a gradual decline of intense public interest, the fourth stage. In the final, the post-problem stage, the current issue is replaced by a new one and moves into a twilight zone of lesser public attention.

Not all issues follow the five stages of the issue-attention cycle (Nisbet and Huge, 2006). As the cyclical patterns of public attention evolve, a wide array of public discourse has been found across multiple issues of climate change (McComas and Shanahan, 1999), emerging technologies (Anderson et al., 2012; Wang and Guo, 2018), and public health risks (Shih et al., 2008; Arendt and Scherr, 2019). There are cultural differences as well (Jung Oh et al., 2012). Despite these fragmented findings to date, issue attention cycle analyses provide valuable insights into how public attention dramatically waxes and wanes. In particular, any issue that has gone through the cycle is different from issues that have not with two respects at least. First, during the time that the issue earned national prominence, new institutions, programs, and measures would have been developed to deal with the situation. These entities are likely to persist even after public attention has shifted elsewhere, thus having persistent societal impacts afterward. Second, the prolonged impacts of these entities are subject to what was heavily discussed when the issue was the primary public concern.

With this regard, it is important for the scholarship to look at the specifics of public conversations about a target issue. Although issue-attention cycle was originally conceived in relation with traditional media, including newspapers and televisions, there is burgeoning literature applying the model to social media platforms, most notably Twitter, where the public is increasingly turning to for information seeking and sharing without the gate-keeping process (David et al., 2016). Twitter conversations as such are more resonating with real-world word-of-mouth. It is not uncommon for journalists to reference social media in their news stories. Research also consistently finds that Twitter takes the initiative and greater control over public discourse especially in the early stages of the issue-attention cycle (Jang et al., 2017; Wang and Guo, 2018). Building on these prior studies, we first analyze the volume of Twitter conversations about COVID-19 to demonstrate an issue-attention cycle on a social media platform. We then examine how the content of Twitter during COVID-19 outbreak evolves as the cycle progresses.

COVID-19-related Analyses. As the current pandemic have had a huge impact on every aspect towards humanity, many researchers from the computer science and communication fields have also initiated various related research. Some researchers have focused on predicting the transmissibility of the virus. One work estimated the viral reproduction number ($R_0$) of the virus and showed $R_0$ of SARS-CoV-2 seems already larger than that of SARS-CoV, which was the cause of the SARS outbreak firstly found in the Guangdong province of southern China in 2002 (Liu et al., 2020). Another work claims that by reducing 90% of travel world wide, the spread of epidemic could be significantly reduced, by constructing a stochastic mathematical prediction model of the infection dynamics (Chinazzi et al., 2020).

Other lines of works are about understanding propagation of misinformation related to COVID-19. Particularly, one study modeled the spread of misinformation about COVID-19 as an epidemic model on various social media platforms like Twitter, Instagram, YouTube, Reddit and Gab; it also showed that users interact each others differently and consume information differently depending on the platforms (Cinelli et al., 2020). In this light, many media platforms like Facebook, YouTube, and Twitter said that they try to bring people back to a reliable source of medical information, and to do so they have direct communication lines with
CDC in the U.S. and WHO (Frenkel et al., 2020). When narrowing down to local-specific matters, one article claims that the fake news online in Japan has led to xenophobia towards patients and Chinese visitors, based on the qualitative analysis upon Japanese online news articles (Shimizu, 2020). Meanwhile, a work conducted a survey with 300,000 online panel members in 2015, South Korea while the MERS outbreak was prevalent in this country, and claimed that if the information from public health officials is untrustworthy, people rely more on online news outlets and communicate more via social media (Jang and Baek, 2019). Vice versa, one consulting firm argues that the general public could not hear well the voice of public health officials due to the prevalence of misinformation, including fake news about cures, conspiracy theories, and misleading information on the spread of the virus (Analytica); they argue that the efficacy of the response to restrain this ‘infodemic’ varies from country to country and depends on public confidence in the authorities. There is also an attempt to compare three countries in terms of political bias. The authors conducted a large-scale survey cross the U.S., the U.K., and Canada and statistically found that although political polarization of COVID-19 exists in the U.S. and Canada, the exact belief in COVID-19 is broadly related to the quality of an individual’s reasoning skills, regardless of political ideology (Pennycook et al., 2020).

In order to support pursuing the aforementioned research, many types of datasets are also released to the public as well as the research communities. One research crawled and opened tweet information from the total 10 languages with the COVID-19-related keywords for around three months (Chen et al., 2020). Another work collated over 59k academic articles, including over 47k full research papers about COVID-19, SARS-CoV-2, and the related Coronavirus issues (Wang et al., 2020).

### Table 1: Statistics of the crawled tweets.

| Language | Duration | Used Keyword\(^1\) | # of Tweets |
|----------|----------|--------------------|-------------|
| Korean   | Jan 1 to Mar 27, 2020 | Corona, Wuhan pneumonia | 1,447,489 |
| Farsi    | Jan 1 to Mar 30, 2020 | #Corona, #Coronavirus, #Wuhan, #pneumonia corona, n-cov, covid, acute pneumonia | 459,610 |
| Vietnamese | Jan 1 to Mar 31, 2020 | Corona | 87,763 |
| Hindi    | Jan 1 to Mar 31, 2020 | Wuhan pneumonia | 1,373,333 |

\(^1\) Keywords are listed here after translated in English from the actual local languages, e.g., “코로나” → “Corona” in Korean.

APIs\(^2\). We have focused on South Korea, Iran, Vietnam, and India in this research. These are all Asian countries, thereby we may control covariates like major differences between Western and Asian cultures. Meanwhile, the four countries all place unique characteristics in terms of dealing with the current outbreak. In Iran, the number of confirmed cases has gradually increased since the first confirmed case, whereas in Vietnam, their number has steadily stayed at a low level. In South Korea, there was a sudden drastic increase of the number after the first confirmed case, but it seems they have successfully flattened the curve, unlike other Western countries. In India, the situation had been relatively mild until mid March, and since then, there has been a drastic surge. We suspect that due to the different offline circumstances, the topics in online social media may vary across countries in addition to the unique cultural backgrounds.

We have set up two keywords, “Corona” and “Wuhan pneumonia,” in general, to crawl tweets (see Table 1 to find exact keywords used for crawling tweets for each country) and collected tweets for the three-month period from January to March 2020. Particularly with Farsi, we have not used keywords but used hashtags, starting with “#”, mainly used among Iranians since otherwise unexpected Arabic tweets could be crawled together.

### 3 Method

#### 3.1 Data

We have crawled the Twitter dataset by using the existing Twint Python library\(^2\) and Twitter search (Chen et al., 2020). Another work collated over 59k academic articles, including over 47k full research papers about COVID-19, SARS-CoV-2, and the related Coronavirus issues (Wang et al., 2020).

2The crawled Twitter dataset and the detailed information about the language-specific tokenizers is explained at https://github.com/dscig/COVID19_tweetsTopic. An advanced twitter scraping tool is written in Python. The detailed information about the scraper is explained at https://github.com/twintproject/twint.

3Official search tweets API for developers. Full-archive endpoint option provides complete access to tweets from the first tweet in March 2006. See also https://developer.twitter.com/en/docs/tweets/search/.
phases as shown in Figure 1. We have repeated the process for the aforementioned four languages.

**Preprocessing Data.** We firstly need tokens, which can be defined as the smallest units that have meaning, in order to extract topics from the collected tweets. We have filtered out unnecessary textual information like stop words, special characters (non-letters), special commands, emojis, etc. We then utilize the existing Python tokenizer libraries corresponding to each specific language. Detailed information about the language-specific tokenizers is also explained at the provided web link.

**Decide Topical Phases.** Next, we also need to set up specific target phases divided by dates to extract topics. It may not be feasible to extract topics from the whole 3-month period since there would be multiple fluctuations and changes on the topics reflecting the real events such as a drastic increase of the COVID-19 confirmed cases. It may also be not acceptable to use the epidemic phases that each government announces because the offline epidemic phases seem not to capture the actual online topic trends, as will be explained at the forthcoming Basic Daily Trends section.

We, therefore, devise a bottom-up approach to detect dates that show the sign of sudden increases in the daily volume of the tweets. In particular, we have set up two learnable parameters of the first derivatives (hereafter `velocity`) and the second derivatives (hereafter `acceleration`) of the daily tweet volumes, as illustrated in the formulas below.

\[
\text{velocity} = \frac{\# \text{ of tweet}_t - \# \text{ of tweet}_{t-1}}{D_t - D_{t-1}} \\
\text{acceleration} = \frac{\text{velocity}_t - \text{velocity}_{t-1}}{D_t - D_{t-1}}
\]

We reckon the `velocity` and `acceleration` values at the date when the first confirmed case being announced are the ground truths (GT) for each country. The intuition of this approach is that the `velocity` and `acceleration` values are proxies to unique communication traits for each country in terms of a specific subject (i.e., COVID-19 in our case) and once they have been computed from the first confirmed date, they would be the same to the following period.

We have set up joint thresholds for `velocity` and `acceleration` in order to find dates that show while velocity is still smaller than the `velocity_{GT}`, acceleration just becomes larger than the `acceleration_{GT}:` 0 < `velocity < velocity_{GT}` & `acceleration > acceleration_{GT}`. In this light, we learn two parameters from the first confirmed date by country then detect other dates that can be considered as start of forthcoming topical phases. When learning parameters, for `velocity`, we round down the `velocity_{GT}` then plus 1, and for `acceleration`, we round down the `acceleration_{GT}`, which are similar to the loss minimization concept of the machine-learning approach (i.e., learning is finished by one step).

Moreover, in terms of canceling noise signals, we have adopted a low-pass filter with 0.2 as the low-frequency threshold and smoothed the data. As a result, we could detect the dates that can divide the collected tweets into certain topical phases. For instance, in the case of South Korea, we could detect three dates and derive four corresponding phases, as shown in Figure 2-bottom, based on the computed daily velocity and acceleration values as shown in Figure 2-top.

**Model Topics.** We have utilized the latent Dirichlet allocation (LDA) for the topic modeling task. The LDA is one of the well-known machine-learning methods to extract topics amid given tex-
tual documents (i.e., a collection of discrete data-points) – tweets in our case (Ostrowski, 2015). The LDA generates and maximizes the joint probability between the word distribution of topics and the topic distribution of documents (Blei et al., 2003). The number of topics for each phase is a hyperparameter. We have set the range of the number of topics is between 2 and 50, and calculate perplexity (PPL), probability of how many tokens can be placed at the next step (i.e., indicating the ambiguity of the possible next token). PPL is a well-known metric to optimize a language model with a training practice (Adiwardana et al., 2020). During the iteration, we have fixed the minimum required frequency of words among the entire tweets for each phase to be 20 and the epoch number for each topic to be 100, respectively. We then decide the optimum number of topics for each phase by choosing the minimum PPLs.

As a result, we have decided on the number of topical phases and the corresponding optimized number of topics for each phase, as presented in Table 2. For example, in the case of South Korea, after the number of phases was decided as four from the ‘Decide Topical Phases’ module, the optimized number of topics is computed for each phase as 2, 41, 15, and 43, respectively.

Label Topics. As the last step, we have labeled the major themes for the extracted topics. This is to allocate semantic meanings to each topic and to analyze the semantic trends. Before labeling, we have sorted all tweets with the estimated topic numbering by descending order (i.e., tweets with larger volumes in terms of the estimated topic numbering list first) and discarded tweets whose volumes are less than 25% percentile. This is to focus on major topics by excluding tweets whose topics are relatively minor.

We then extract the top 1,000 most retweeted tweets and the top 30 highest probable keywords for each topic. We provide these datasets to domain experts for each language and ask them to label themes for each topic based on the given datasets. If several topics could be labeled as the same theme for each phase, those topics were merged as one theme label, and the eventual number of merged theme labels is shown in the third row for each country in Table 2. In addition, if one topic has more than one themes then it is labeled to have multiple classes. The maximum number of multiple cases within a topic were weighed as 0.5 when plotting the daily trends of the number of tweets.

In regard to the local/global news themes, we have narrowed down the labels since people talk about different issues under the news category. In particular, we have sub-labeled them as _confirmed if tweets are about the confirmed/death cases, _hate if about the hate crimes towards certain races, _economy if about the economic situations/policies, _cheerup if about supporting each others, _education if about when to reopen schools, and none if about general information, respectively.

4 Result

4.1 Basic Daily Trends

We depict the daily trends by plotting the daily number of tweets, and the daily number of the COVID-19 confirmed cases simultaneously. Adding to the two trends, we include official epidemic phases announced by each government as vertical lines (see Figure 3). By seeing the tweet and confirmed case trends together, we could confirm that the tweet trends are somewhat associated with the confirmed case trends. However, the official epidemic phases do not explain the tweet trends well.

Figure 3: Daily trends on the Four countries: X-axis is dates and Y-axis is trends of # of tweets with log-scale.

South Korea. The first confirmed case was identified on January 20, 2020. From early January till January 20, the daily numbers of tweets were relatively small, whereas the number sharply increased on January 25, as depicted in Figure 4. January 25 was the date when the Korean government increased the travel warning level on Wuhan city and Hubei province to suggest to evacuate from there, and this sign may affect the communication on Twitter. On February 18, the number sharply increased that had not been shown before, and it may be due to the 31st confirmed case related to a cult religious group in Daegu city. After the 31st

COVID-19 pandemic in Korea. https://bit.ly/3fy48Zp.
confirmed case has been found, the quarantine authority tried rigorous testing focusing on Daegu, and the number of the confirmed cases was drastically increasing until mid-March. The tweet trends also follow the same pattern. However, the official epidemic phases announced by the government, divided by the vertical dash lines in the figure, seem lag from the increasing number of tweets, and therefore we could say that the epidemic phases may not explain well enough the online communication trends in Korea.

**Iran.** On February 19, two people tested positive for SARS-CoV-2 in the city of Qom. After this date, we see a significant surge in the number of tweets and it reaches a peak in a few days. On February 23, the government changed the alert from white to yellow. Although the number of confirmed cases keeps increasing, the number of tweets starts to decrease gradually with a little fluctuation as shown in Figure 5. Therefore, the trends of these two numbers show different patterns in contrast
to Korean tweets. In the meantime, the government gradually increased preventive measures, and a number of cities with the highest rate of infection were announced hot spots or red zones. Overall, they didn’t place the whole country under the red alert. However, the government announced new guidance and banned all trips on 25 March. The president, on 28 March, said that 20 percent of the country’s annual budget would be allocated to fight the virus, which might be implicitly a sign of the red alert.

Vietnam. On January 23, 2020, Vietnam officially confirmed the first two COVID-19 patients, who come from Wuhan, China. After that, the number of tweets increased sharply and reached to peak in early February as shown in Figure 6. Although a few new cases were detected, the number of tweets tended to decrease and remained stable. On the second half of February, there are no new cases, however, the number of tweets increased rapidly and create a new peak. This peak could not remain for a long time. This trend can be explained by two possible reasons. The first is that the pandemic has spread over the world. The second is that the last cases in Vietnam were treated successfully. After a long time with no new cases, Vietnam had constantly confirmed new cases in Hanoi and many other cities from March 6. The number of tweets of this phase increased again and remain stable at a relatively higher level than the initial phase.

India. The first case of COVID-19 was confirmed on January 30, 2020. The number of cases quickly rose to three on account of students returning from the city of Wuhan, China. Throughout February, no new cases were reported and the first weeks of March also saw a relatively low number of cases. The number of cases however picked up numbers from the fourth week of March, notable were the 14 confirmed cases of Italian tourists in the Rajasthan province. This eventually led to the government of India declaring a complete lockdown of the country. The daily number of tweets followed a similar trend as that of the number of cases as depicted in Figure 7. First confirmed cases around January 30, 2020, caused a sudden spike in the number of tweets, that subsided in February. First COVID-19 fatality on March 12 and some other COVID-19 local events led to an exponential increase in the number of tweets. The tweets peaked on March 22 when the government declared lock-down of areas with infected cases and started trending downwards after that. It is strange that the declaration of nationwide lock-down by the government on March 24 only caused a small spike in the number of tweets and trend continued downwards. However, March 31 saw a large spike in the number of tweets owing to confirmation of mass infections in a religious gathering. Overall, the tweet trends seem to be synonymous with the release of official information by the government (e.g., number of confirmed cases, fatalities on COVID-19.)
4.2 Extracted Topical Trends

We have summed the theme labels acquired from the ‘Label Topics’ module as a daily basis and analyzed the topic changes across time with the three types of plots for the three target countries below: The first plot shows the daily topical trends based on proportions; the second shows the trends based on the number of tweets; the third shows the trends based on the number of tweets that country names like the U.S where explicitly mentioned. Overall, as people talk more on the COVID-19 outbreak (i.e., the daily # of tweets increases), the topics people talk about become less diverse.

**South Korea.** We have derived a total of four topical phases and plotted daily topical proportions as well as daily topical frequencies (see Figure 8-top and -mid). At first, there was no related topic on Phase 0. Then from Phase 1 to Phase 3, the number of topic diverged as 8, 5, and 11. On Phase 1, people talk much on personal thoughts and opinions linked to the current outbreak, and also they cheered up each other. On Phase 2, as the crisis going up to its peak, people talked less on personal issues and mainly talked on political and celebrity issues. The political issues were about shutting down the borders of South Korea towards China and of other countries towards Korea. On Phase 3, as the daily number of tweets becomes smaller than Phase 2, people tended to talk on more diverse topics including local and global news. In particular, people worried about hate crimes happened towards Asians in Western countries. People might be interested in different subjects as they think the crisis seems to be off the peak.

We also see the daily trends talked about various countries by counting the tweets remarking on each country name either by their local languages or by English. Korea, China, and Japan were mostly mentioned, and we suspect it is mainly triggered by the geopolitical relationships. Meanwhile, the U.S. and Italy also steadily mentioned across the three-month period, and the media outlets broadcasting global news may affect this phenomenon.

**Iran.** Figure 9-top and -mid illustrate two topical phases, their proportions, and daily topical frequencies in Farsi tweets. Phase 0 includes global news about China as well as unconfirmed local news that reflects the fear of virus spread in the country. Political issues form a remarkable portion of tweets in this phase, as the country has been struggling with various internal and external conflicts in recent years, and also, there was a congressional election in Iran. In phase 1, a significant increase in the number of tweets takes place, where local news regarding the virus outbreak constitutes the majority. An intriguing finding is that informational tweets about preventive measurements overshadow global news, which can be explained by the sociology of disaster that when people in a less developed country are at risk they naturally tend to share more information. However, political tweets are still widespread because of the aforementioned reasons and public dissatisfaction about the government response to the epidemic. This fact is also highlighted in Figure 9-bottom that after Iran and China, the U.S. is the most mentioned name. One possible explanation is that the outbreak puts another strain on the frail relationship between Iran and the U.S.

**Vietnam.** There are six topical phases with Vietnam and they are visualized as in Figure 10-top
and -mid. Phase 0 totally related to global news because in this period, Vietnam did not have any cases. From phase 1 to phase 5, topics diverged separately but they focused on local news except phase 3. Phase 3 is the phase when no new cases in Vietnam were detected. We can see a common point of phase 0 and phase 3 is no new cases in Vietnam (local news) so tweets tended to talk more about global news. Especially, in phase 3, we can see the increase of personal topics that most did not have in other phases. It was because a conflict event that related to Korean visitors made a huge of personal tweets.

Next, we show the number of tweets that mentioned countries as in Figure 10-bottom. The most three countries mentioned are Vietnam, Korea, and China. Vietnam and China were mentioned frequently across phases because Vietnam is the local and China is the original place of the pandemic.

Besides, Korea was mentioned in a large number of tweets but they only concentrated on Phase 3. This is totally similar to topics changes due to the Korean visitor event in Vietnam.

India. We have established three topical phases for tweets in Hindi in India (Figure 11-top and -mid). In the starting phase, the tweets are focused on sharing information about COVID-19, and global news about COVID-19 in China. People want to share the news about COVID-19 and information on how to be safe. Thereafter in Phase 1, the number of topics become more diverse. Although a large portion of the topics is concerned with information about the virus and global news, especially China, a major portion is formed by rumors or misinformation. The number of tweets spike on January 30, 2020, when the first case was confirmed in India. Towards the end of Phase 1, there is a further spike in the number of tweets, primarily due to the
beginning of announcements of some measures by the government to contain the virus (such as halting issuing new Visas to India). Lastly, in phase 3, a huge spike in the number of tweets is witnessed. The proportion of informational tweets decreases, whereas local news tweets confirming new cases increase. Regrettably, a marked portion of the tweets still consists of hateful content and misinformation. Interestingly enough, although the situation continued to worsen, tweets with people expressing dissatisfaction with the government are negligible. Phase 3 also witnesses mentions of other countries, especially Brazil and Europe, in addition to China and understandably, India. This could be attributed to a growing number of cases in Italy and Spain, Brazil, as well as the news surrounding the use of Hydroxychloroquine in Brazil. The U.S. also finds considerable mention due to the same reasons.

5 Discussion

We have analyzed tweets in order to understand what people are actually talking about related to the COVID-19 pandemic. In South Korea, the daily numbers of tweets tend to reach their peak as they are synchronized with sudden offline events. However, in case of Iran and Vietnam, the daily numbers of tweets tend to be not well synced with the offline events as in Iran, the government strongly control the online and offline media outlets and in Vietnam, people do not use Twitter much so the tweet trends may not resonate the actual flow of the public opinions. In all countries, we conclude that the epidemic phases or the national disaster stage announced by the governments did not well match the actual public opinion flows on social media, and therefore, we explore the topical phases which resonate with the flow of the public opinions with a bottom-up approach.

After extracting the topical phases, which those numbers were 4 in South Korea, 2 in Iran, and 6 in Vietnam, respectively, we have used the LDA and found the optimum number of topics for each topical phase and then labeled the corresponding themes for each derived topic. In general, as people talk more about COVID-19, the topics they refer to tend to be concentrated in a small number. This observation could become clearer if we consider the tweet depth value by phase as presented in Table 2. Tweet depth is defined as the number of retweets per day divided by the number of tweets per day. It can be deemed as a standardized cascading depth, and therefore, the larger value means the greater extent the depth for one tweet. From the case from South Korea and Vietnam, we could verify the observation as tweet depth also tends to get larger when people communicate more on COVID-19. However, For the Iran and India case, the number of phases were too small to observe the general characteristics of the topical trends.

Moreover, once the daily tweet volume has its highest peak then the forthcoming trend tends to go down in every country as shown in Figure 4–7 as the trend related to COVID-19 may also follow the Issue Attention Cycle. In this light, we observe that for some countries, the peak of the daily tweet trend precedes the peak of the daily confirmed case up to a few weeks, whereas for other countries, the two peaks are close to each other. No countries showed that the peak of the daily tweet trend succeeds that of the daily confirmed case.
When comparing South Korea and Vietnam, there is an intriguing point to discuss. The topic on Phase 0 in Korea was not related to COVID-19 whereas that in Vietnam was about the global news with confirmed cases. It may be cautious to generalize due to the small tweet volumes in Phase 0 for both countries, but Vietnamese users may concern more on the global epidemic issue from the first place and this tendency may affect on successful defending against pandemic later on.

To be specific to each country, in case of South Korea, when the offline situation has become severe (Phase 2), the number of topics becomes smaller, which may mean people more focus on a handful of specific issues. There is a unique trait at the initiated period (phase 0) such that people cheered up each others to hustle up and be in solidarity at the difficult times. Whereas in case of Iran, the number of topics has been relatively steady across time and the mentioned major topics have been skewed as news and information. In case of Vietnam, as the similar to the case of South Korea, at Phase 4, where the tweet traffic is relatively than Phase 3, the number of topics becomes larger and the themes of topics becomes less direct to the confirmed and/or death tolls, e.g., people talked on economics at Phase 2 and 4. Meanwhile, the Indian case indicates some unique characteristics such that many topics were related to misinformation, which was not shown much on other countries.

6 Concluding Remark

There are several limitations to be considered. First, we analyzed tweets solely from the four countries, and therefore, we need to be cautious about addressing explanations and insights that can be applied in general. We plan to extend the current study by including more countries. Second, there could be other ways to decide the topical phases. However, our approach can be aligned with the Issue Attention Cycle as we compute unique communication traits (i.e., velocity and acceleration by country) that would be relatively constant by issue, which is the COVID-19 outbreak issue in our case.

Despite the existing limitations, the current research could provide an important implication to fight against Infodemic. We find several topics that were uniquely manifested in the recent pandemic crisis by country. For instance, we could discover the emergence of misinformation on Hindi tweets. Our findings shed light on understanding public concerns and misconceptions under the crisis and therefore can be helpful in determining which misinformation to be discredited. This attempt may help eventually defeat the disease.

References

Daniel Adiwardana, Minh-Thang Luong, David R So, Jamie Hall, Noah Fiedel, Romal Thoppilan, Zi Yang, Apoorv Kulshreshtha, Gaurav Nemade, Yifeng Lu, et al. 2020. Towards a human-like open-domain chatbot. arXiv preprint arXiv:2001.09977.

Oxford Analytica. Misinformation will undermine coronavirus responses. Emerald Expert Briefings, (oxan-db).

Ashley A Anderson, Dominique Brossard, and Dietram A Scheufele. 2012. News coverage of controversial emerging technologies: Evidence for the issue attention cycle in print and online media. Politics and the Life Sciences, 31(1-2):87–96.

Florian Arendt and Sebastian Scherr. 2019. Investigating an issue—attention—action cycle: A case study on the chronology of media attention, public attention, and actual vaccination behavior during the 2019 measles outbreak in Austria. Journal of health communication, 24(7-8):654–662.

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. Journal of machine Learning research, 3(Jan):993–1022.

Emily Chen, Kristina Lerman, and Emilio Ferrara. 2020. Covid-19: The first public coronavirus twitter dataset. arXiv preprint arXiv:2003.07372.

Matteo Chinazzi, Jessica T Davis, Marco Ajelli, Corrado Gioannini, Maria Litvinova, Stefano Merler, Ana Pastore y Piontti, Kumpeng Mu, Luca Rossi, Kaiyuan Sun, et al. 2020. The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak. Science, 368(6489):395–400.

Matteo Cinelli, Walter Quattrociocchi, Alessandro Galeazzi, Carlo Michele Valensise, Emanuele Brugnoli, Ana Lucia Schmidt, Paola Zola, Fabiana Zollo, and Antonio Scala. 2020. The covid-19 social media infodemic. arXiv preprint arXiv:2003.05004.

Clarissa C David, Jonathan Corpus Ong, and Erika Fille T Legara. 2016. Tweeting supertyphoon haiyan: Evolving functions of twitter during and after a disaster event. PloS one, 11(3).

Anthony Downs. 1972. Up and down with ecology: The issue attention cycle. The Public Interest, 28:38–51.

Sheera Frenkel, Davey Alba, and Raymond Zhong. 2020. Surge of virus misinformation stumps facebook and twitter. The New York Times.
Kyungeun Jang and Young Min Baek. 2019. When information from public health officials is untrustworthy: The use of online news, interpersonal networks, and social media during the mers outbreak in south korea. *Health communication*, 34(9):991–998.

S Mo Jang, Yong Jin Park, and Hoon Lee. 2017. Round-trip agenda setting: Tracking the intermedia process over time in the ice bucket challenge. *Journalism*, 18(10):1292–1308.

Hyun Jung Oh, Thomas Hove, Hye-Jin Paek, Byoungkwon Lee, Hyegyu Lee, and Sun Kyu Song. 2012. Attention cycles and the h1n1 pandemic: A cross-national study of us and korean newspaper coverage. *Asian Journal of Communication*, 22(2):214–232.

Jooyeon Kim, Behzad Tabibian, Alice Oh, Bernhard Schölkopf, and Manuel Gomez-Rodriguez. 2018. Leveraging the crowd to detect and reduce the spread of fake news and misinformation. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pages 324–332.

Sejeong Kwon, Meeyoung Cha, Kyomin Jung, Wei Chen, and Yajun Wang. 2013. Prominent features of rumor propagation in online social media. In *2013 IEEE 13th International Conference on Data Mining*, pages 1103–1108. IEEE.

David MJ Lazer, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer, Miriam J Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, et al. 2018. The science of fake news. *Science*, 359(6380):1094–1096.

Ying Liu, Albert A Gayle, Annelies Wilder-Smith, and Joacim Rocklöv. 2020. The reproductive number of covid-19 is higher compared to sars coronavirus. *Journal of travel medicine*.

Katherine McComas and James Shanahan. 1999. Telling stories about global climate change: Measuring the impact of narratives on issue cycles. *Communication research*, 26(1):30–57.

Matthew C Nisbet and Mike Huge. 2006. Attention cycles and frames in the plant biotechnology debate: Managing power and participation through the press/policy connection. *Harvard International Journal of Press/Politics*, 11(2):3–40.

David Alfred Ostrowski. 2015. Using latent dirichlet allocation for topic modelling in twitter. In *Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015)*, pages 493–497. IEEE.

Gordon Pennycook, Jonathon McPhetres, Bence Bago, and David Rand. 2020. Predictors of attitudes and misperceptions about covid-19 in canada, the uk, and the usa.

Tsung-Jen Shih, Rosalynda Wijaya, and Dominique Brossard. 2008. Media coverage of public health epidemics: Linking framing and issue attention cycle toward an integrated theory of print news coverage of epidemics. *Mass Communication & Society*, 11(2):141–160.

Kazuki Shimizu. 2020. 2019-ncov, fake news, and racism. *The Lancet*, 395(10225):685–686.

Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science*, 359(6380):1146–1151.

Lucy Lu Wang, Kyle Lo, Yoganand Chandrasekhar, Russell Reas, Jiangjiang Yang, Darrin Eide, Kathryn Funk, Rodney Kinney, Ziyang Liu, William Merrill, et al. 2020. Cord-19: The covid-19 open research dataset. *arXiv preprint arXiv:2004.10706*.

Weirui Wang and Lei Guo. 2018. Framing genetically modified mosquitoes in the online news and twitter: Intermedia frame setting in the issue-attention cycle. *Public Understanding of Science*, 27(8):937–951.