The evolution of rumors on a closed platform during COVID-19

Andrea Wen-Yi Wang, Jo-Yu Lan, Ming-Hung Wang, Chihhao Yu

Submitted to: Journal of Medical Internet Research
on: May 21, 2021

Table of Contents

The evolution of rumors on a closed platform during COVID-19

Andrea Wen-Yi Wang[1] MSc; Jo-Yu Lan[2] BEng; Ming-Hung Wang[2] PhD; Chihhao Yu[1] MFA

[1]Information Operations Research Group Taipei, Taiwan TW
[2]Department of Information Engineering and Computer Science Feng Chia University Taichung TW

Corresponding Author:
Chihhao Yu MFA
Information Operations Research Group
7F-13, No. 103, Sec. 1, Fuxing S. Rd., Da'an Dist.
Taipei, Taiwan
TW

Abstract

Background: In 2020, the COVID-19 pandemic put the world in crisis on both physical and psychological health. Simultaneously, a myriad of unverified information flowed on social media and online outlets. The situation was so severe that the World Health Organization identified it an infodemic on February 2020.

Objective: We want to study the propagation patterns and textual transformation of COVID-19 related rumors on a closed-platform.

Methods: We obtained a dataset of 114 thousand suspicious text messages collected on Taiwan's most popular instant messaging platform, LINE. We also proposed an algorithm that efficiently cluster text messages into groups, where each group contains text messages within limited difference in content. Each group then represents a rumor and elements in each group is a message about the rumor.

Results: 114 thousand messages were separated into 937 groups with at least 10 elements. Of the 936 rumors, 44.5% (417) were related to COVID-19. By studying 3 popular false COVID-19 rumors, we identified that key authoritative figures, mostly medical personnel, were often quoted in the messages. Also, rumors resurfaced multiple times after being fact-checked, and the resurfacing pattern were influenced by major societal events and successful content alterations, such as changing whom to quote in a message.

Conclusions: To fight infodemic, it is crucial that we first understand why and how a rumor becomes popular. While social media gives rise to unprecedented number of unverified rumors, it also provides a unique opportunity for us to study rumor propagations and the interactions with society. Therefore, we must put more effort in the areas.

(JMIR Preprints 21/05/2021:30467)
DOI: https://doi.org/10.2196/preprints.30467

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Original Paper

The evolution of rumors on a closed platform during COVID-19

Abstract

Background: In 2020, the COVID-19 pandemic put the world in crisis on both physical and psychological health. Simultaneously, a myriad of unverified information flowed on social media and online outlets. The situation was so severe that the World Health Organization identified it as an *infodemic* on February 2020.
Objective: We want to study the propagation patterns and textual transformation of COVID-19 related rumors on a closed-platform.
Methods: We obtained a dataset of 114 thousand suspicious text messages collected on Taiwan's most popular instant messaging platform, LINE. We also proposed an algorithm that efficiently clusters text messages into groups, where each group contains text messages within a limited difference in content. Each group then represents a rumor and elements in each group is a message about the rumor.
Results: 114 thousand messages were separated into 937 groups with at least 10 elements. Of the 936 rumors, 44.5% (417) were related to COVID-19. By studying 3 popular false COVID-19 rumors, we identified that key authoritative figures, mostly medical personnel, were often quoted in the messages. Also, rumors resurfaced multiple times after being fact-checked, and the resurfacing pattern were influenced by major societal events and successful content alterations, such as changing whom to quote in a message.
Conclusions: To fight infodemic, it is crucial that we first understand why and how a rumor becomes popular. While social media gives rise to an unprecedented number of unverified rumors, it also provides a unique opportunity for us to study rumor propagations and the interactions with society. Therefore, we must put more effort in the areas.

Keywords: COVID-19; rumors; rumors diffusion; rumors propagations; social listening; infodemic; social media; closed platform; natural language processing; machine learning; unsupervised learning; computer and society

Introduction

Background

Online social media has democratized content. By creating a direct path from content producer to consumers, the power of production and sharing of information has been

redistributed from limited parties to general populations. However, social media platforms have also given rise to the proliferation of misinformation and enabled the fast dissemination of unverified rumors [1-3]. In 2020, the COVID-19 pandemic put the world in crisis on both physical and psychological health. Simultaneously, a myriad of unverified information flowed on social media and online outlets. The situation was so severe that the World Health Organization identified it as an *infodemic* on February 2020 [4]. According to studies, rumors and claims regarding erroneous health practices can have long-lasting effects on physical and psychological health, and it even interfered with the control of COVID-19 in various parts of the world [5, 6].

Prior Works

In light of the *infodemic*, several investigations have been carried out to look at the COVID-19 misinformation issue in various aspects. Several survey studies revealed that people relied on social media to gather COVID-19 information and guidelines [7, 8]. Misinformation on social media has since been a keen interest of the research community.

Efforts have been put into studies of true and false rumors on social media [9]. For example, Cinelli et al. [10] compared feedbacks to the reliable and questionable information across five platforms, including Twitter, YouTube, and Gab. The study showed that users on the less regulated platform, Gab, responded to questionable information 4 times more than those on the reliable ones. YouTube users were more attracted to reliable contents, and Twitter users reacted to both contents more equally. Gallotti et al. [11] looked at how much unreliable information Twitter users were exposed to across countries. While the level of exposure was country dependent, they revealed that the exposure to unreliable information decreased globally as the pandemic aggravated.

Machine learning and deep learning techniques have been used to detect rumors [12] and study the topics and sentiments for COVID-19 misinformation [13, 14]. For example, Jelodar et al.

[15] used Latent Dirichlet Allocation to extract topics from 560 thousands of COVID-19 Twitter posts and then used LSTM neural network to classify sentiments of posts. By applying Structure Topic Model and Walktrap Algorithm, Jo et al. [16] classified questions and answers from South Korea's largest online forum and discovered that questions related to COVID-19 symptoms and related government policies revealed the most fear and anxiety. Furthermore, by employing a multimodal deep neural network for demographic inference and VADER model for sentiment analysis, Zhang et al. [17] performed a cross sectional study on Twitter users. They found that older people exhibited more fear and depression toward COVID-19 than their younger counterparts, and females were generally less concerned about the pandemic.

Previous investigations on rumors indicated that individuals are more likely to believe in questionable statements after seeing them repeatedly [18, 19], and that rumors became more powerful after being shared multiple times [20]. These studies highlight the severe consequences of an *infodemic* during a pandemic period, when false information would involve many erroneous health practices that have direct consequences on people's well-beings. However, the majority of the studies focused on data collected from public social media platforms such as Twitter, Facebook, or Weibo. Explorations on closed messaging platforms, such as WhatsApp, WeChat, Telegram or LINE, remained extremely scarce. While popular social media platforms are indeed important targets to study online behaviors and expressions, closed platforms remain an integral place to look at, given its more private settings. Furthermore, most studies only look at the broad topics of misinformation. For example, some looked at reliable versus unreliable information [10, 11, 21], and others employed natural language processing techniques to reduce thousands of social media posts into 10 to 20 groups of topics [10, 13, 15, 16]. Studies that put focus on rumors themselves and investigated the diffusion pattern of specific rumors is rare in COVID-19 *infodemic* research. A similar one would be the study of the content change and temporal diffusion pattern of 17 popular

political rumors on twitter by Shih et al [22]. They found that false information came back repeatedly, usually becoming more extreme and intense in wordings, while true information did not resurface at all.

Contributions of this study

Our contribution to the current research is in three ways. First, by investigating COVID-19 messages on LINE, we added to the limited research of COVID-19 rumors on closed messaging platforms [23, 24]. We looked into a dataset of 114,124 suspicious messages reported by LINE users in Taiwan between January and July 2020.

Secondly, we proposed an efficient algorithm that could cluster a large number of text messages according to their text similarity without having to decide how many groups beforehand. The results were clusters where each one only contains messages that are within limited alterations among each other. Thus, each resulting cluster is one specific rumor.

Third, by using the results from the algorithm, we were able to look at the dynamics of each rumor over time. To the best of our knowledge, we are the first to study not only how the content of a specific COVID-19 rumor evolved over time but the interaction between content change and popularity. We found that some form of content alterations had been successful in aiding the spread of false information.

The major findings of this work are three-fold:

1. By combining Hierarchical Clustering and K-Nearest Neighbors, we could reduce computational time of clustering to linear time. This would enable the large-scale study of rumor transformation.

2. Fact-check did not effectively alleviate the spread of COVID-19-related false information. In fact, several peaks of false information messages were observed after repeated fact-check reports across different rumors.

3. Key authoritative figures are an integral part of false COVID-19 information. Most messages falsely quoted medical personnel and such practice help with the popularity of a message.

In the following sections, we used *clusters* and *groups* interchangeably. And we described a group of suspicious messages as one *rumor*, since belonging to the same group meaning they were seen as one narrative. And then we referred to rumors that are verified false as *misinformation* or *false information*.

Method

Data

LINE, similar to Telegram, is the most popular instant message platform in Taiwan. According to the survey by Taiwan Communication Survey in 2018, 98.5% of people in Taiwan used LINE as their primary messaging tool.

In Taiwan, LINE users can voluntarily forward suspicious messages to fact-checking LINE bots such as Cofacts or MyGoPen. These bots archive the messages and check against their existing databases to reply with the fact-checked results.

We obtained a dataset of 210, 221 suspicious messages forwarded by LINE users to a fact-checking LINE bot between January to July 2020. The dataset included rumors related to COVID-19 as well as other topics. To do clustering, we preprocessed each message by the following steps:

1. Removed non-Simplified or non-Traditional Chinese Characters.
2. Tokenized with Jieba [25].
3. Removed tokens that are Chinese stopwords.

In the following sections, we focused on longer texts. We only looked at 114, 124 messages having at least 20 tokens. The character distribution is presented in (Table 1) and number of messages reported per date is in (Figure 1).

Along with the text content of each reported message, we also obtained the report time of

each message and a unique identifier for the LINE user that reported the message. It is to note that the user identifiers we received were scrambled, therefore, it was not possible for us to use the identifiers to attribute any message back to any actual LINE user.

Table 1. Components of messages having at least 20 tokens. "Others" include characters such as punctuation marks and emojis.

	All	Chinese	Digits	Alphabets	Others
Min	24	24	0	0	0
Median	233	145	7	2	38
Max	10012	8132	3252	7014	5532

Problem Definition

A text message, during natural sharing among users, would undergo some degrees of content transformation, such as with addition of phrases like “Please share it with your friends and family”. Some may have some other kinds of transformation, but the main content itself stays roughly the same. What we would like to do is to identify what messages are in essence the same, ignoring the minor differences. In the following, we explained the technical definition and algorithm. To align with terms in Natural Language Processing, we would use a *document* to refer to a *message*.

Given a set of *n* documents, we would like to group them into *m* clusters, of which each cluster is made up of documents very similar in usage of terms, only within a limited degree of text alterations. Note that *m* is unknown beforehand.

For example, given two documents A and B, they should be in the same cluster if the overlapping terms of A and B constitute a large part of both A and B. However, if the overlapping terms make up a large part of A but not B, then they should be in different clusters, because that means B is made up of A plus a significant number of other terms.

Formally, we defined the terms in a document to be its token set after tokenization. And the distance between two documents A and B to be

$$d(A,B)=1-\frac{\left|tok(A)\sqcap tok(B)\right|}{max\left(\left|tok(A)\right|,\left|tok(B)\right|\right)} \quad (1)$$

where $tok(\cdot)$ is the set of tokens of one document. And $¿\cdot\vee¿$ denotes the number of elements in a set.

The Cluster-Classification, "Hybrid", Algorithm

We presented an algorithm that combines a popular unsupervised learning algorithm, Hierarchical Clustering, and a supervised learning algorithm, K-Nearest Neighbor, to achieve fast clustering of large numbers of documents. See (Figure 2) for algorithm flow chart.

Notation

1. $(A)_j$: jth element of set A.

2. $Label(x)$: The label of element x.

Input

1. D: the set of all documents to be grouped.

2. D^T : the set of tokenized documents. The order is preserved as D.

3. *train portion* ***p***: a number in (0, 1].

4. *distance threshold* **λ**: a number in (0, 1].

Algorithm

1. Select ***p*** × |D^T| elements from D^T, denoted as D_p^T, and the rest not selected as set D_q^T.

2. Construct distance matrix M for D_p^T, where $M_{ij}=d¿$ by formula (1). Note that M is symmetric.

3. Feed M into Hierarchical Clustering with a distance threshold of **λ**. We would get back a sequence of labels L_p, where $(L_p)_i$ is the label of element $(D¿¿p^T)_i¿$. Elements with the same label are in the same cluster. Since the label itself does not carry meaning, manipulate them so they are all non-negative whole numbers.

4. $\forall unique\, l \in L_p$, if $¿\{k \vee k=l \forall k \in L_p\} \vee ¿1$, then replace the value of l to −1. Denote the updated label set as $L_p^{'}$.

5. Train a K-Nearest Neighbors classifier ***K*** using the training set $(D_p^T, L_p^{'})$. Then use ***K*** to predict the labels of (D_q^T). Denote the prediction as L_q.

6. Construct L from $L_p^{'}$ and L_q, where $(L)_i=Label((D^T)_i)¿$.

7. Construct $D_O^T=\{d_i \vee Label(d_i)=-1 \forall d_i \in D^T\}$.

8. Redo step 2 and 3 for D_o^T. Denote the output as L_o. Make sure the values of L_o do not overlap with the values of L from step 6.

9. Update L from step 6 with L_o.

Algorithm Output

Output is L. The i[th] element of L, denoted as $(L)_i$, is the label of $(D^T)_i$. Note that the value of the label itself does not carry any meaning. However, elements in D^T with the same label belong to the same cluster.

Dataset ground truth

We randomly selected 50,000 messages from the dataset and used pure Hierarchical Clustering algorithm to perform clustering. The messages were separated into 7,401 groups. The largest group had 1,082 messages, and the smallest group contained only 1. There were 5, 231 groups with only 1 message, meaning the rest of 44, 796 messages were separated into 2170 groups. There were 12 groups with at least 500 messages. Refer to (Table 2) for further statistics.

Table 2. Statistics of group sizes (50, 000 documents).

Groups with at least X messages	Mean	Standard Deviation	Maximum	Q_3	Q_2
1	6.756	39.190	1082	2	1
2	20.631	70.478	1082	10	3

Performance comparisons with other common models

Evaluation Metrics

We opted precision, recall and F-score as evaluation metrics. In the sense of information retrieval, precision is the number of correct results returned divided by all results returned from search. Hence, high precision means the predictions are very relevant. On the other hand, *recall* measures the number of correct results returned divided by the total number of correct results. High recall corresponds to the completeness of returned results. Note that simply by returning all documents, one could achieve 100% of recall, but that will result in very low precision. Therefore, precision and recall need to be taken together to determine the quality of classification. F-score, defined as the harmonic mean of precision and recall, is one such measure that combines precision and recall.

Experiments Settings

We compared speed and performances among 4 models:

1. Hierarchical Clustering only (**clustering**). The result from this model is ground truth.

2. The Cluster-Classification Algorithm (**hybrid**). This is our proposed algorithm.

3. Latent Dirichlet Allocation (**LDA**).

4. KMeans with PCA dimensionality reduction (**pca+kmeans**).

Throughout the experiments we used the distance threshold $\boldsymbol{\lambda} = 0.6$. Both **LDA** and **pca+kmeans** require a predefined number of groups, a requirement which does not really fit our use case. However, for the sake of comparison, we would use the number of groups outputted by **clustering** model as input to both models.

Measuring model performances

Suppose the input is a tokenized set of k documents D^T and the **clustering** model put k documents into n groups, (g_1, g_2, ... g_n). g_1 is the group having the largest number of documents and

g_n the least. Another model M put D^T into m groups: (l_1, l_2, ..., l_m). We calculated precision, recall and F-score of model M by the algorithm in (Textbox 1):

In each experiment, we did 5 iterations. In each iteration, we randomly selected k messages from our dataset. We would get 1 precision and recall after each iteration, and we used the results of 5 iterations to calculate 95% confidence intervals.

Textbox 1. Algorithm for calculating Precision, Recall and F-score.

Initialization: $i \leftarrow 1, c \leftarrow 0, p \leftarrow 0, r \leftarrow 0, f \leftarrow 0;$

while c < $\frac{k}{2}$ **do**

Find l_k where l_k has the most overlapping components with g_i;
calculate precision p_k, recall r_k and F-score f_k of l_k by comparing with g_i;
$r \leftarrow r + r_k$;
$p \leftarrow p + p_k$;
$f \leftarrow f + f_k$;
$i \leftarrow i + 1$;
$c \leftarrow c + |g_i|$;

Result: precision ← p/i; recall ← r/i F-score ← f/i;

Experiments Results

As shown in (Figure 3), the **hybrid** model greatly reduced the time required especially when ***p*** was equal or less than 0.6. Furthermore, the performance metrics remained greater than 99% across levels of ***p*** (Figures 4, 5, 6). It showed that the **hybrid** model's assignments of groups were very complete (measured by recall), and that the classification of K-Nearest Neighbors did not introduce too many errors in each group (measured by precision). From (Table 3), we observed that **LDA** is much slower than other models. Furthermore, the precision was very low, meaning that predicted groups could have many false positives. On the other hand, **pca+kmeans** were 10 times slower than **clustering**. While the precision was comparable to that of **hybrid** methods, recall was only 73%. This showed that **pca+kmeans** would miss out many messages due to minor differences in content.

Table 3. Performance comparison for 10,000 documents.

Model	Mean Runtime (seconds)	Mean Precision	Mean Recall	Mean F-score
clustering	6.594	-	-	-
hybrid, $\boldsymbol{p}$ = 0.2	2.172	0.993	0.982	0.986
hybrid, $\boldsymbol{p}$ = 0.4	2.502	0.995	0.996	0.995
hybrid, $\boldsymbol{p}$ = 0.6	3.418	0.997	0.998	0.997
hybrid, $\boldsymbol{p}$ = 0.8	4.697	0.998	0.999	0.999
LDA	1788.981	0.624	0.939	0.704
pca+kmeans	41.143	0.993	0.734	0.823

Results

Overview

By using the **hybrid** algorithm with train portion $\boldsymbol{p}$ = 0.4 and distance threshold $\boldsymbol{\lambda}$ = 0.6, 114,124 messages were separated into 12,260 groups. Among those, 8,529 groups only had 1 message. Therefore, the rest of 105, 595 messages were separated into 3, 731 groups. The largest group had 2, 546 messages. There were 15 groups with at least 1000 elements. We presented the statistics of group sizes in (Table 4). Searching with keywords listed in (Textbox 2), we identified that among 936 rumors with at least 10 reported messages, 44.5% (417) were related to COVID-19. We presented case studies of three popular rumors discovered in the dataset. All three rumors were fact-checked to be false and misleading during the period our dataset covered. We picked out some major societal events in (Table 5) that precede each peak in messages reported in (Figure 7, 9, 10). While there are multiple important events happening every day, we picked out incidents that were the “first” occurrences.

Table 4. Statistics of group sizes (114 thousand documents).

Groups with at least X messages	Mean	Standard Deviation	Maximum	Q_3	Q_2
1	9.309	71	2546	2	1
2	28.302	126.907	2546	10	3
10	102.96	238.31	2546	75	27

Date	Event
Feb 9, 2020	• First asymptomatic laboratory-confirmed case in Taiwan.
Feb 15, 2020	• First COVID-19 death case in Taiwan.
Feb 21, 2020	• Passengers on Diamond Princess returned to Taiwan.
Mar 11, 2020	• COVID-19 declared a global pandemic by WHO.
Mar 18, 2020	• Director of CECC, Chen, Shih-Chung, went to the Legislative Yuan for interpellation about the pandemic for the first time. • Reached a total of 100 confirmed cases. Single day confirmed cases hit record high for 3 consecutive days.
Mar 26, 2020	• CECC released the first report on the analysis of confirmed cases in Taiwan.
Mar 30, 2020	• First death case in Taiwan's first hospital cluster infection.
Apr 1, 2020	• A day before 4-day long weekend. • First day of mask requirement on public transportation.
Apr 5, 2020	• Last day of a 4-day long weekend.

Table 5. Major societal events related to COVID-19.

Textbox 2. 22 keywords used to identify a COVID-related message.

□□, □□, □□, □□, □□, 2019-ncov, covid, □□, □□, □□, □□, □□, □□, □□□, ibuprofen, □□, □□, □□, □□□□, □□□, □□, □□

Case 1: Do not go outside!

The rumor content is presented in (Table 6). This rumor first appeared in the dataset on Feb 2, 2020. Over the course of 3.5 months, there were a total of 2,119 messages reported. As shown in (Figure 7), the reported messages went viral at least four times: it peaked on Feb 22 with 80 messages, Mar 16 with 68. Then, on Apr 2, it welcomed the highest with 205. And lastly, 166 messages on Apr 6. During this period, we observed several content changes (See (Table 7)).

First, the time-sensitive information in the messages evolved. At its early stage, "Lantern Festival" (Feb 8, 2020), was spotted in most of the messages. However, on Feb 18, there was the first one that replaced "Lantern Festival" with "March". Then, after Mar 10, most reported messages used "Mid-Autumn Festival (June 25, 2020)".

Secondly, 98.87% of the reported messages falsely quoted authority, with the most prevalent being Zhong, Nan-Shan (46%) and Chen, Shih-Chung (52.7%). Efforts were made to emphasize the authoritativeness of the quoted party as well. For example, titles for *The Mainland Academian Zhong, Nan-Shan* became longer: from *"Expert in Pandemic from Mainland China"*, *"Expert in Coronavirus"*, to *"Expert in Coronavirus from Mainland China, 78-year-old Academian Zhong, NanShan"*. Then starting from Apr 1, 2020, every reported message had Zhong replaced by Chen, Shih-Chung (□□□) (See (Figure 8)). As the Minister of Health and Welfare and Director of Taiwan's CECC, Chen's popularity skyrocketed during the pandemic through his daily press conference. This was also when we observed the largest number of reported messages.

Due to the prevalence of this message spreading on web and closed platforms, the Ministry of Health and Welfare and CECC both sent out a press release [26] on Apr 2, 2020, reminding the public that this was false. Nevertheless, this did not stop another viral spread of the same message at the end of a four-day long holiday in Taiwan, where crowds were seen in every tourist attraction on the island. For days people were worried that the long weekend would lead to another outbreak of the pandemic, which explained why the message bearing the key topic "do not go out" would become a

big hit.

Table 6. Content of Case 1 rumor.

English Translation	Original
Academian Zhong, Nan-Shan emphasized repeatedly, “Do not go outside! At least wait until at least the Lantern Festival.” Be warned that even if cured, you would suffer the rest of your life. This is a plague worse than SARS. The side effects of the drugs are more severe. Even if there is special medicine, it could only save your life, nothing more. Think about your family before stepping outside. […] This is a war, not a game [...] No one is an outsider in this war [...] Please share it with others. By Zhong, Nan-Shan.	[illegible] 17 [illegible]——[illegible]

Table 7. Change log for Case 1 rumor contents.

Date	Previous	New
Feb 17, 2020	**Academian Zhong, Nan-Shan** stressed □□□□□□□□□	**Pandemic expert from Mainland China, Academian Zhong, Nan-Shan** stressed □□□□□□□□□□□□□□□
Feb 18, 2020		**Coronavirus expert from Mainland China, 78-year-old Academian Zhong, Nan-Shan** stressed □□□□□□□□□□□□ 78 □□□□□□
Feb 27, 2020		**Coronavirus expert from Mainland China, 84-year-old Academian Zhong, Nan-Shan** stressed □□□□□□□□□□□□ 84 □□□□□□
Apr 1, 2020		**Director of Taiwan's Ministry of Health and Welfare, Chen, Shih-Chung**, reminded everyone □□□□□□□□□□□□□
Feb 18, 2020	Do not go outside! At least wait until the **Lantern Festival**. □□□□□□□□□□□□□□□□	Do not go outside! At least wait until the **Mid-Autumn Festival**. □□□□□□□□□□□□□□□□□

Case 2: Drink salty water can prevent the spread of COVID-19.

This rumor promoted drinking salt water to prevent the coronavirus. In fact, we investigated two such rumors and the combination of them as shown in (Table 8).

Both Message (A) and (B) were promoting "drinking water to protect oneself from the virus". This, in fact, was a popular myth about COVID-19 internationally. This practice was fact-checked several times in March by Taiwan's fact-checking agencies[27, 28], and even WHO had fact-checked a similar claim about rinsing nose from saline[29]. However, this did not stop the piece from receiving attention (See (Figure 9)). To make matters worse, several translations of Message (A+B) were seen in April, including but not limited to English, Indonesian, Filipino and Tibetan.

Similar to the previous case, 94.2% of reported messages in Case 2 included misquotations of medical professionals, with the more popular ones being *Director of The Veteran Hospital* (22.94%) and *Dr. Wang of Tung Hospital* (71.28%).

The lifespan of this "drink salted water" rumor was rather long. One famous fact-checking platform in Taiwan, MyGoPen, released an article to disprove this false medical advice again in October 2020 [30], 7 months after it was first seen in our dataset.

Table 8. Content of Case 2 rumor.

	English Translation	Original
(A)	This is a 100% accurate information... Why did we see a huge decline of confirmed cases in China during the last few days? They simply forced their citizens to rinse mouths with salted water 3 times a day and then drink water for 5 minutes. The virus would attack throats before the lungs, and when getting in touch with salted water, the virus would die or get destroyed in lungs. This is the only way to prevent the spread of COVID-19. There is no need to buy medicine as there is nothing effective on the market.	□□ 100%□□□□□... □□□□□□□□□□□ □□□□□□□□□□□□□□□□□□□□□ 3 □ □□□□□□□□ 5 □□□□□□□□□□□□ □□□□□□□□□□□□□□□□□□□□□□□□ □□□□□□□□□□□□□□□□□□□□□□□ □□□□□□□□□□□□□□
(B)	Before reaching the lungs, the Novel Coronavirus would survive in throats for four days. At this stage, people would experience sore throats and start coughing. If one can drink as much warm water with salt and vinegar, the virus could be destroyed. Share this information to save people's lives.	□□□□□□□□□□□□□□□□□□□□□ 4 □□□□□□□□□□□□□□□□□□□□□□□□ □□□□□□□□□□□□□□□□□□□□□□□ □□□□□□□□□□□
(A+B)	Why did Mainland China show a huge decline of confirmed cases over the last few days? Besides wearing masks and washing hands, they simply rinse mouths with salted water 3 times a day and then drink water for 5 minutes [...] Dr. Wang of Tung Hospital stated that the Novel Coronavirus would survive in throats for four days before reaching the lungs [...] If one can drink as much warm water with salt and vinegar, the virus could be destroyed [...]	□□□□□□□□□□□□□□□□□□□□□□□□ □□□□□□□□□□□□□ 3 □□□□□□□□□ 5 □□[...] □□□□□□□□□□□□□□□□□ □□□□ 4 □[...] □□□□□□□□□□□□□□□ □□□□□□[...]

Case 3: This is a critical period, here are some suggestions...

The content of this rumor is presented in (Table 9). This rumor first appeared in the dataset on Feb 6, 2020 and had a total of 2121 reports. Over the 1.5 months of its most popular time, it went viral at least two times: one on Feb 17 with 394 reports and on Mar 19 with 543 (Figure 10). The content started with an authoritative tone that announced, "We are at the most critical period of COVID-19", and then provided a list of *suggestions*. While some *suggestions* made medical sense for hygiene, others did not.

It was not stated explicitly in the message what the critical period was referring to, however, when taking together the listed "guidelines" into account, we could deduce that it hinted at the "critical period to prevent community spread". Community spread is a phase in a pandemic where many people who tested positive in an area cannot be determined how they got infected. It is not hard to imagine that people would be concerned and worried about this significant phase where the risk of getting infected is greatly increased. In fact, we observed that such concerns co-occurred with the spread of this piece of message in February.

On Feb 15, 2020, Taiwan's CECC reported that a taxi driver, infected by a person traveled back from China, was tested positive with the virus. He died on the same day and became the first death case in Taiwan. Over the next four days, four of his family members were also tested positive, forming the first COVID-19 local cluster in Taiwan. During that time, people's concerns for community spread were looming. In fact, Google trend for search term "□□□□ (Community Spread)" sharply increased on Feb 16 (Figure 11). Also, during this period, the number of the reported messages sharply increased (Figure 10).

Content-wise, 88% of the reported messages had quoted several authorities to *endorse* the information, with the majority being Taiwan Medical Association (81.5%) and CECC Director Chen, Shih-Chung (18.7%) (Figure 12). We spotted a major revision of the message (Table 10) on Feb 12, 6 days after the first report, where the 18 bullets were pruned to 14, and strong words were modified to

gentler tone. Moreover, the message added a signature of *"Regards from Taiwan Medical Association"* on the last line. This became the most widespread version afterwards. Out of the 394 messages reported on Feb 17, 84.5% (333) were of this version. Another key event in content transformation occurred on Mar 18 when Chen, Shih-Chung, went to the Legislative Yuan (similar to the Congress in the US) for interpellation about the pandemic. On the same day, messages started to misquote the director: "Chen, Shih-Chung explained in Legislative Yuan on Mar 18 (3/18 □□□□□□□□)" (Table 10). The next day, we saw another sharp increase of reported messages, reaching the highest peak. Of the 543 messages reported on Mar 19, 51.2% (280) misquoted Chen.

In terms of fact-checking, from a rather early period between Feb 10 to Feb 15, several fact-checking agencies published reports pointing out the falsity of the message, [31-33]. The misquoted party, Taiwan Medical Association, also released a clarifying statement on Feb 12 [34]. However, like what we observed in the previous two cases, such fact-checking efforts did not avoid the message from getting widespread attention in later time.

Table 9. Content of Case 3 rumor.

English Translation	Original
10 days from now, Taiwan is in a critical period to combat COVID-19. Here are some suggested measures. 1. Strictly prohibited going to public places. 2. Takeout from restaurants. 3. Eat outside in open spaces. 4. Wash your hands the right way (extremely important). 5. When taking the subway or bus, choose the seats at the first half of the vehicle. 6. Do not wear contact lenses. 7. Eat warm food and more vegetables. 8. Avoid constipation. 9. Drink warm water. 10. Do not visit hair salons. 11. Hang the clothes you're wearing outside for two hours the first thing you get home. 12. Do not wear jewelry. 13. Wash your hands immediately after touching cash or coins. Put coins you just received inside a	□□□□ 10 □□□□□□□□□□□□□□□□□□□□□□: 1.□□□ □□□□□□ 2.□□□□□□□□□□ 3.□□□□□□□□□ 4.□ □□□□□□(□□□□)□ 5.□□□(□□)□□□□□□□□ 6.□ □□□□□□ 7.□□□,□□□□□□,□□□□ 8.□□□□□□ 9. □□□□□ 10.□□□□□□ 11.□□□□□(□□,□□),□□□□ □□□□ 2 □□ 12.□□□□□□ 13. □□□□□□,□□□□□, □□□□□□□,□□□□□□□□□,□□□,□□□□. 14.□□□□ □□□□□□□□□□□□□□□□□15. □□□□□□□. 16.□□ □□□□□□□. 17.□□□□□□18.□□□□□□□□

plastic bag for one day before using them. 14. Do not use a colleague’s phone when working. If you have to, disinfect before using. 15. Avoid taking public transportation during rush hour. 16. Do not visit night markets or traditional markets. 17. Exercise. 18. Avoid going to the gym.	

Table 10. Change log for Case 3 rumor content.

Date	Previous	New
Feb 12, 2020	1. Strictly prohibited going to public places. 1.□□□□□□□□□	1. Reduce going to public places. 1.□□□□□□□□□
	3. Eat outside in open spaces. 5. When taking the subway or bus, choose the seats at the first half of the vehicle. 10. Do not visit hair salons. 16. Do not visit night markets or traditional markets. 3.□□□□□□□□□ 5.□□□(□□)□□□□□□□□ 10.□□□□□□ 16.□□□□□□□□□.	*deleted*
		[add] Regards from Taiwan Medical Association □□□□□□□□
Mar 18, 2020	10 days from now, Taiwan is in a critical period to combat COVID-19. Here are some suggested measures. [...] □□□ 10 □□□□□□□□□□□□□□□□□□□□ [...]	10 days from now, Taiwan is in a critical period to combat COVID-19 **(Explained by Chen, Shi-Chung in Legislative Yuan on Mar 18, 2020)**. Here are some suggested measures[...] □□□ 10 □□□□□□□□□□□□□□□□□□(3/18 □□□□□□□□) □□□□ [...]
	Regards from Taiwan Medical Association □□□□□□□□	*deleted*

Discussion

Principal findings

In this paper, we investigated a dataset of 114,124 messages from a closed-messaging platform, LINE. We proposed an algorithm that reduced the computational time of clustering from exponential to linear time. The algorithm enabled us to identify 417 COVID-related rumors with at least 10 messages and investigate the evolution of such rumors.

Looking into 3 popular false rumors, we observed some commonalities. First observation is the frequent appearances of key authorities. Given the nature of the pandemic, the authorities were usually medical personnel. At times, the change of quoted figures signaled the paradigm shift of whom the public looked up to. For example, from Zhong Nan-Shan to Chen Shi-Chung. Other times, the quoted party did not seem to make any sense. For example, the Dr. Wang in Case 2 was in fact an Orthopedist. In addition, similar to the findings of Wood et al. [35], it is quite obvious that the current practice of fact-check did not reduce the spread of false information. In all three cases, we observed that false COVID-19 rumors resurfaced multiple times even after being fact-checked, and with different degrees of content alterations. By identifying major societal events preceding each resurfacing peak, we suspected that resurfacing patterns were influenced by major societal events and textual transformation. However, each peak of popularity would not last long and it was often without good explanation about how one wave of widespread attention ended.

In addition to the above 3 case studies, we went through 5 other COVID-19 related rumors and manually identified common patterns of textual changes during rumor propagation. It was quite common to observe messages having a line or two disclaimers that expressed the uncertainty of truthfulness of the forwarded messages. For example, "*The following is for your reference only, I do not guarantee the truthfulness of the message. (□□□□□□□□□□□□□□□)"* was often seen. Secondly, most messages also included Simplified Chinese characters and terms that are rarely used in Taiwan. For example, while in Taiwan, people refer to the SARS pandemic as "SARS", but many messages used "□□", a term popularly used in China. We also noticed messages that were a merge of other previously independent ones, and messages that included translation to other non-Chinese languages.

These characteristics could serve as rules to discover possible false information as early detection mechanisms. Although we identified these characteristics manually this time, it is quite possible to employ techniques such as Natural Language Processing to automatically recognize these textual changes in the future, making it possible to have an automatic early warning system of possible misinformation before fact-check efforts by professionals.

Limitations

This study had several limitations. First, this data was collected by LINE user's reports. Therefore, it was impossible to infer the true distribution of messages without making some assumptions. That is, if we saw more health-related misinformation in our

data, it did not necessarily translate to more health-related rumors circulating in the platform. In fact, it could also be that people were more alerted and skeptical at truthfulness health-related information. In addition, we only looked at text messages, therefore, information distributed visually or in audio was not covered. Lastly, our algorithm to group messages does not work well with short texts.

Conclusion

While social media may give rise to an unprecedented number of unverified rumors, it also provided a unique opportunity to study rumor propagation. In fact, to combat with *infodemic*, we need to first understand how and why some rumors became popular. In our studies, we proposed an algorithm that enables the research community to perform large-scale studies on the evolution of text messages at rumor-level rather than topic-level. Moreover, we showed textual commonalities in widespread rumors in Taiwan during COVID-19. And that the attention one rumor received was tied with major societal events and content changes, for example, changing whom to quote in the message. To the best of our knowledge, this is one of the few works that study COVID-19 misinformation on closed-messaging platforms and the first to study textual evolution of COVID-19 related rumors during its propagation. We would hope that this would further spark more studies in rumor propagation patterns as an effort to fight with the *infodemic*.

Acknowledgements

We cited a survey from Taiwan Communication Survey in Method/Data section and it is

required of us to have the following acknowledgement: “Data were collected by the research project of the Taiwan Communication Survey (TCS), which is supported by the Ministry of Science and Technology of R.O.C. The author(s) appreciate the assistance in providing data by the aforementioned institute. The views expressed herein are the authors’ own. doi: 10.6141/TW-SRDA-D00176-1”

Abbreviations

CECC - Central Epidemic Command Center

References

1. Lazer D, Baum M, Benkler Y, Berinsky A, Greenhill K, Menczer F, et al. The science of fake news. Science. 2018;359(6380):1094-6.
2. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science. 2018;359(6380):1146-51. doi: 10.1126/science.aap9559.
3. Del Vicario M, Bessi A, Zollo F, Petroni F, Scala A, Caldarelli G, et al. The spreading of misinformation online. Proceedings of the National Academy of Sciences. 2016;113(3):554. doi: 10.1073/pnas.1517441113.
4. World Health Organization. Novel coronavirus (2019-nCoV) situation report. 2020 [updated 2020/02/022021-05-20]; Available from: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-ncov-v3.pdf.
5. Abdoli A. Gossip, rumors, and the COVID-19 crisis. Disaster Medicine and Public Health Preparedness. 2020;14(4):e29-e30. doi: 10.1017/dmp.2020.272.
6. Tasnim S, Hossain M, Mazumder H. Impact of rumors and misinformation on COVID-19 in social media. J Prev Med Public Health. 2020 5;53(3):171-4. doi: 10.3961/jpmph.20.094.
7. Mat Dawi N, Namazi H, Hwang H, Ismail S, Maresova P, Krejcar O. Attitude toward protective behavior engagement during COVID-19 pandemic in Malaysia: the role of e-government and social media. Frontiers in Public Health. 2021 2021/03/01;9(113). doi: 10.3389/fpubh.2021.609716.
8. Mubeen S, Kamal S, Kamal S, Balkhi F. Knowledge and awareness regarding spread and prevention of COVID-19 among the young adults of Karachi. J Pak Med Assoc. 2020 May;70(Suppl 3)(5):S169-s74. PMID: 32515406. doi: 10.5455/jpma.40.
9. Pulido C, Villarejo-Carballido B, Redondo-Sama G, Gómez A. COVID-19 infodemic: more retweets for science-based information on coronavirus than for false information. International Sociology. 2020;35(4):377-92. doi:

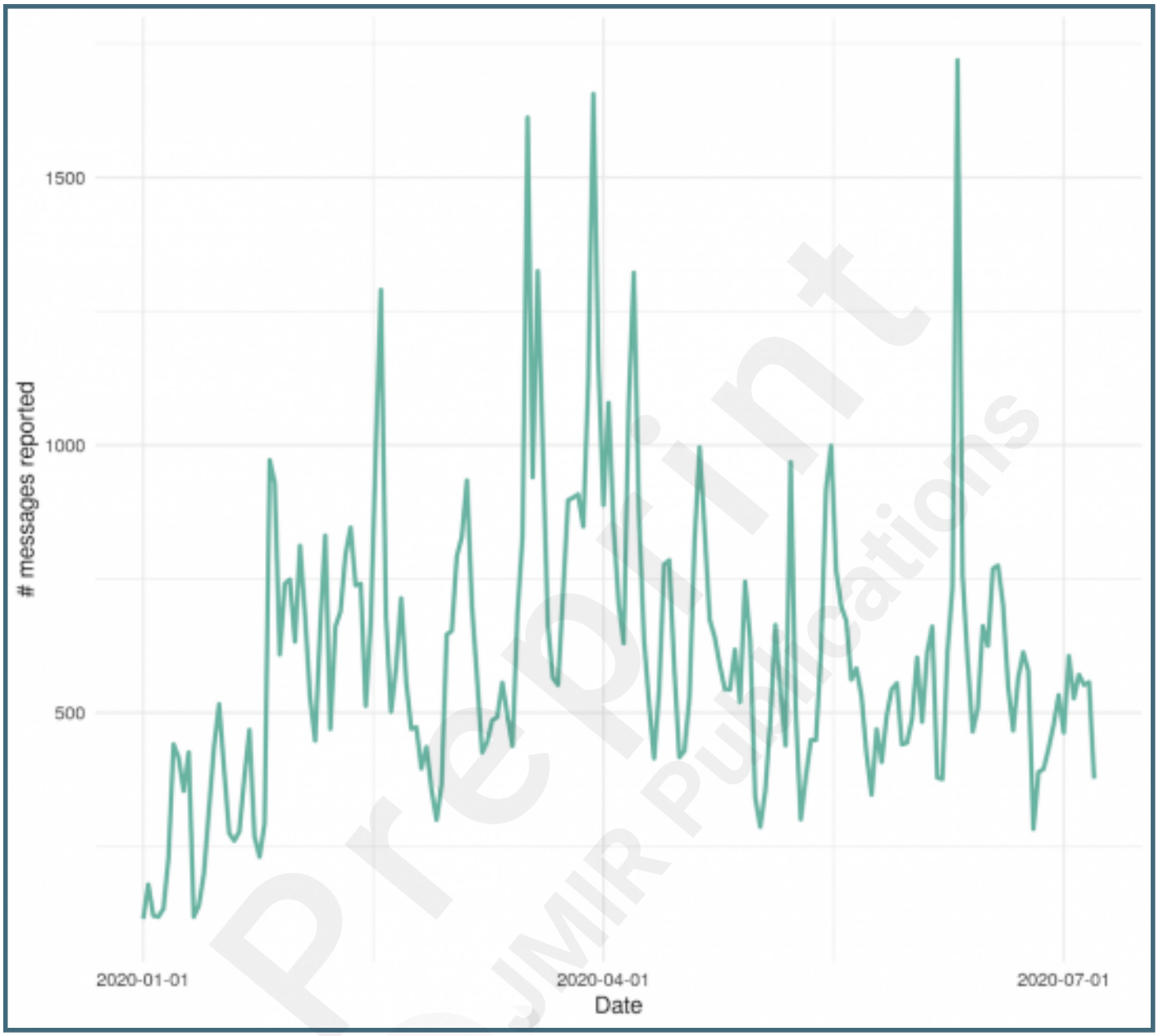
10.1177/0268580920914755.
10. Cinelli M, Quattrociocchi W, Galeazzi A, Valensise C, Brugnoli E, Schmidt A, et al. The COVID-19 social media infodemic. Scientific Reports. 2020 2020/10/06;10(1):16598. doi: 10.1038/s41598-020-73510-5.
11. Gallotti R, Valle F, Castaldo N, Sacco P, De Domenico M. Assessing the risks of 'infodemics' in response to COVID-19 epidemics. Nature Human Behaviour. 2020 2020/12/01;4(12):1285-93. doi: 10.1038/s41562-020-00994-6.
12. Shi A, Qu Z, Jia Q, Lyu C, editors. Rumor detection of COVID-19 pandemic on online social networks. 2020 IEEE/ACM Symposium on Edge Computing (SEC); 2020 12-14 Nov. 2020.
13. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: infoveillance study. J Med Internet Res. 2020 2020/4/21;22(4):e19016. doi: 10.2196/19016.
14. Kwok S, Vadde S, Wang G. Tweet Topics and Sentiments Relating to COVID-19 Vaccination Among Australian Twitter Users: Machine Learning Analysis. J Med Internet Res. 2021 2021/5/19;23(5):e26953. doi: 10.2196/26953.
15. Jelodar H, Wang Y, Orji R, Huang S. Deep sentiment classification and topic discovery on novel coronavirus or COVID-19 online discussions: NLP using LSTM recurrent neural network approach. IEEE J Biomed Health Inform. 2020 Oct;24(10):2733-42. PMID: 32750931. doi: 10.1109/jbhi.2020.3001216.
16. Jo W, Lee J, Park J, Kim Y. Online Information exchange and anxiety spread in the early stage of the novel coronavirus (COVID-19) outbreak in South Korea: structural topic model and network analysis. J Med Internet Res. 2020 2020/6/2;22(6):e19455. doi: 10.2196/19455.
17. Zhang C, Xu S, Li Z, Hu S. Understanding concerns, sentiments, and disparities among population groups during the COVID-19 pandemic via Twitter data mining: large-scale cross-sectional study. J Med Internet Res. 2021 Mar 5;23(3):e26482. PMID: 33617460. doi: 10.2196/26482.
18. Boehm L. The validity effect: a search for mediating variables. Personality and Social Psychology Bulletin. 1994 1994/06/01;20(3):285-93. doi: 10.1177/0146167294203006.
19. Berinsky A. Rumors and health care reform: experiments in political misinformation. British Journal of Political Science. 2017;47(2):241-62. doi: 10.1017/S0007123415000186.
20. DiFonzo N, Bordia P. Rumor psychology: social and organizational approaches: American Psychological Association; 2007. x, 292-x, p. ISBN: 1-59147-426-4 (Hardcover); 978-159147-426-5 (Hardcover).
21. Yang K, Torres-Lugo C, Menczer F. Prevalence of low-credibility information on twitter during the COVID-19 outbreak. arXiv preprint arXiv:200414484. 2020.
22. Shin J, Jian L, Driscoll K, Bar F. The diffusion of misinformation on social media: Temporal pattern, message, and source. Computers in Human Behavior. 2018

2018/06/01/;83:278-87. doi: 10.1016/j.chb.2018.02.008.
23. Ng L, Loke J. Analyzing public opinion and misinformation in a COVID-19 telegram group chat. IEEE Internet Computing. 2021;25(2):84-91. doi: 10.1109/MIC.2020.3040516.
24. Bastani P, Bahrami M. COVID-19 related misinformation on social media: a qualitative study from Iran. J Med Internet Res. 2020 Apr 5. PMID: 32250961. doi: 10.2196/18932.
25. "Jieba" (Chinese for "to stutter") Chinese text segmentation: built to be the best Python Chinese word segmentation module. 0.42.1 ed. GitHub.
26. Ministry of Health and Affair. "Director Chen said do not go out before Dragonboat festival" is a false information. 2020 [2021-05-20]; Available from: https://www.mohw.gov.tw/cp-4633-52577-1.html.
27. MyGoPen. Coronavirus will stay at your throat for four days? Fake picture and a false rumor! 2020 [2021-05-20]; Available from: https://www.mygopen.com/2020/03/gargling-eliminate-coronavirus.html.
28. Taiwan FactCheck Center. Viral on the internet, "Drinking warm water with salt and vinegar could eradicate coronavirus" is a false information. 2020 [2021-05-20]; Available from: https://tfc-taiwan.org.tw/articles/3207.
29. World Health Organization. FACT: Rinsing your nose with saline does NOT prevent COVID-19. [2021-05-20]; Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters#saline.
30. MyGoPen. Misleading false rumor - "Drink salt water to prevent from coronavirus? 100% correct?". 2020 [2021-05-20]; Available from: https://www.mygopen.com/2020/10/salt-water.html.
31. Taiwan FactCheck Center. Clarification about "From Taiwan Medical Association, 10 days from today, Taiwan enters the critical period of COVID-19". 2020 [2021-05-20]; Available from: https://tfc-taiwan.org.tw/articles/2547.
32. MyGoPen. "10 days from today, Taiwan enters the critical period of COVID-19" is a misleading information. 2020 [2021-05-20]; Available from: https://www.mygopen.com/2020/02/10-key.html.
33. Rumor & Truth. Rumor has it goes "10 days from today, Taiwan enters the critical period of COVID-19", so what date is today? 2020 [2021-05-20]; Available from: https://www.rumtoast.com/12842.
34. Taiwan Medical Association. Taiwan Medical Association clarification statement. 2020 [2021-05-20]; Available from: https://www.tma.tw/meeting/meeting_info04.asp?/9112.html.
35. Wood T, Porter E. The elusive backfire effect: mass attitudes' steadfast factual adherence. Political Behavior. 2019 2019/03/01;41(1):135-63. doi: 10.1007/s11109-018-9443-y.

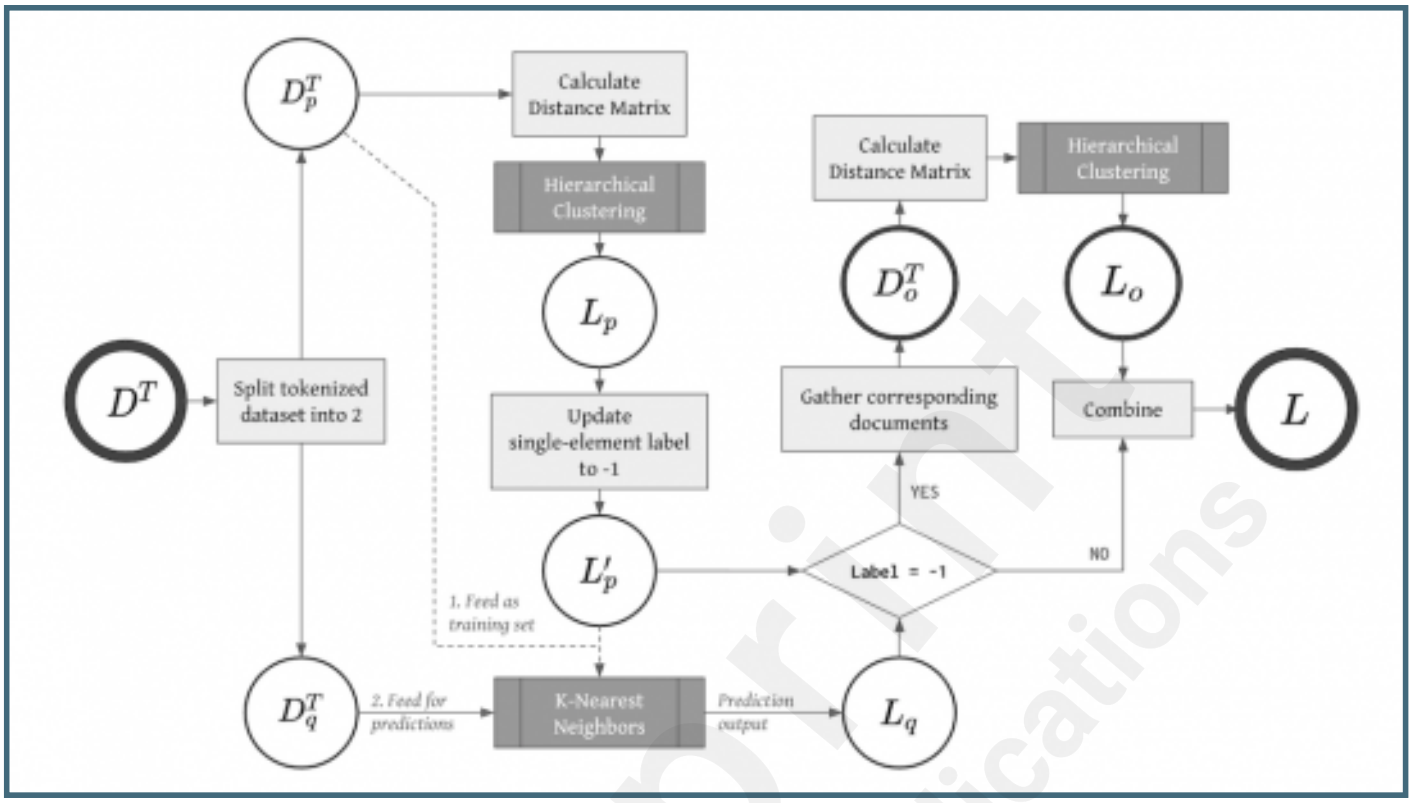
Supplementary Files

Figures

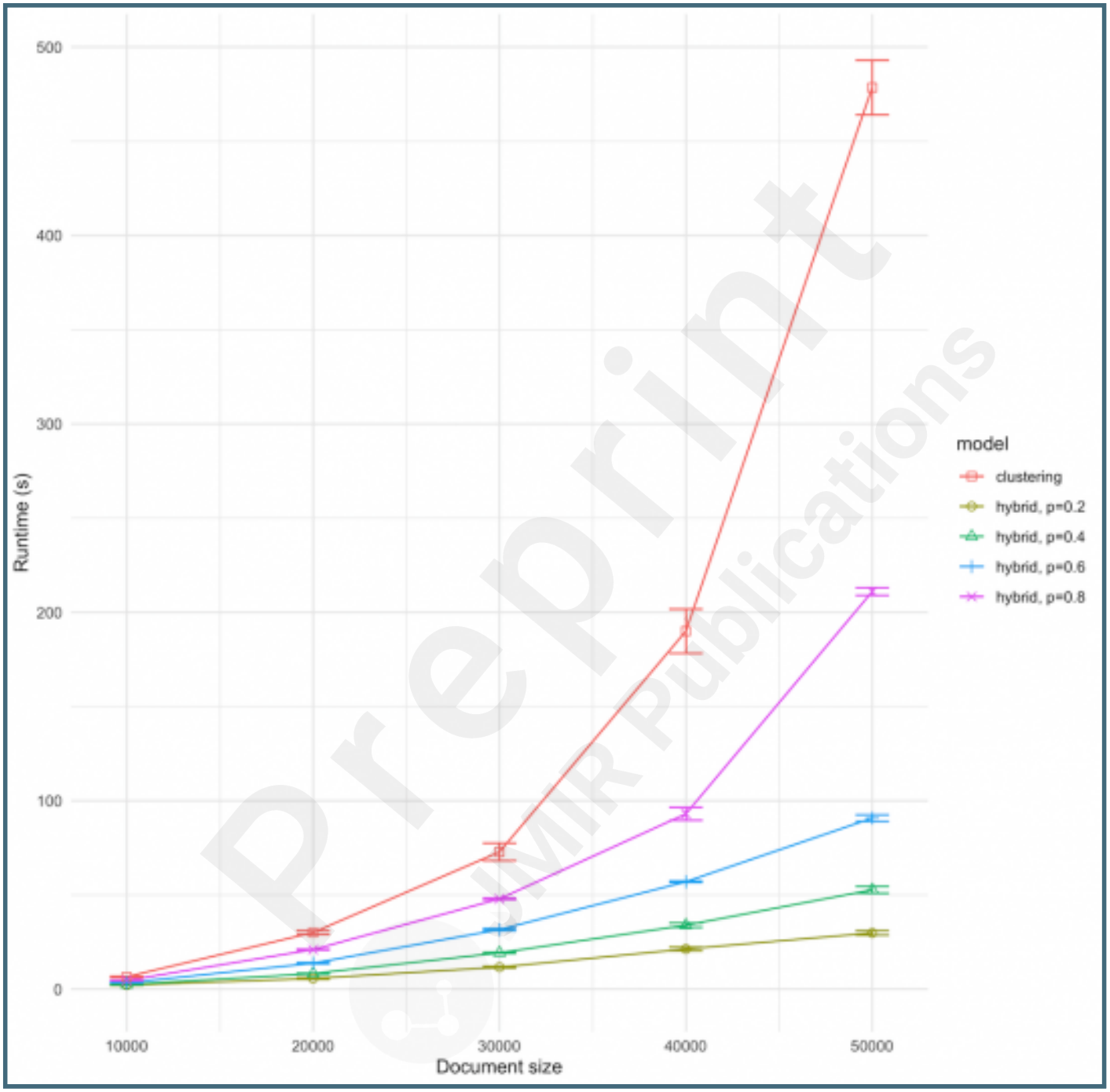
Number of suspicious messages reported by date.



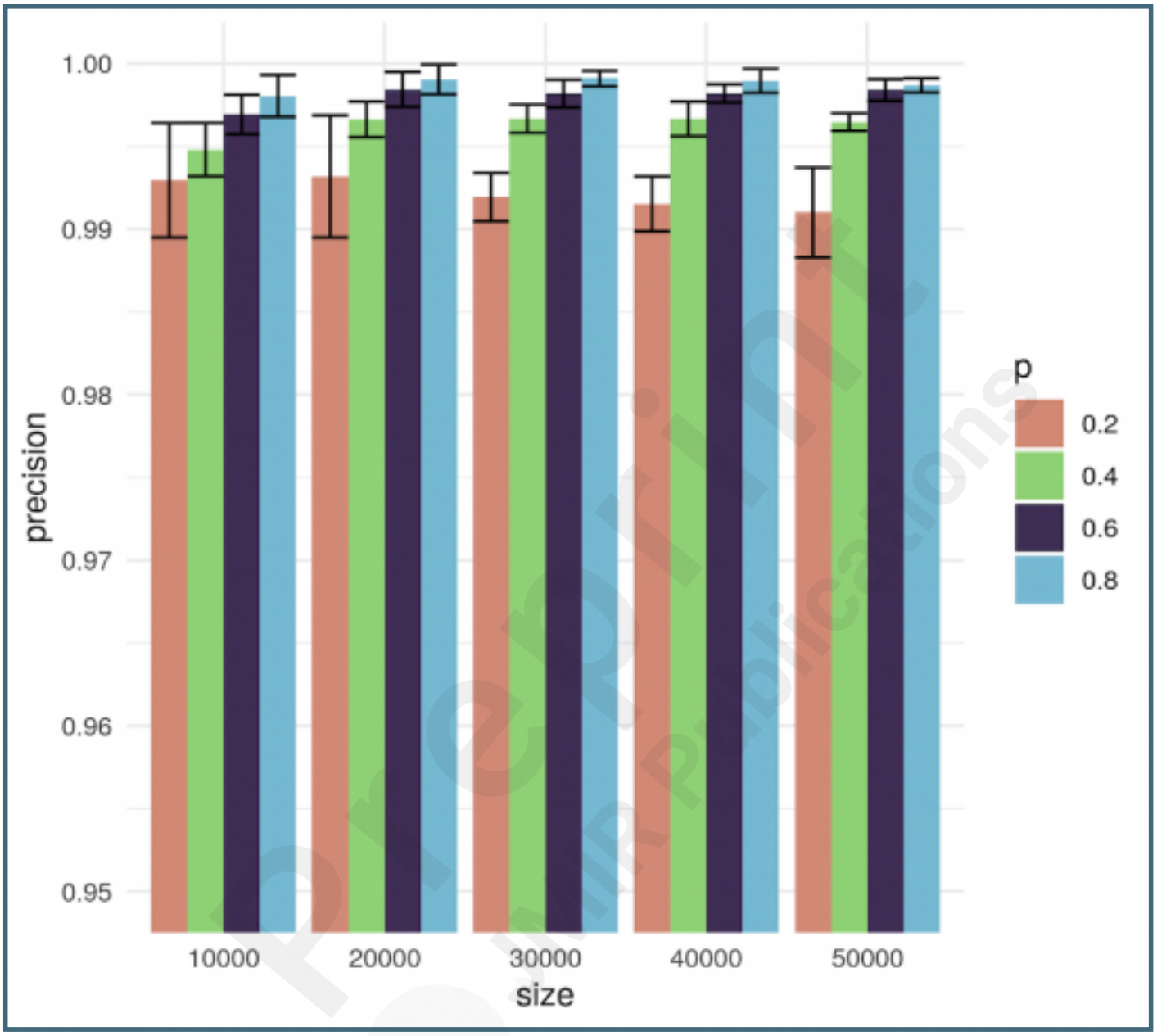
The Cluster-Classification, "Hybrid", Algorithm flow diagram.



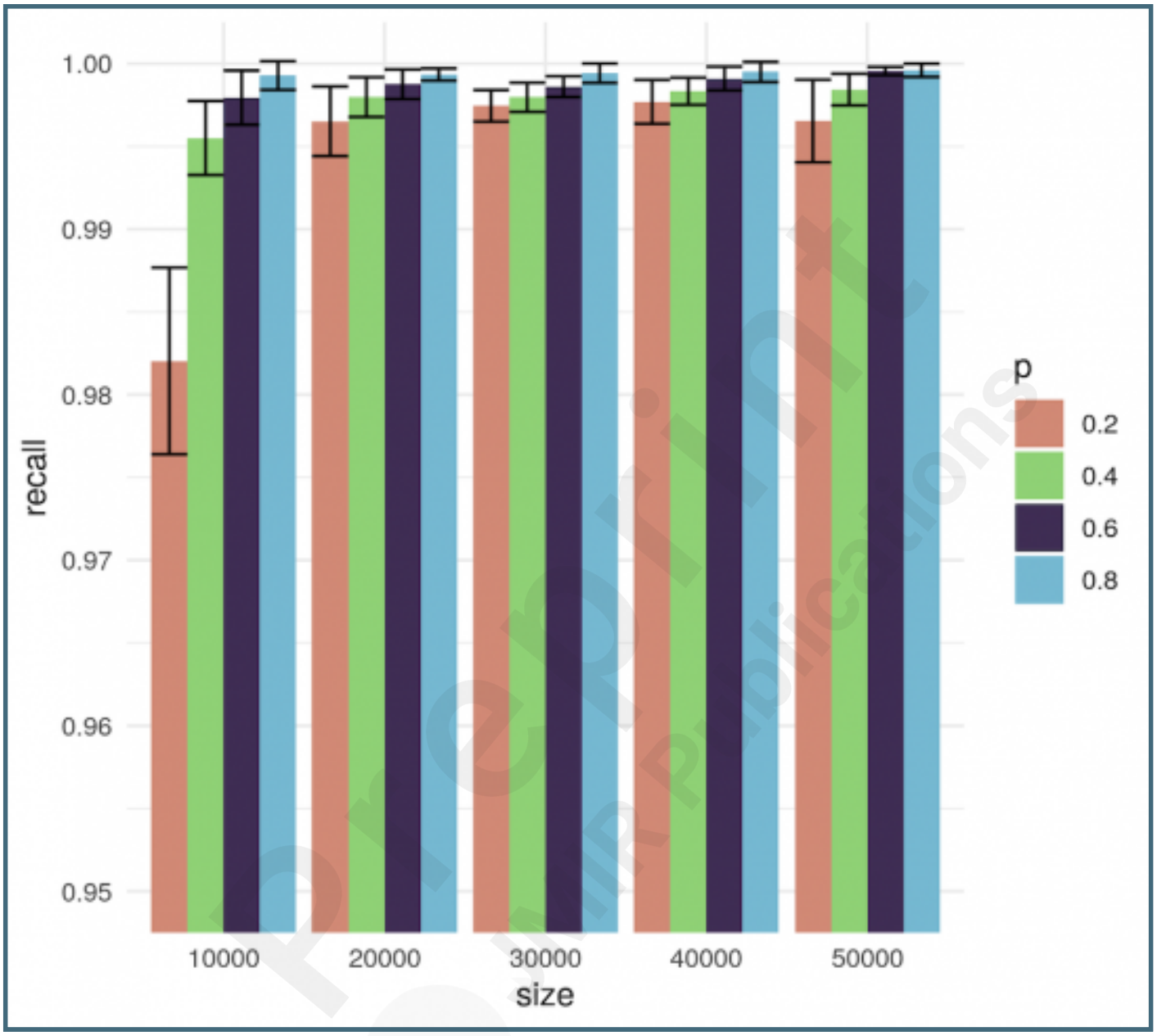
Speed comparisons with 95% confidence interval between clustering and hybrid algorithm, across different levels of p. Using hybrid with p lower than 0.6 reduced the runtime from exponential to linear time.



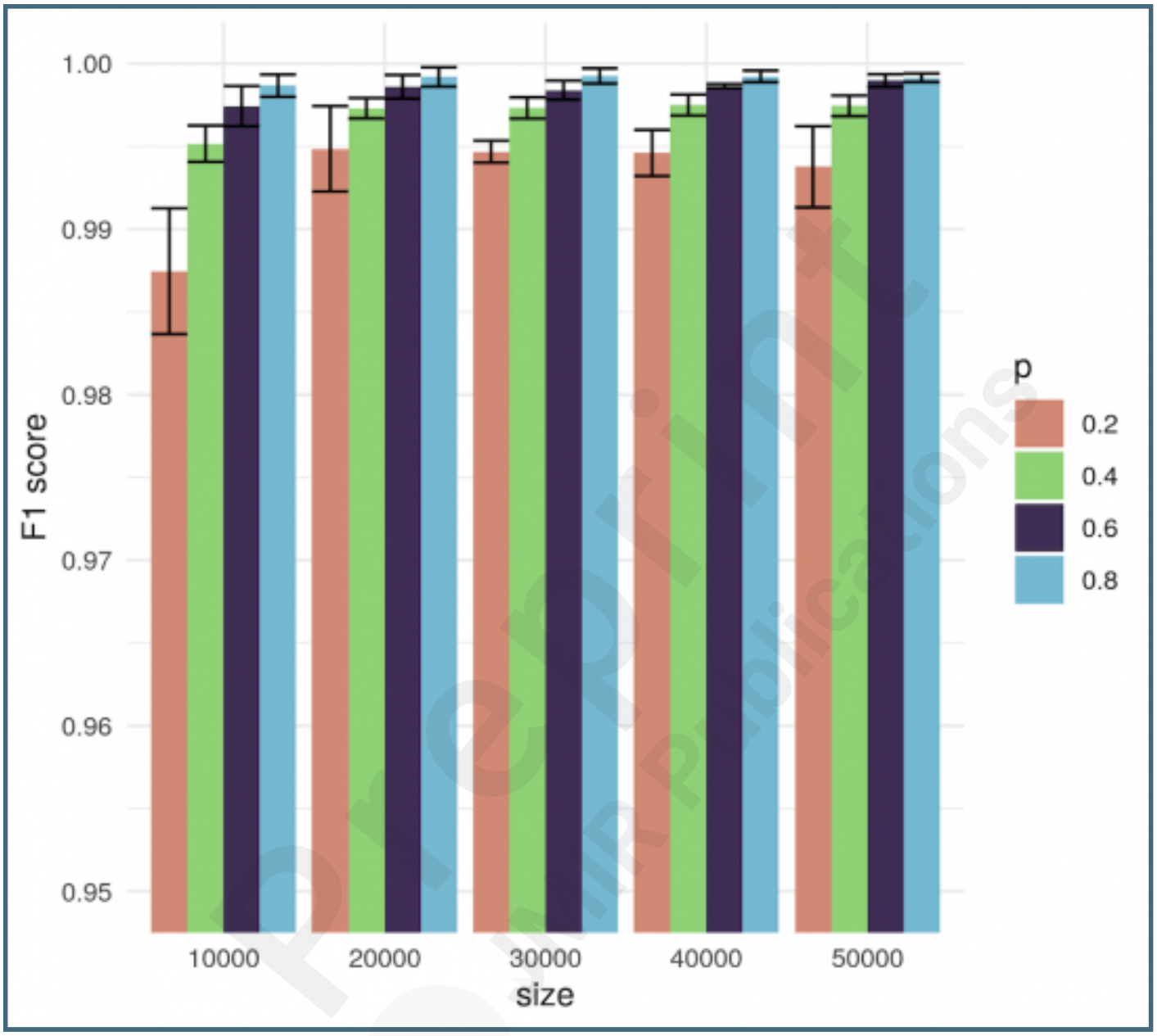
Precisions and 95% confidence interval of the hybrid algorithm across different document sizes and train portion p.



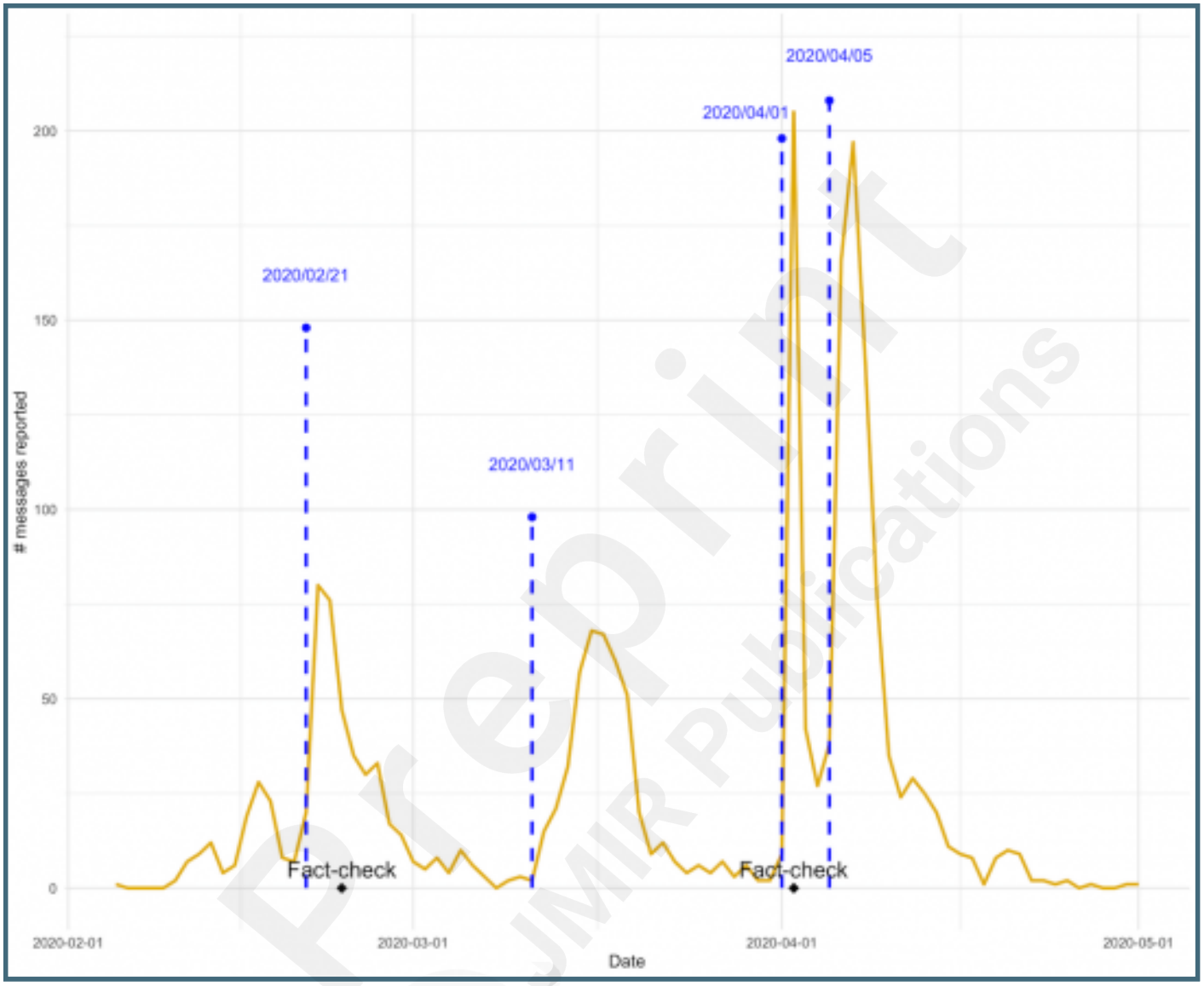
Recalls and 95% confidence interval of the hybrid algorithm across different document sizes and train portion p.



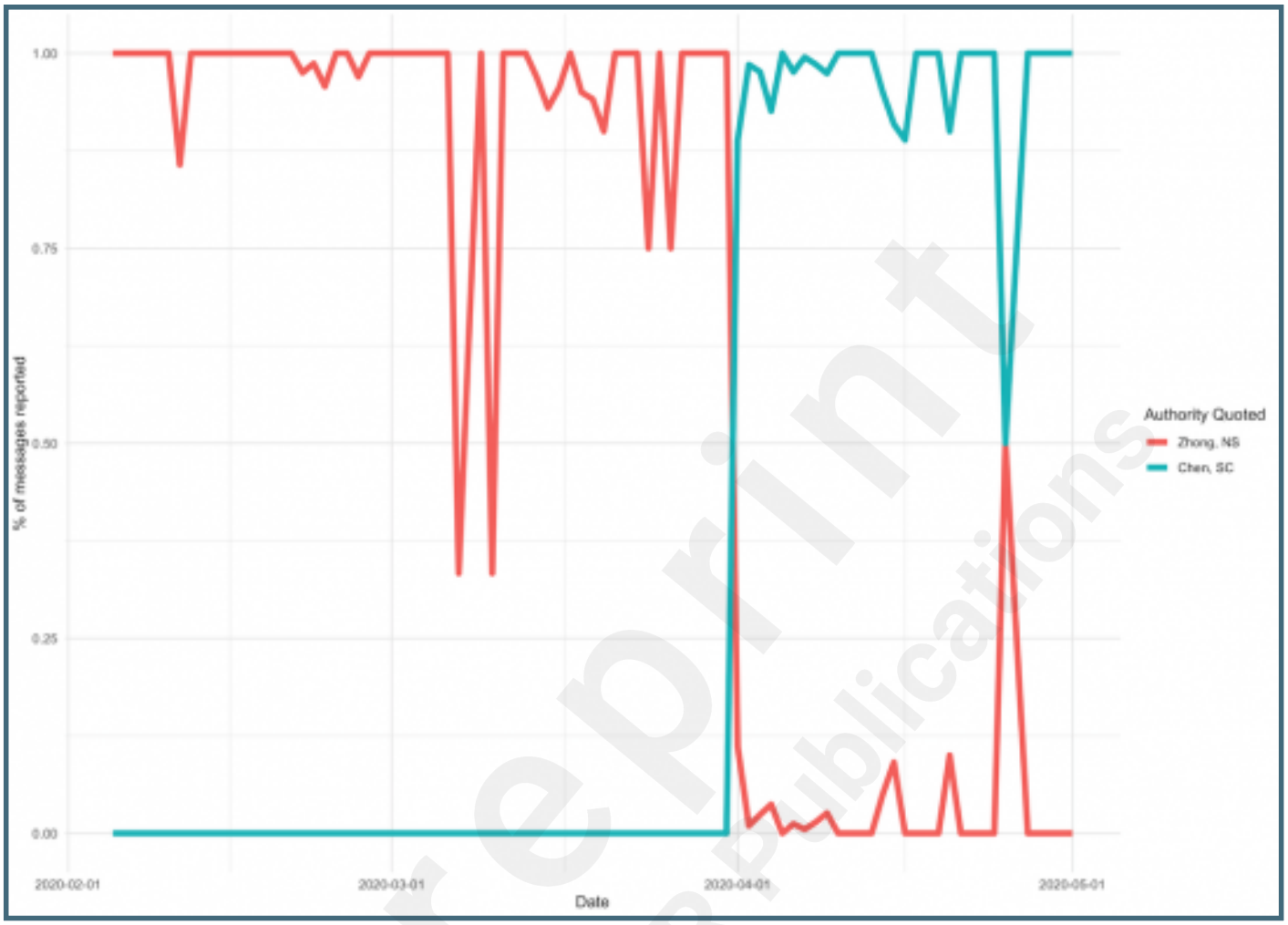
F-scores and 95% confidence interval of the hybrid algorithm across different document sizes and train portion p.



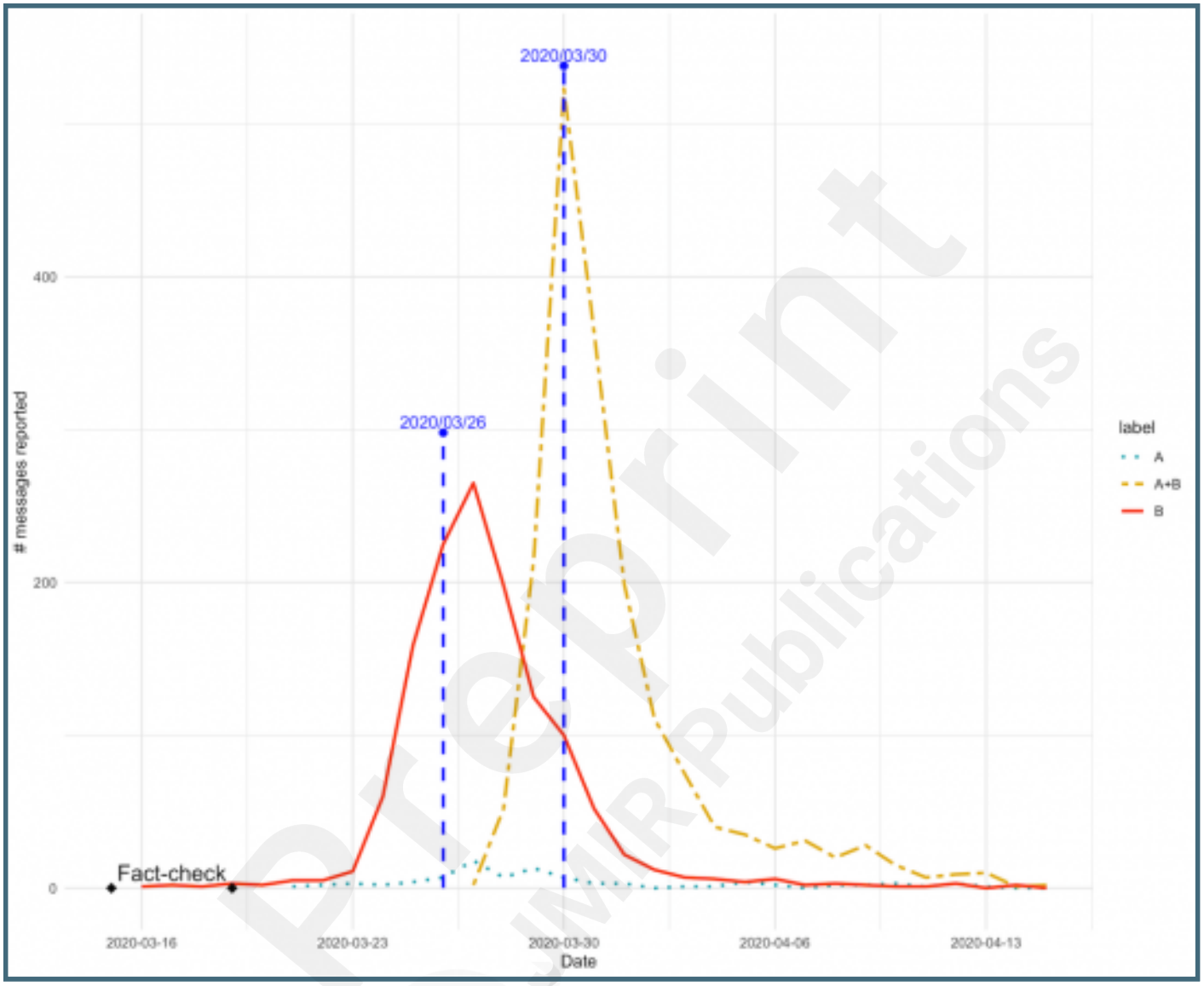
Number of Case 1 messages reported by date. The number peaked on Apr 2 with 205 reports, when messages started misquoting the Director Chen of CECC. There were also a large number of reports after a 4-day long weekend, on Apr 6 with 166 and Apr 7 with 197. Refer to Table 5 for major societal events in blue dashed lines.



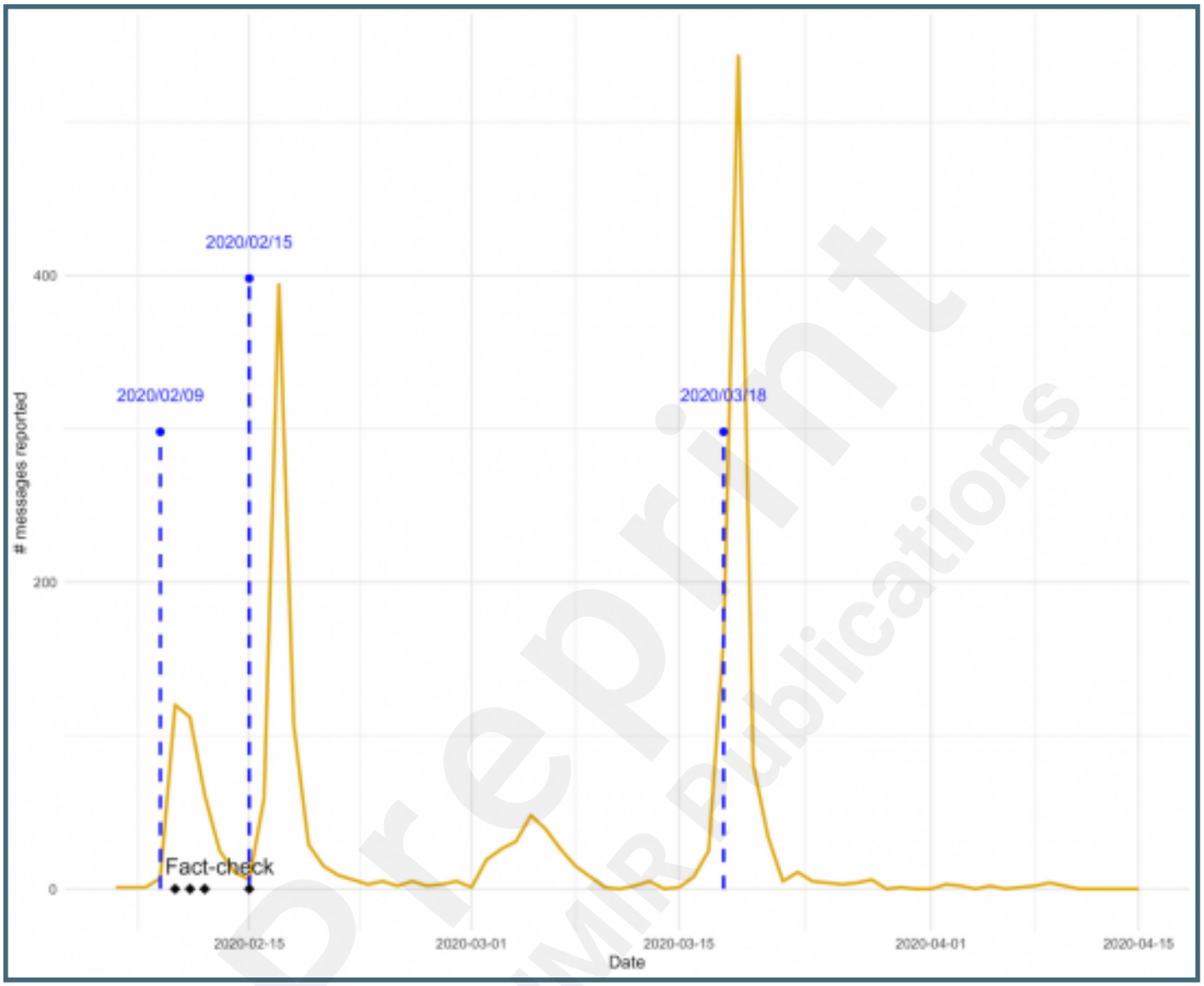
Director Chen, Shih-Chung replaced Zhong, Nan-Shan as the most quoted party in Case 1 rumor after Apr 1, 2020.



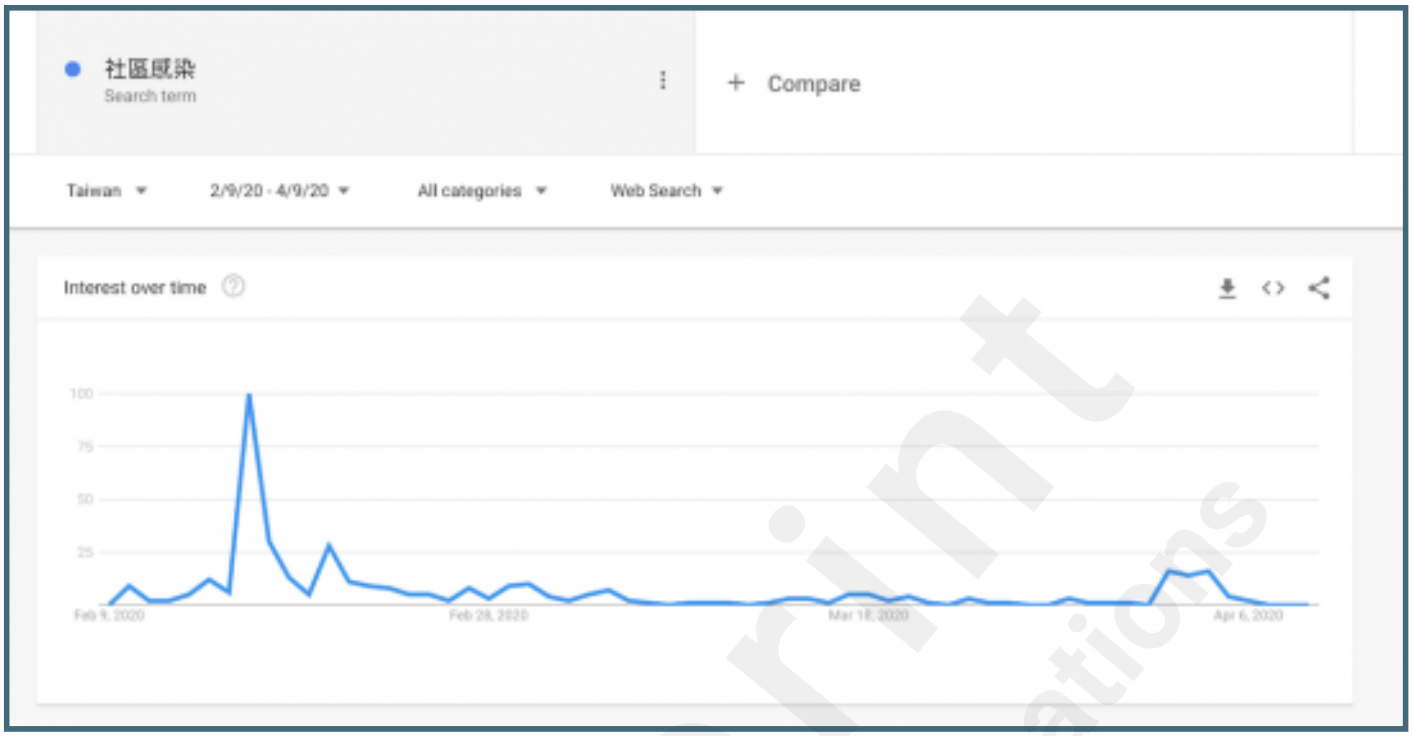
Number of Case 2 messages reported by date. The rumor had been fact-checked rather early, however, the information still received widespread attention. Message (B) peaked on Mar 26 with 225 reports, and the combined message peaked on Mar 30, 2020 with 523. Refer to Table 5 for major societal events in blue dashed lines.



Number of documents of Case 3 reported by date. The higher peaks were on Feb 17 with 394 reports and March 19 with 543 reports. Refer to Table 5 for major societal events.



Google search trend for "community spread" in Taiwan sharply increased during days around Feb 16.



The messages containing authority sharply increased in a week after the rumor first appeared in our dataset, mostly quoting Taiwan Medical Association. From Mid March, at least half of the reported messages quoted Director Chen.