

The Evolution of Rumors on a Closed Platform during COVID-19

Andrea W Wang ^{*1}, Jo-Yu Lan ^{†2}, Chihhao Yu ^{‡3}, and Ming-Hung Wang ^{§4}

^{1,3}Information Operations Research Group (IORG)[¶]

^{2,4}Department of Information Engineering and Computer Science,
Feng Chia University

1 Introduction

Online social media has democratized contents. 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 [24] [14] [8]. 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 [26]. 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 [2] [23].

*wenyi@iorg.tw

†m0907618@mail.fcu.edu.tw

‡chihhao@iorg.tw

§tonymhwang@gmail.com

¶<https://iorg.tw/>

In light of the *infodemic*, several investigations have been carried out to look at the COVID-19 misinformation issue in various aspects. Topics included but not limited to, the types and contents of COVID-19 misinformation [27] [5], the spread and prevalence of rumors on social media platforms [7], [13], [10], [27], [19], [20], the consequences of misinformation [6], and the application of machine learning algorithms on rumor analyses [21] [11]. 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, or LINE, remained extremely scarce. While popular social media platforms are indeed important targets to study online behaviours and expressions, closed platforms remain an integral place to look at, given its more private settings.

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 [17] [18]. According to the survey by Taiwan Communication Survey in 2018, 98.5% of people in Taiwan used LINE as their primary messaging tool, making LINE the most popular instant message platform in Taiwan.¹ We looked into a dataset of 114,124 suspicious messages reported by LINE users in Taiwan between January, 2020 to July, 2020

Secondly, we proposed an efficient algorithm that could cluster a large number of text messages according their topics and narratives 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 cluster is one specific rumor.

Third, by using the results from the algorithm, we were able to look at the dynamics of each particular 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 were 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

¹Data 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 institute aforementioned. The views expressed herein are the authors' own. doi: 10.6141/TW-SRDA-D00176-1

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, the popularity of rumors were more influenced by major societal events.
3. Key authoritative figures were often falsely mentioned or quoted in misinformation, and such practice helped with the popularity of a message.

This paper is organized as followed: we introduced our data in Section 3. Next, we presented the proposed algorithm to cluster text data in Section 4 and subsequently compared the proposed algorithm with other clustering techniques in Section 5.2. Finally we reviewed 3 high-volume COVID-19 false information in Section 5.3. We discussed and concluded this work in Section 6 and 7.

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*.

2 Related Works

From the inception of the pandemic, several survey studies revealed that people relied on social media to gather COVID-19 information and guidelines [15] [16]. 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 [19]. For example, Cinelli et al. 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 [7]. Gallotti et al. looked at the how much unreliable information Twitter users were exposed to across countries. While the level of exposure was country dependant, they revealed that the exposure to unreliable information decreased globally as the pandemic aggravated [10].

Machine learning and deep learning techniques have been used to study the topics and sentiments for COVID-19 misinformation [1]. For example, Jelodar et al. 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 [11]. By applying Structure Topic Model and Walktrap Algorithm, Jo et al. 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 [12]. Furthermore, by employing a multimodal deep neural network for demographic inference and VADER model for sentiment analysis, Zhang et al. 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 [28].

Previous investigations on rumors indicated that individuals are more likely to believe in questionable statements after seeing repeatedly [4] [3], and that rumors became more powerful after being shared multiple times [9]. Most studies only look at the broad topics of misinformation. For example, some looked at reliable versus unreliable information [7] [10] [27], and others employed natural language processing techniques to reduce thousands of social media posts into 10 to 20 groups of topics [1] [11] [12] [7]. Shih et al. instead investigated the content change and temporal diffusion pattern of 17 popular political rumors on twitter [22]. They found that false rumors came back repeatedly, usually becoming more extreme and intense in wordings, while true information did not resurface at all. To the best of our knowledge, there has not been similar study at COVID-19 rumors.

3 Data

In Taiwan, LINE users can voluntarily forward suspicious messages to fact-checking LINE bots such as Cofacts ² or MyGoPen ³. The bots archive the messages and check against their existing databases. If such message has been fact-checked, the bots would reply with the fact-checked results.

We obtained a dataset of 210,221 suspicious messages forwarded by LINE

²<https://cofacts.tw/>

³<https://mygopen.com/>

users to a fact-checking LINE bot between January to July, 2020. The dataset included rumors related to COVID-19 and also some 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 ⁴.
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 distributions is presented in Table 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 identifier 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.

	All	Chinese Characters	Digits	English Alphabets	Others	Number of Tokens
Min	24	24	0	0	0	20
Median	233	145	7	2	38	58
Max	10012	8132	3252	7014	5532	2971

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

⁴<https://github.com/fxsjy/jieba>

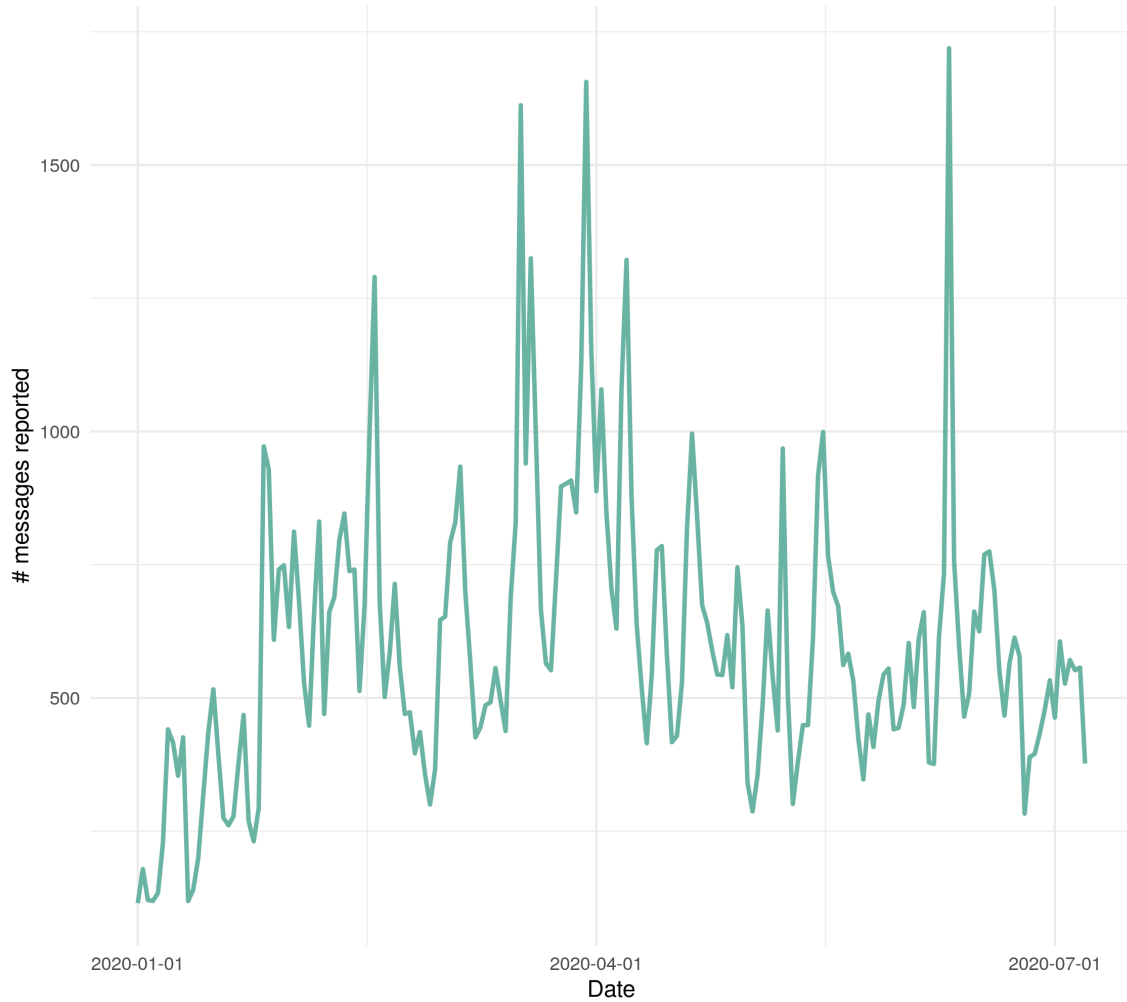


Figure 1: number of suspicious messages reported by date

4 Method

In this section, we described our problem and the proposed clustering algorithm. To follow the terminology of Natural Language Processing, in this section we used *document* to refer to one *message* in our dataset.

4.1 Problem Definition

Given a set of n documents, we would like to group them into m clusters, of which each cluster are made up of documents very similar in usage of terms, only within a limited degrees of text alterations. Intuitively, we wanted the same cluster to have documents that *talked about the same thing in the same way*. 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 and also some 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

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

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

4.2 The Cluster-Classification, "Hybrid", Algorithm

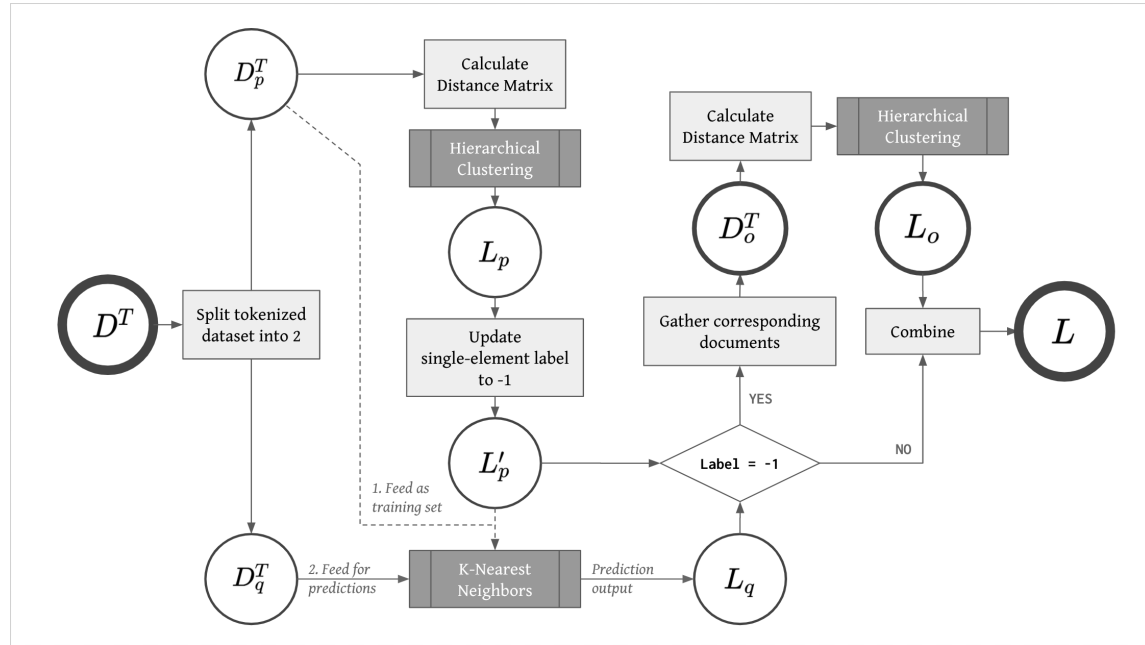


Figure 2: Algorithm flow diagram

Notation

1. $(A)_j$: j^{th} 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. Each element $(D^T)_i$ is the token set of document $(D)_i$.
3. *train portion* p : a number in $(0, 1]$.
4. *distance threshold* λ : a number in $(0, 1]$.

Algorithm

1. Select $p \times |D^T|$ 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 $M_{i,j} = \mathbf{d}((D_p^T)_i, (D_p^T)_j)$ by Formula 1. Note that M is symmetric.
3. Feed M into Hierarchical Clustering with distance threshold of λ . We would get back a sequence of numbers 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 number itself does not carry meaning, manipulate them so they are all non-negative whole numbers.
4. $\forall (L_p)_i \in L_p$, if $|\{l | l = (L_p)_i \forall l \in L_p\}| = 1$, then replace the value of $(L_p)_i$ to -1 . Denote the updated label set as L'_p .
5. Train a K-Nearest Neighbors classifier \mathbf{K} using the training set (D_p^T, L'_p) . And then use \mathbf{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 $(L)_i = \text{Label}((D^T)_i)$.
7. Construct $D_O^T = \{d_i | \text{Label}(d_i) = -1 \forall d_i \in D^T\}$.
8. Redo step 2 and 3 for D_O^T . Denote the resulting sequence 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 .

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.

5 Results

5.1 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.

	mean	std	max	Q_3	Q_2	min
All Groups	6.756	39.190	1082	2	1	1
Groups with at least 2 elements	20.631	70.478	1082	10	3	2

Table 2: Group size statistics

5.2 Model Comparisons

5.2.1 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 combine precision and recall.

5.2.2 Experiments Settings

We compared speed and performances among 4 models:

1. Hierarchical Clustering only (**clustering**). The result from this model is considered to be ground truth.
2. Cluster-Classification Model (**hybrid**). This is our proposed algorithm.
3. Latent Dirichlet Allocation (**LDA**).

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

Throughout the experiments we used distance threshold $\lambda = 0.6$.

Both **LDA** and **pca+kmeans** clustering required a predefined number of groups, which doesn't really fit our purposes. However, for the sake of comparison, we would use the number of groups outputted by **clustering** model as input to both models.

5.2.3 Measuring model performances

Suppose the input is tokenized set of k documents D^T and the **clustering** model put k documents into n groups, (g_1, g_2, \dots, g_n) . g_1 is the group having largest number of documents and g_n the least. Another model M put D^T into m groups: (l_1, l_2, \dots, l_m) . We calculated precision, recall and F-score of model M by the following algorithm:

Algorithm 1: Calculating Precision, Recall, F-score

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

while $c < k/2$ **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 $\leftarrow p/i$;

recall $\leftarrow r/i$;

F-score $\leftarrow f/i$;

In each experiments, 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 confidence intervals.

5.2.4 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 much 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 transformations of a message.

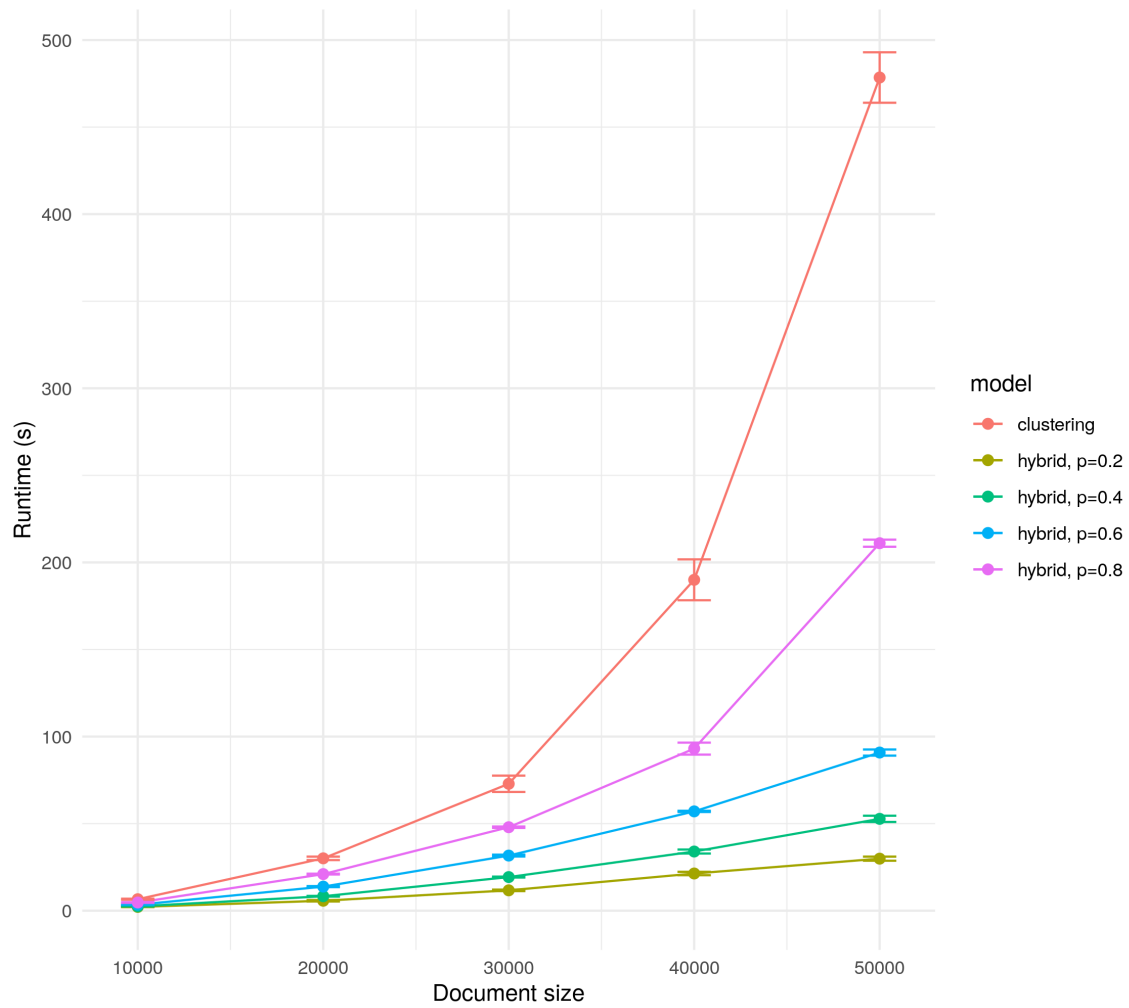


Figure 3: speed comparison between **clustering** and **hybrid** across different levels of p . Using **hybrid** with p lower than 0.6 reduced the runtime from exponential to linear time.

Model	Runtime (s) mean	Precision mean	Recall mean	F-score mean
clustering	6.594	-	-	-
hybrid, $p = 0.2$	2.172	0.993	0.982	0.986
hybrid, $p = 0.4$	2.502	0.995	0.996	0.995
hybrid, $p = 0.6$	3.418	0.997	0.998	0.997
hybrid, $p = 0.8$	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

Table 3: Performance comparison (10,000 documents)

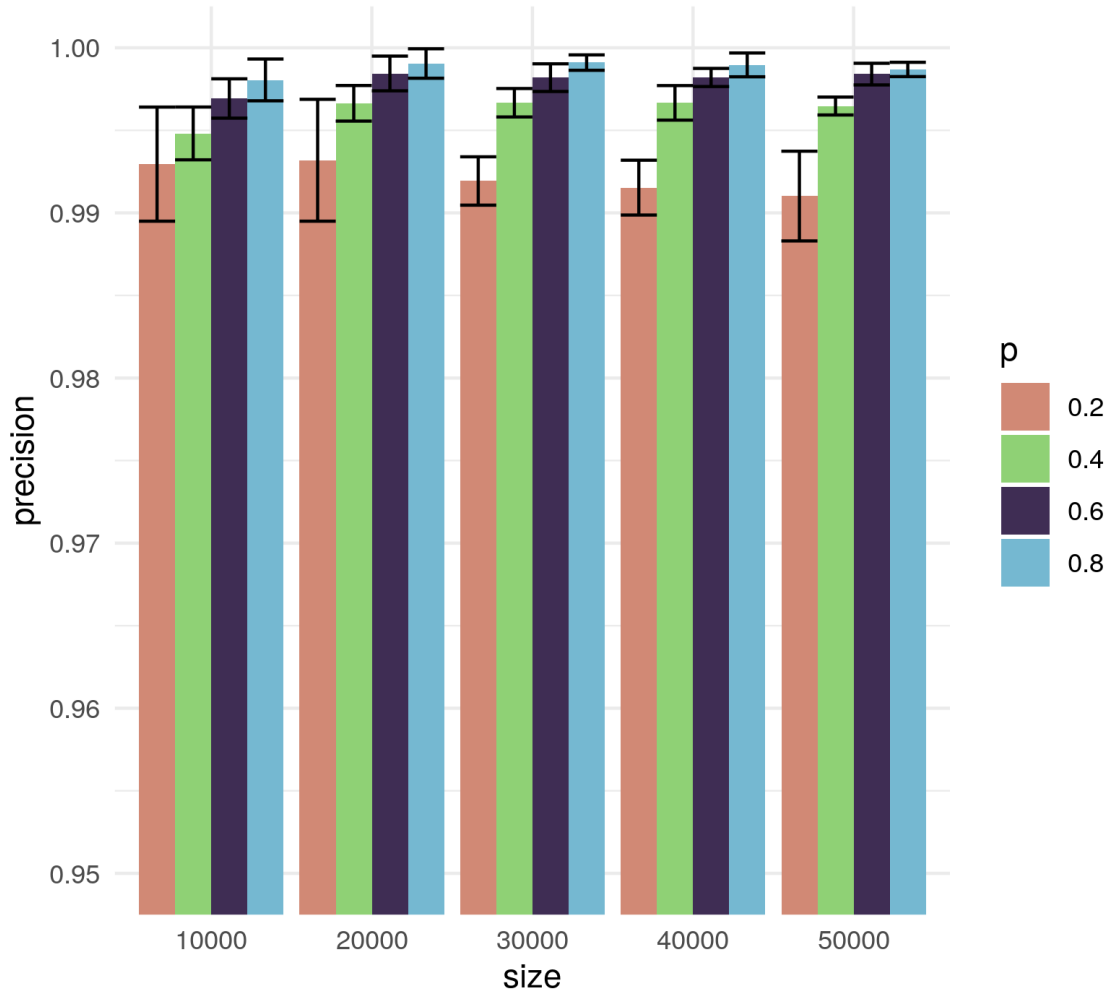


Figure 4: Precision

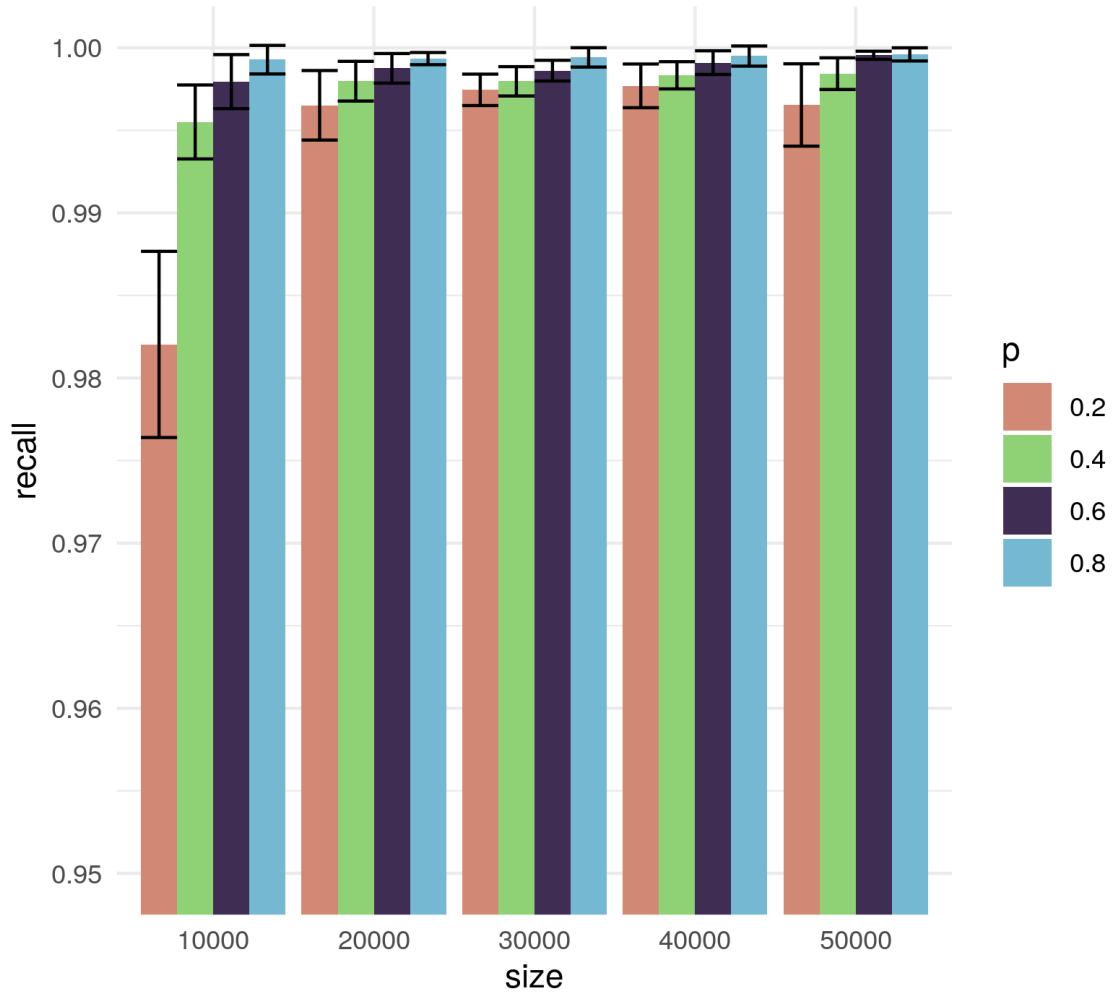


Figure 5: Recall

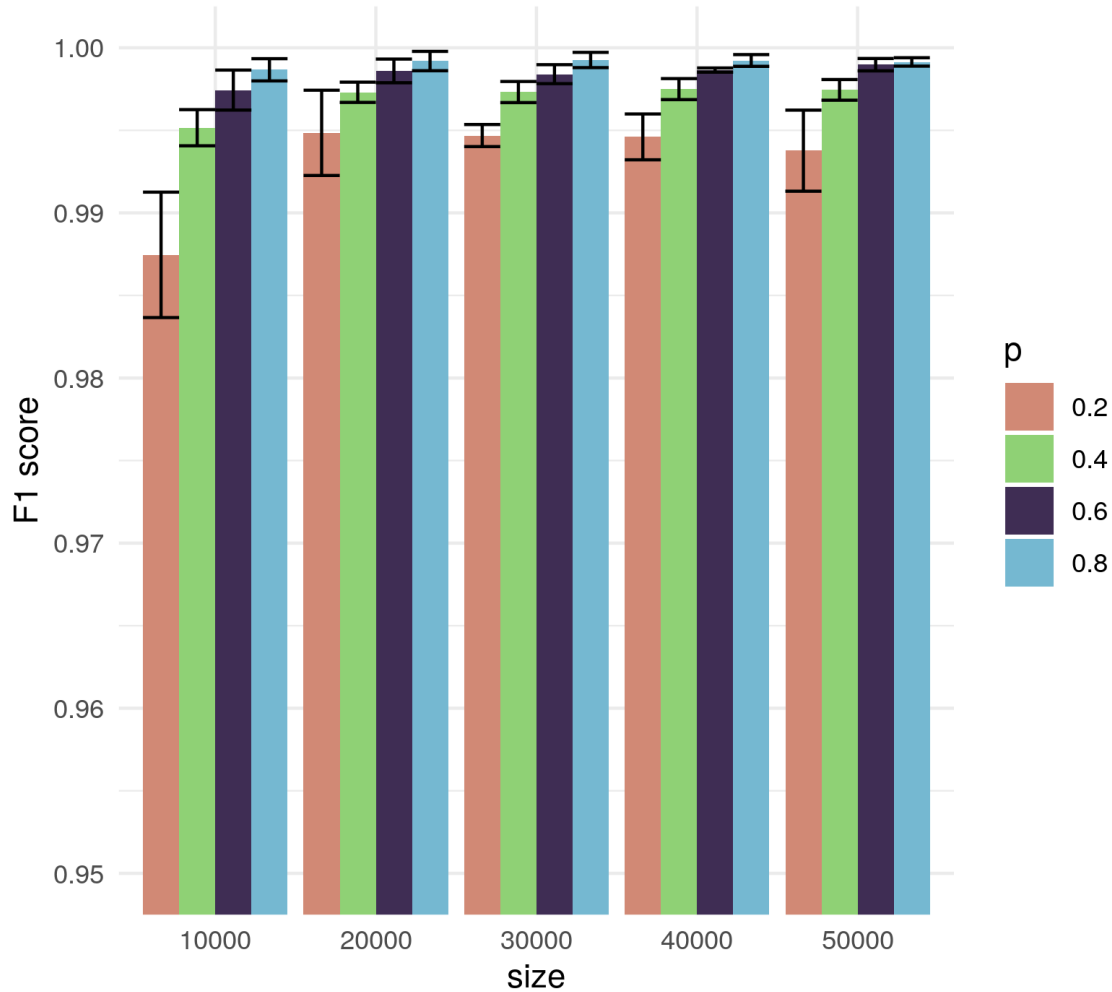


Figure 6: F-Score

5.2.5 Clustering 114K messages using the hybrid method

We used hybrid methods with train portion $p = 0.4$ and distance threshold $\lambda = 0.6$ to cluster the whole set of 114 thousands messages. The 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

	mean	std	max	Q ₃	Q ₂	min
All	9.309	71	2546	2	1	1
Groups with at least 2 elements	28.302	126.907	2546	10	3	2

Table 4: Group Size statistics

5.3 Case Studies

In this section we presented some high-volume suspicious messages related to COVID-19, obtained from the previous section 5.2.5.

5.3.1 Case 1: Do not go outside!

English Translation	Original
<p>Academian Zhong, Nan-Shan emphasized repeatedly, 'Do not go outside! Wait until at least the Lantern Festival to assess the situation of the epidemic.' Be warned that even if you're cured, you would suffer the rest of your life. This is a plague worse than SARS. The side effect of the drugs are more severe...This is a war, not a game ... There is no outsider in this war ...</p>	<p>鐘南山院士再次強調：別出門，元宵後，再看疫情控制情況！警告：一旦染上，就算治癒了，後遺症也會拖累後半生！這場瘟疫比17年前的非典更嚴重，用的藥副作用更大。如果出了特效藥，也只能保命，僅此而已！出門前想想你的家人，別連累家人，能不出門就不出門，大家一起轉發吧！這是一場戰役，不是兒戲，收起你盲目的自信和僥倖心理，也收起你事不關己高高掛起的態度，在這場戰役中沒有局外人！在家！在家！在家！不要點贊！求轉發——鐘南山</p>

Table 5: Case 1 Message Content

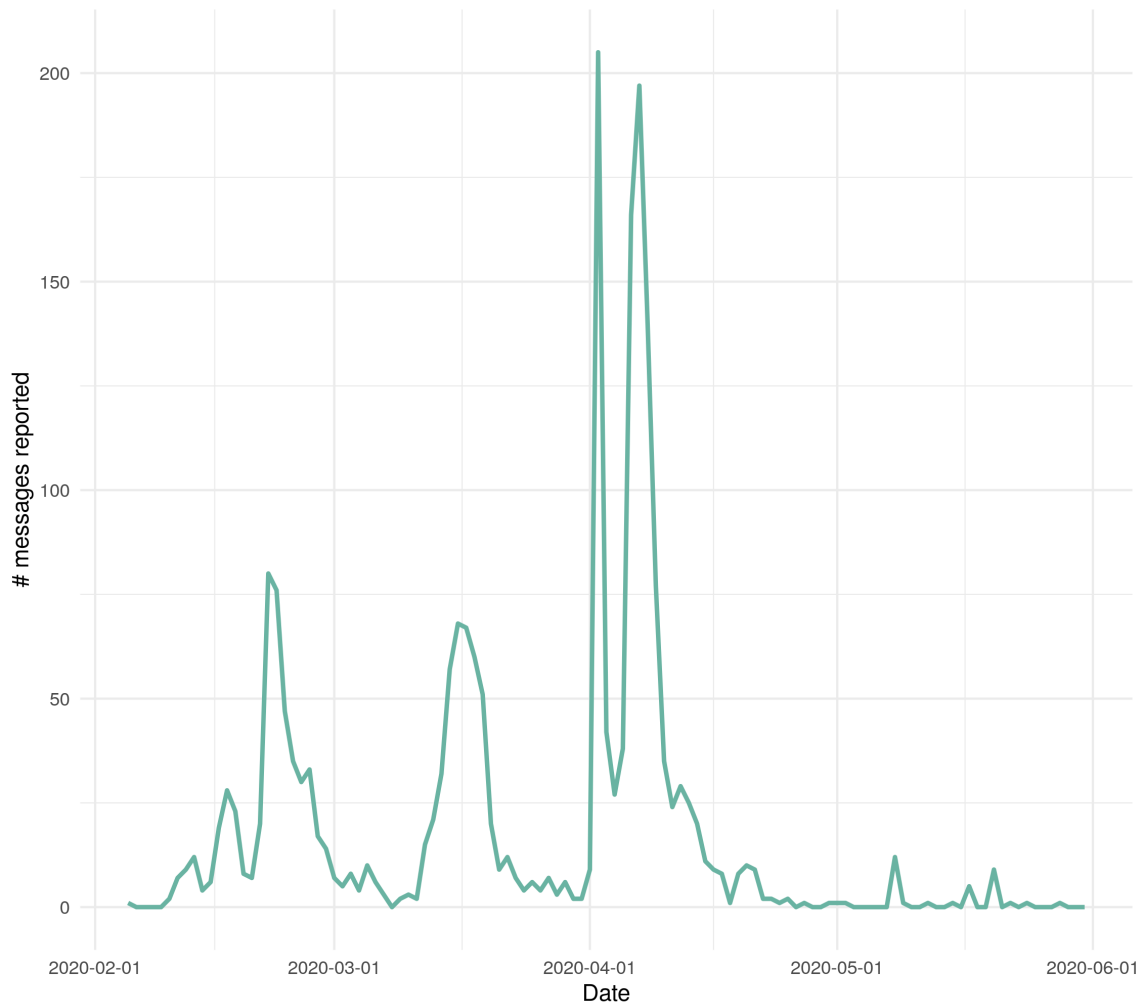


Figure 7: number of documents of Case 1 reported by date. The number peaked on Apr 2nd (205 documents), the day the Ministry of Health and Welfare announced that this was a misinformation. We subsequently saw another peak on Apr 6th (166 documents), the day after a 4-day long weekend.

This case first appeared in the dataset on Feb 2nd, 2020. Over the course of 3 and a half months, there were a total of 2,119 messages reported. The reporting went viral at least four times: it peaked on February 22nd (80 documents), March 16th (68 documents), welcomed the highest peak on Apr 2nd (205 documents), then the last one on Apr 6th with 166 documents. We observed a number of key characteristic changes in the texts itself over the life of this message.

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

Secondly, the efforts were put to emphasize the authoritativeness from whom the message was *quoted*. The first form of this message started with quotation from *The Main-land Academician Zhong, Nan-Shan*, who gained fame during the SARS pandemic in 2003⁵. Other titles, such as "Expert in Pandemic from Mainland China" or "Expert in Coronavirus", were also observed in some transformations. Then later, on Feb 18th, age was first seen in the message: "*Expert in Coronavirus from Mainland China, 78-year-old Academician Zhong, Nan-Shan, emphasized...*". Starting March 10th to March 31st, almost every message included age. Then starting from April 1st, every reported message has Zhong replaced by Chen, Shih-chung. As the Director of Taiwan's Central Epidemic Command Center (CECC), Chen's popularity has skyrocketed during the pandemic through his daily press conference. This was also when we observed the highest peaks of the reported messages.

Due to the prevalence of this message spreading on web and closed platforms, the Ministry of Health and Welfare as well as CECC sent out a press release and a facebook post^{6 7} on April 2nd, reminding the public that this was a false information. 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 tourists 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.

⁵https://en.wikipedia.org/wiki/Zhong_Nanshan

⁶<https://www.mohw.gov.tw/cp-4633-52577-1.html>

⁷<https://www.facebook.com/470265436473213/posts/1524703107696102/>

Date	Previous	New
Feb 17, 2020	Academian Zhong, Nan-Shan stressed again 鍾南山院士再次強調	Pandemic expert from Mainland China , Academian Zhong, Nan-Shan stressed again 大陸防疫專家鍾南山院士再次強調
Feb 18, 2020		Coronavirus expert from Mainland China, 78-year-old Academian Zhong, Nan-Shan stressed again 大陸，冠狀病毒專家鐘南山78院士再次強調
Feb 27, 2020		Coronavirus expert from Mainland China, 84-year-old Academian Zhong, Nan-Shan stressed again 大陸，冠狀病毒專家鐘南山84院士再次強調
Apr 1st, 2020		Director of Taiwan's Ministry of Health and Welfare, Chen, Shih-Chung, reminded everyone 台灣衛福部長陳時中提醒大家
Feb 18, 2020	Do not go outside! Wait until the Lantern Festival to reassess pandemic situation. 別出門，元宵後，再看疫情控制情況	Do not go outside! Wait until March to reassess pandemic situation. 別出門，三月後再看疫情控制情況
		Do not go outside! Wait until the Mid-Autumn Festival to reassess pandemic situation. 別出門，端午節過後，再看疫情控制情況

Table 6: Content Change Log for Case 1

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

In this case we looked at the messages that promoted drinking salt water to prevent the coronavirus. In fact, we investigated two messages and the combination of the them (Table 7).

We first observed Message (B) in our dataset on March 16th. Over the course of its evolution, several medical personnel, such as *Director of The Veteran Hospital* or *Dr. Wang of Tung Hospital* (who, in fact, is an Orthopedist), were misquoted. This showed the use of authoritative power to spread this piece of false medical information. The highest peak was on March 27th, where 265 documents were reported. Around the same time, a small number of Message (A) were also lurking, however, it did not get as much attention as Message (B) before both messages merged into 1 on March 27th and went viral shortly after on March 30th (Orange line in Figure 8). In fact, Message (B) was fact-checked by Taiwan FactChecking Center ⁸ rather early, on March 19th ⁹ and announced it a misinformation, however, this did not stop the piece from misquoting doctors and continued spreading. As a matter of fact, several translations of Message (A+B) were reported in April, including but not limited to English, Indonesian, Filipino and Tibetan. The lifespan of this "drink salted water" message was rather long, as the another famous fact-checking platform in Taiwan, MyGoPen ¹⁰, released an article to disprove this false medical advice again in October 2020 ¹¹, 7 months after it was first seen in our dataset.

⁸<https://tfc-taiwan.org.tw/>

⁹<https://tfc-taiwan.org.tw/articles/3207>

¹⁰<https://mygopen.com/>

¹¹<https://mygopen.com/2020/10/salt-water.html>

English Translation	Original
<p>(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.</p>	<p>這是100%準確的信息... 為什麼中國過去幾天大大減少了感染人數？他們只是簡單地強迫他們的人民每天漱口3次鹽水。完成後，喝水5分鐘。因為該病毒只能在喉嚨中侵襲，然後再侵襲肺部，當受到鹽水侵襲時，該病毒會死亡或從胃中流下來並在胃中銷毀，這是預防冠狀病毒流行的唯一方法。市場上沒有藥品，所以不要購買</p>
<p>(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.</p>	<p>新冠肺炎在還沒有來到肺部之前，它會在喉嚨部位存活4天。在這個時候，人們會開始咳嗽及喉痛。如果他能儘量喝多溫開水及鹽巴或醋，就能消滅病菌。儘快把此訊息轉達一下，因為你會救他人一命！</p>
<p>(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.[...]</p>	<p>為什麼中國大陸過去幾天大大減少了感染人數？除了戴口罩勤洗手外，他們只是簡單地每天漱口3次鹽水。完成後，喝水5分鐘[...] 新冠肺炎在還沒有來到肺部之前，它會在喉嚨部位存活4天[...] 如果他能儘量喝多溫開水及鹽巴或醋，就能消滅病菌[...]</p>

Table 7: Case 2 Message Content

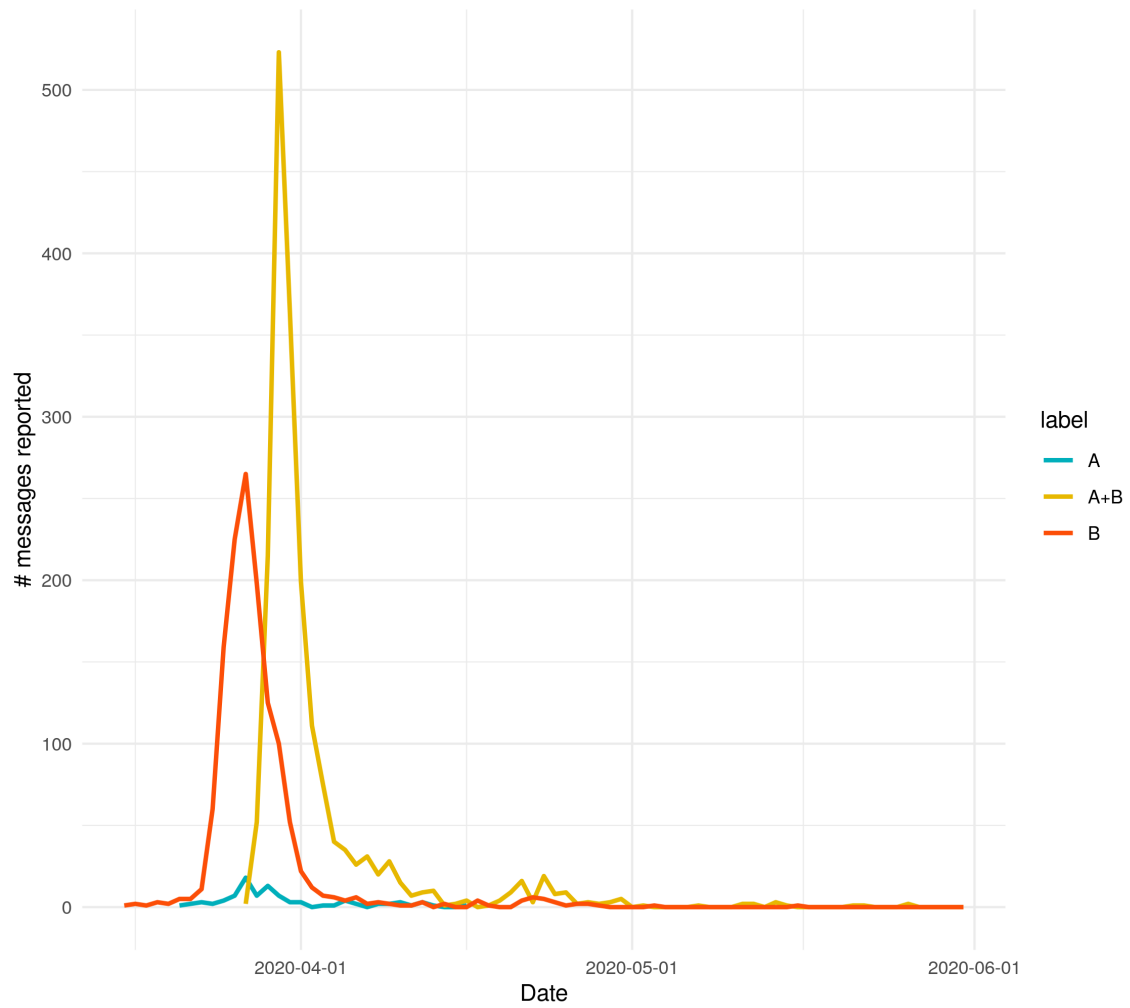


Figure 8: number of documents of Case 2 reported by date.

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

English Translation	Original
<p>10 days from now, Taiwan is in a critical period combating COVID-19. Here are some suggested measures.</p> <ol style="list-style-type: none"> 1. Strictly prohibited going to public places. 2. Choose takeout from restaurants. 3. Eat outside in open spaces. 4. Wash your hands the right way (extremely important). 5. When taking 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 plastic bag for one day before using them. 14. Do not use colleague's phone when working. Disinfect before you have to use one. 15. Avoid taking public transportation during rush hour. 16. Do not visit night market or traditional market. 17. Exercise. 18. Avoid going to the gym. 	<p>今天開始10天，台灣正式進入武漢肺炎關鍵期。建議如下：1.嚴禁進入公共場所。2.用餐儘量將食物外帶。3.用餐環境儘量在外。4.正確方式的洗手(特別重要)。5.坐捷運(公車)，選擇在車前頭。6.避免戴隱形眼鏡7.吃熱食,避開生涼食物,多吃蔬菜8.保持腸胃暢。9.多喝溫水。10.暫停去髮廊。11.穿過的衣服(外套,長褲),回家先單獨吊在外2小時12.暫停戴首飾。13.一有接觸錢幣,一定要洗手,剛拿進來的錢幣,先單獨放在塑膠袋中,一天後,才拿出來。14.在公司不要使用別人的電話筒。電話筒的消毒。15.避開峰時間坐車。16.不去傳統市場及夜市。17.適當的運動。18.暫停進入健身房。</p>

Table 8: Case 3 Message Content

This rumors first appeared in the dataset on February 6th and has a total of 2121 reports in our dataset. Over the 1.5 months of its most popular time, it went viral at least two times: one on February 17th with 394 reports, and on March

19th with 543 reports. It was fact-checked by the Taiwan FactCheck Center on February 15th, 2020¹², however, the fact-check did not avoid the message from getting attention. The content started with authoritative tone that announced "We are at the most critical period of COVID-19", and then provided a list of "do's and don't's". While some *suggestions* made medical sense in terms of hygiene, others didn't¹³. 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 (社區感染) is a phase in a pandemic where many people who tested positive in an area cannot be determined how they got infected¹⁴. 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 February 15th, 2020, Taiwan's Central Epidemic Command Center (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 4 days, four of his family members were also tested positive, forming the first COVID-19 cluster in Taiwan. During that time, people's concerns for community spread was looming. In fact, Google trend for search term "社區感染 (Community Spread)" sharply increased on February 16th (Figure 10). Also, during this period, the number of the reported messages sharply increased (Figure 9).

Content-wise, like what we observed in the first two cases, authorities, especially medical personnel, were used in several versions of the same message to "endorse" the content (Table 9). We spotted a major revision of the message on Feb 12th, 6 days after the first report, where the 18 bullets were pruned to 14, and strong words were modified to gentler tone. Last but not least, the message added a signature of "*Regards from Medical Association*" on the last line. This became the most widespread version afterwards. Out of the 394 documents reported on Feb 17th, 333 documents were of this version. Another key event in content transformation occurred on March 18th. On March 18th, Chen, Shih-chung, the CECC director, went to the Legislative Yuan (similar to Congress in

¹²<https://tfc-taiwan.org.tw/articles/2547>

¹³<https://tfc-taiwan.org.tw/articles/2547>

¹⁴<https://www.cdc.gov/coronavirus/2019-ncov/faq.html#Spread>

the US) to answer interpellation about COVID-19. On the same day, messages started to have "Chen, Shih-Chung explained in the Legislative Yuan on March 18th (3/18陳時中立法院明)" before giving the list of *suggestive measures*. The next day, we saw another sharp increase of reported messages, reaching the highest peak. Of the 543 messages reported on March 19th, 280 has quoted Chen.

Date	Previous	New
Feb. 12, 2020	1. Strictly prohibited going to public places. 1.嚴禁進入公共場所。	1. Reduce going to public places. 1.減少進入公共場所。
	3. Eat outside in open spaces. 5. When taking subway or bus, choose the seats at the first half of the vehicle. 10. Do not visit hair salons. 16. Do not visit night market or traditional market. 3.用餐環境儘量在外。 5.坐捷運(公車), 選擇在車前頭。 10.暫停去髮廊。 16.不去傳統市場及夜市。	<i>deleted</i>
		Regards from Medical Association 醫師全聯會關心您
Mar. 18, 2020	10 days from now, Taiwan is in a critical period combating COVID-19. Here are some suggested measures. 今天起10天, 台灣正式進入武漢肺炎關鍵期, 建議如下	10 days from now, Taiwan is in a critical period combating COVID-19 (Explained by Chen, Shi-Chung in Legislative Yuan on March 18th). Here are some suggested measures. 今天起10天, 台灣正式進入武漢肺炎關鍵期, (3/18陳時中立法院明) 建議如下

Table 9: Content Change Log for Case 3

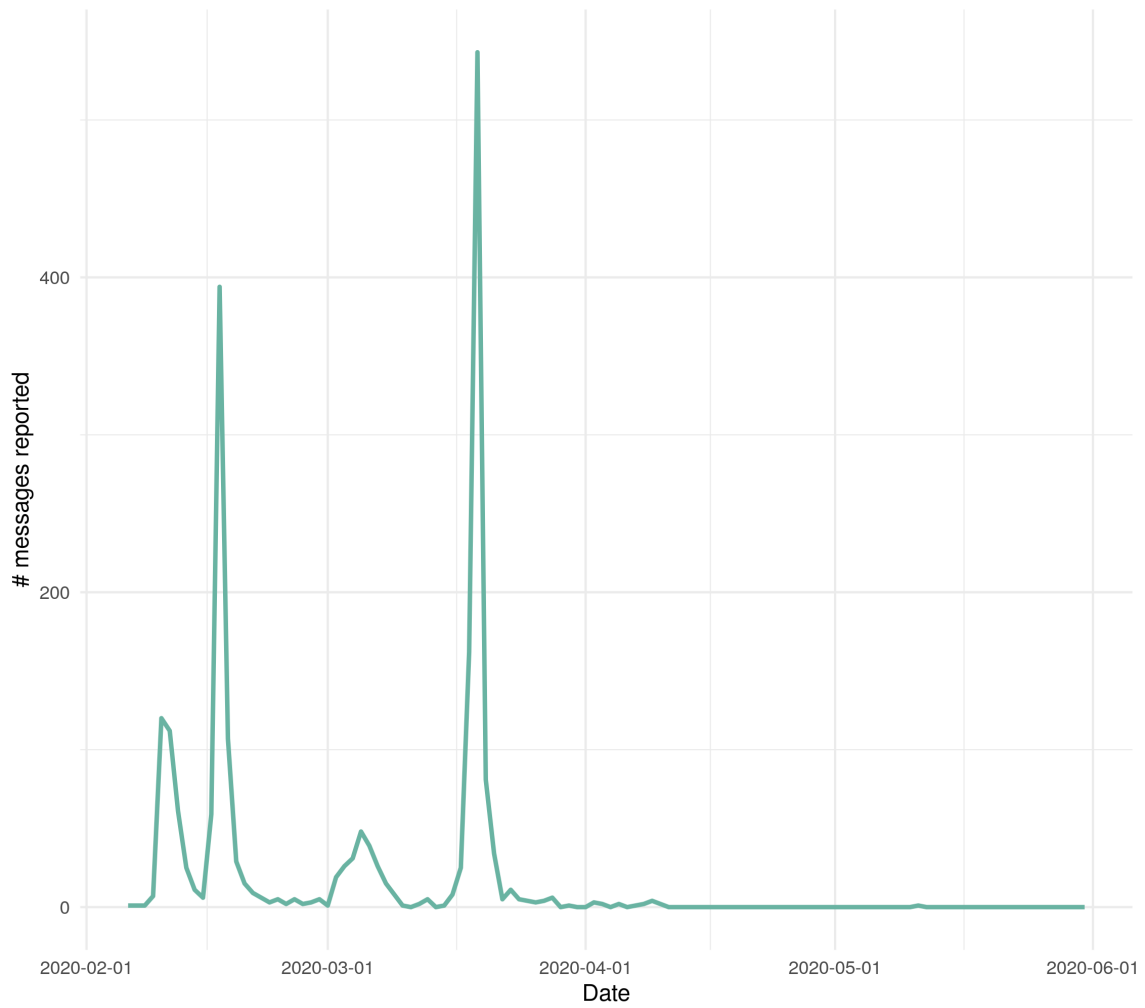


Figure 9: number of documents of Case 3 reported by date. The higher peaks were on Feb 17th with 394 reports and March 19th with 543 reports.

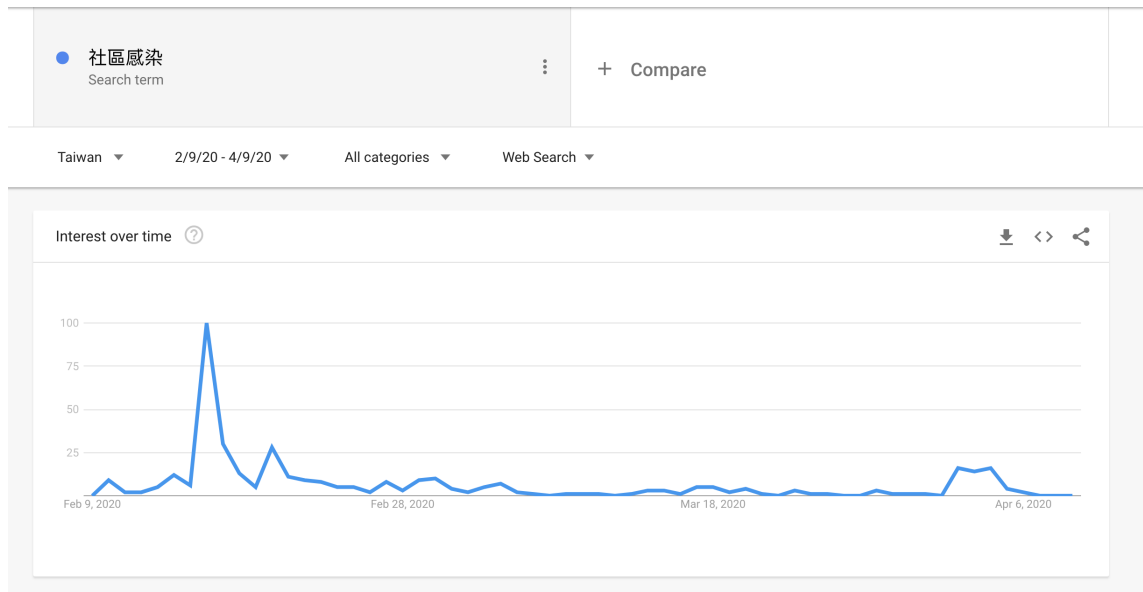


Figure 10: Google Trend of the interest in "community spread (社區感染)" in Taiwan between Feb 9th, 2020 and Apr 9th, 2020. The interest showed a sharp increase from Feb 15th to Feb 16th, where it peaked.

6 Discussion

Similar to the findings of [25], we found that fact-check did not effectively alleviate the spread of false information. The popularity of rumors were more associated with major societal events or content changes. In addition to the above 3 case studies, we went through five other COVID-19 related rumors and manually identified common patterns of textual changes in their propagation. First of all, we observed that key authoritative figures were often (falsely) mentioned or quoted. For example, COVID-19 rumors often included medical-related persons, such as doctors or head of CECC. In addition, 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.* (以下謹提供參考不代表是否正確) was seen in some messages during propagation. Many messages also included simplified Chinese characters or terms that are rarely used in Taiwan. For example, while in Taiwan, people refer to SARS pandemic as "SARS", a large

number of messages use "非典", which is a term more 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 mechanism. 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 a automatic early warning system of misinformation that does not involve fact-check by professionals.

This study had several limitations. First, this data was collected by people'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.

7 Conclusion

In this paper, we analyzed COVID-19 related rumors on a closed-messaging platform, LINE. We proposed a clustering algorithm that reduced the computational time from exponential to linear time. The algorithm enabled us to investigate the evolution of text messages. In fact, the algorithm enabled the research community to perform large-scale studies on the evolution of text messages at message-level rather than topic-level. Similar to what [22] discovered in its study of 17 political rumors, we found that false COVID-19 rumors tend to resurface multiple times even after being fact-checked, and with different degrees of content alterations. Furthermore, the messages often falsely quoted or mentioned authoritative figures, and such practice was helpful for the rumor to reach broader audiences. Also, the resurfacing patterns seemed to be influenced by major societal events and content change. However, each peak of popularity would not last long and it was often without good explanation about how one wave of propagation ended. 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.

References

- [1] Alaa Abd-Alrazaq et al. "Top concerns of tweeters during the COVID-19 pandemic: infoveillance study". In: *Journal of medical Internet research* 22.4 (2020), e19016.
- [2] Amir Abdoli. "Gossip, Rumors, and the COVID-19 Crisis". In: *Disaster Medicine and Public Health Preparedness* 14.4 (2020), e29–e30. DOI: 10.1017/dmp.2020.272.
- [3] Adam J Berinsky. "Rumors and health care reform: Experiments in political misinformation". In: *British journal of political science* 47.2 (2017), pp. 241–262.
- [4] Lawrence E Boehm. "The validity effect: A search for mediating variables". In: *Personality and Social Psychology Bulletin* 20.3 (1994), pp. 285–293.
- [5] J Scott Brennen et al. "Types, sources, and claims of COVID-19 misinformation". In: *Reuters Institute* 7 (2020), pp. 3–1.
- [6] Aengus Bridgman et al. "The causes and consequences of COVID-19 misperceptions: Understanding the role of news and social media". en-US. In: *Harvard Kennedy School Misinformation Review* 1.3 (June 2020). DOI: 10.37016/mr-2020-028. URL: <https://misinforeview.hks.harvard.edu/article/the-causes-and-consequences-of-covid-19-misperceptions-understanding-the-role-of-news-and-social-media/> (visited on 03/10/2021).
- [7] Matteo Cinelli et al. "The covid-19 social media infodemic". In: *Scientific Reports* 10.1 (2020), pp. 1–10.
- [8] Michela Del Vicario et al. "The spreading of misinformation online". In: *Proceedings of the National Academy of Sciences* 113.3 (2016), pp. 554–559.
- [9] Nicholas DiFonzo and Prashant Bordia. *Rumor psychology: Social and organizational approaches*. American Psychological Association, 2007.

- [10] Riccardo Gallotti et al. "Assessing the risks of 'infodemics' in response to COVID-19 epidemics". In: *Nature Human Behaviour* 4.12 (2020), pp. 1285–1293.
- [11] Hamed Jelodar et al. "Deep sentiment classification and topic discovery on novel coronavirus or covid-19 online discussions: Nlp using lstm recurrent neural network approach". In: *IEEE Journal of Biomedical and Health Informatics* 24.10 (2020), pp. 2733–2742.
- [12] Wonkwang Jo et al. "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". In: *Journal of medical Internet research* 22.6 (2020), e19455.
- [13] Ramez Kouzy et al. "Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter". In: *Cureus* 12.3 (). ISSN: 2168-8184. DOI: 10.7759/cureus.7255. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7152572/> (visited on 03/10/2021).
- [14] David MJ Lazer et al. "The science of fake news". In: *Science* 359.6380 (2018), pp. 1094–1096.
- [15] Norazryana Mat Dawi et al. "Attitude Toward Protective Behavior Engagement During COVID-19 Pandemic in Malaysia: The Role of E-government and Social Media". In: *Frontiers in Public Health* 9 (2021), p. 113.
- [16] Syed Muhammad Mubeen et al. "Knowledge and awareness regarding spread and prevention of COVID-19 among the young adults of Karachi". In: *J Pak Med Assoc* 70.5 (2020), S169–74.
- [17] Hui Xian Lynnette Ng and Jia Yuan Loke. "Analysing Public Opinion and Misinformation in a COVID-19 Telegram Group Chat". In: *IEEE Internet Computing* (2020).
- [18] Bastani P and Bahrami Ma. "COVID-19 Related Misinformation on Social Media: A Qualitative Study from Iran." English. In: *Journal of Medical Internet Research* (Apr. 2020). ISSN: 1439-4456, 1438-8871. DOI: 10.2196/18932. URL: <https://europepmc.org/article/med/32250961> (visited on 03/10/2021).

- [19] Cristina M Pulido et al. "COVID-19 infodemic: More retweets for science-based information on coronavirus than for false information". In: *International Sociology* 35.4 (2020), pp. 377–392. DOI: 10.1177/0268580920914755. eprint: <https://doi.org/10.1177/0268580920914755>. URL: <https://doi.org/10.1177/0268580920914755>.
- [20] Gautam Kishore Shahi, Anne Dirkson, and Tim A. Majchrzak. "An exploratory study of COVID-19 misinformation on Twitter". en. In: *Online Social Networks and Media* 22 (Mar. 2021), p. 100104. ISSN: 2468-6964. DOI: 10.1016/j.osnem.2020.100104. URL: <https://www.sciencedirect.com/science/article/pii/S2468696420300458> (visited on 03/10/2021).
- [21] A. Shi et al. "Rumor Detection of COVID-19 Pandemic on Online Social Networks". In: *2020 IEEE/ACM Symposium on Edge Computing (SEC)*. Nov. 2020, pp. 376–381. DOI: 10.1109/SEC50012.2020.00055.
- [22] Jieun Shin et al. "The diffusion of misinformation on social media: Temporal pattern, message, and source". In: *Computers in Human Behavior* 83 (2018), pp. 278–287.
- [23] Samia Tasnim, Md Mahbub Hossain, and Hoimonty Mazumder. "Impact of rumors and misinformation on COVID-19 in social media". In: *Journal of preventive medicine and public health* 53.3 (2020), pp. 171–174.
- [24] Soroush Vosoughi, Deb Roy, and Sinan Aral. "The spread of true and false news online". In: *Science* 359.6380 (2018), pp. 1146–1151.
- [25] Thomas Wood and Ethan Porter. "The elusive backfire effect: Mass attitudes' steadfast factual adherence". In: *Political Behavior* 41.1 (2019), pp. 135–163.
- [26] World Health Organization. *Novel Coronavirus (2019-nCoV). Situation Report 13, 2 Feb 2020*. 2020. URL: <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-ncov-v3.pdf> (visited on 03/16/2021).
- [27] Kai-Cheng Yang, Christopher Torres-Lugo, and Filippo Menczer. "Prevalence of low-credibility information on twitter during the covid-19 outbreak". In: *arXiv preprint arXiv:2004.14484* (2020).

- [28] Chunyan Zhang et al. "Understanding Concerns, Sentiments, and Disparities Among Population Groups During the COVID-19 Pandemic Via Twitter Data Mining: Large-scale Cross-sectional Study". In: *Journal of medical Internet research* 23.3 (2021), e26482.