

COVID 19 Misinformation / Fake News Records

Thi Tran

4/29/2020

Background Information

- Coronavirus Pandemic is a global healthcare crisis originated from Wuhan, China in December 2019 and then exponentially widespread to various countries around the world in the early to middle 2020. Reported deaths increased from 26 at the end of January to around 3,000 at the beginning of March and then dramatically expanded to more than 200,000 deaths at the end of April 2020. Daily confirmed new cases jumped from around 1,800 cases per day on March 1 to around 100,000 cases per day, and total reported cases worldwide are at more than 3 million cases at the end of April, 2020. Many countries have tried various efforts including stay-at-home orders, economic lockdown, suspended airports and transportations or mandatory prevention methods applied partially or entirely to their citizens with a lot of consequences.
- In that situation, when people are panic and confused with unusual situation, they seek for as much information as possible to save themselves and their families, friends and possibly the community. Hence, it creates an ideal environment for misinformation or fake news to bloom and widespread. Fake news are spread sometimes much faster than the actual pandemic.
- Defined as misleading information, misinformation might have the motives of ethical intentions to inform the community and minimize harms in such a crisis but does not have enough time and efforts to confirm the legitimacy of contents, or have the motives of bad intentions such as gaining benefits for the information creators or simply causing unnecessary panics, worrisome, confusion and sufferings in the community of victims. Regardless the good or bad intention, fake news or misinformation messages especially in contexts of crises can create significant harmful consequences or damages toward targetted victims in various ways, as some researchers already pointed out such as physical harms, financial harms, psychological harms, social harms, reputational harms, safety harms, privacy harms, and confusion harms (Thi Tran et al., 2019 conference papers).
- It is necessary that we analyze characteristics of fake news or misinformation messages to help the efforts of recognizing fake news, eliminating or limiting the spread of fake news, detecting targetted communities or quantifying possible harms from those fake news.
- This research is therefore built to advance knowledge from current efforts regarding facing COVID 19 or coronavirus pandemic misinformation.
- Some examples of Coronavirus - COVID fake news or misinformation messages:
 - This is a list of fake news with extracted features and sources (https://github.com/thitran82/COVID19_FakeNews_Visualization/blob/master/Coronavirus%20fake%20news%201.png).
 - This is a list of summarized fake news messages (https://github.com/thitran82/COVID19_FakeNews_Visualization/blob/master/Coronavirus%20fake%20news%202.png).

Introducing the dataset:

The data used in this research project was obtained from a publicly revealed dataset from researchers of Princeton University through this link (<https://esoc.princeton.edu/files/covid-019-disinformation-data>) File Type: Other Publications File extension: xls Publication Year: 2020 Author: Jacob N. Shapiro, Jan Oledan

The dataset captured various features from more than 620 fake news or misinformation messages or claims related to coronavirus or COVID 19 pandemic.

Considered Research Questions:

#Question group 1: Changes of records in different categories through the time:

- **Q1. Changes in numbers of total fake news posts in different languages (presented by primary languages) through the time.**
- **Q2. Changes in numbers of total fake news posts in different categories of main narratives through the time.**
- **Q3. Changes in numbers of total fake news posts in different motives through the time.**
- **Q4. Changes in numbers of total fake news posts in different sources through the time (individual actors, state sponsors, media, companies, political actors, and some combined types).**
- **Q5. Changes in numbers of total fake news posts in different misinformation types through the time.**

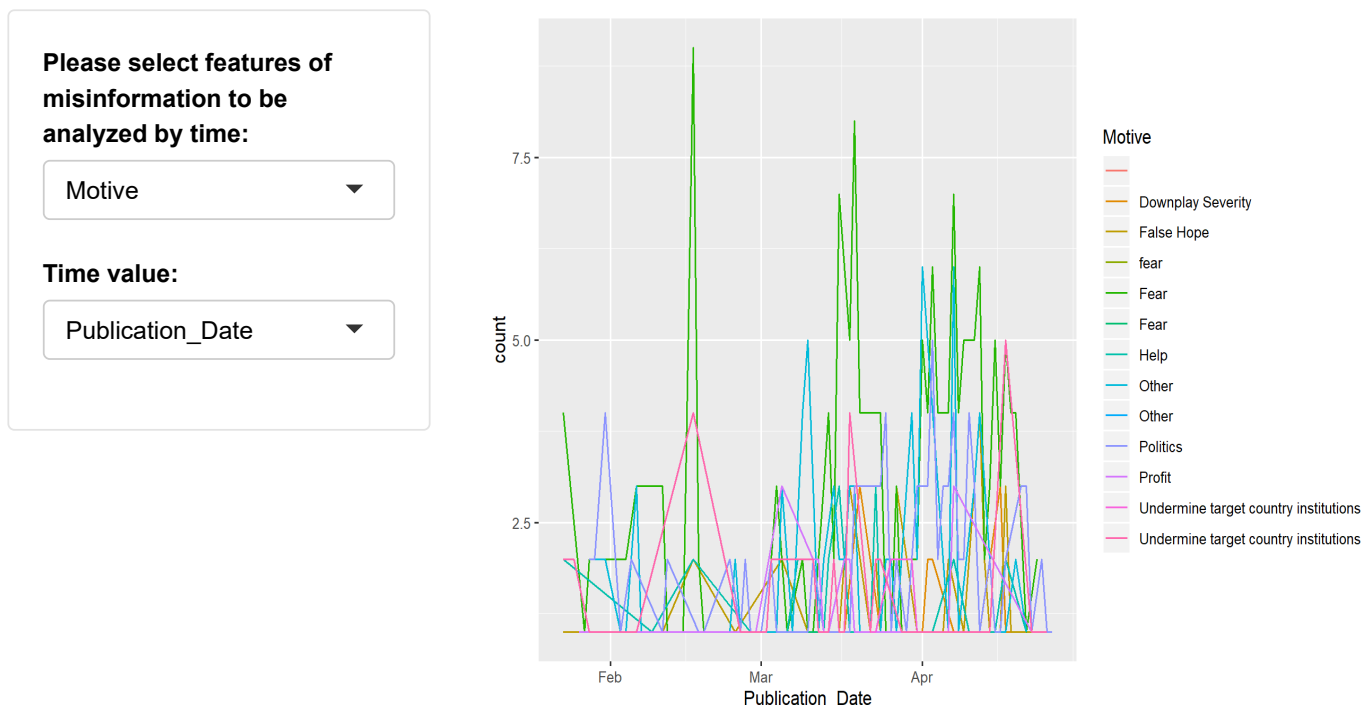
#Question group 2: Changes of records in different categories through the time:

- **Q6. Changes in numbers of total fake news posts in groups of motives based on different languages (presented by primary languages).**
- **Q7. Changes in numbers of total fake news posts in groups of motives based on different categories of main narratives.**
- **Q8. Changes in numbers of total fake news posts in groups of motives based on different sources (individual actors, state sponsors, media, companies, political actors, and some combined types).**
- **Q5. Changes in numbers of total fake news posts in different misinformation types through the time.**

Theoretical background

- Number of fake news records will change through the timeline of developments of the coronavirus outbreaks.
- Both numbers and relatively comparable proportions of misinformation records based on certain categories such as motives (intentions of misinformation creators) or narratives (covered main topics of those messages) through the timeline of the pandemic developments. We want to examine those differences and timely changes.

Launching the Shiny App for interactive visualization:



Discussion:

#Question group 1: Changes of records in different categories through the time:

Research question 1: Changes in numbers of total fake news posts in different languages (presented by primary languages) through the time. Trends fluctuated a lot. Records generally increased at the end of February and then significantly decreased in April. English and Mandarin dominated others in spreading fake news. German and Japanese also contributed a lot in fake news diffusion through time.

Research question 2: Changes in numbers of total fake news posts in different categories of main narratives through the time. Records gradually increased from January to the beginning of March then sharply increased until the beginning of April before falling down significantly. Misinformation about “false cures and preventative methods” dominated others until the end of March but then sharply decreased. 2 groups of fake news related to “emergency responses” and “COVID-19 status of individuals” got highest values compared to others since the beginning of April even though they already fell down sharply. Misinformation about the nature of the virus also accounted for large proportions in April.

Research question 3. Changes in numbers of total fake news posts in different motives through the time. Although the numbers fluctuated a lot, generally COVID 19 fake news reached the peak and stayed at high values in the entire March before falling down in April, similar to the status of other features. As we can expect, fake news that has the motive to create fear and based on politics dominated the situation all the time, while fake news about help is one of the lowest. Interestingly, misinformation about getting profits was one of the lowest proportions, indicating that financial intention is not a big encouragement to spread fake news compared to others in this COVID pandemic.

Research question 4. Changes in numbers of total fake news posts in different sources through the time (individual actors, state sponsors, media, companies, political actors, and some combined types). As we can see from the plot, it is interesting that individual actors accounted for most of the sources of misinformation, sometimes around 90%, while other sources, including all media channels, accounted for much less amounts. That can be the results of popular and extremely active social networks with billions of active users that constantly create and spread fake news. Media and other sources can generate the fake news, but social network users are the people who quickly and widely misinterpret and spread those misleading messages.

Q5. Changes in numbers of total fake news posts in different misinformation types through the time. Similar to others, records based on types of misinformation also significantly increased in March before falling down in April. As we can see from the graph, similar to Research Question 4, the only dominating type of fake news is false reporting, while the other 2 big ones as conspiracy and fake remedy just got much lower proportions, and others got nearly zero values.

#Question group 2: Changes of records in different categories through the time:

Q6. Changes in numbers of total fake news posts in groups of motives based on different languages (presented by primary languages). Generally, we can see that the overall records of motives have the highest value falls into the category Fear, followed by politics. Regarding the languages, similar to research question 1, Mandarin and English dominates others. Both Mandarin and English reached the highest percentages in the motive Fear, which is also the highest category.

*Q7. Changes in numbers of total fake news posts in groups of motives based on different categories of main narratives.** Similar to research question 2, Emergency response and false cures / preventative measures lead the proportions of fake news in most of the motives, except for only one value of the narrative COVID 19 status of individuals, which got the highest share in motive Fear, the biggest influencer of Motive.

Q8. Changes in numbers of total fake news posts in groups of motives based on different sources (individual actors, state sponsors, media, companies, political actors, and some combined types). Similar to Research Question 4, we can see the source “Individual actors” got highest values in nearly all categories of misinformation regarding motives, which the biggest proportion around 90% was also in the biggest contributor of motives: Fear. Contributions of politics varied a lot, and nearly disappeared in some motives such as false hope, profit, severity, or fear.

Q9. Changes in numbers of total fake news posts in different misinformation types through the time. As mentioned in Research question 5, false reporting was the type that had most misinformation records. In detail, false reporting got highest shares in all the 2 top highest motives: fear and politics. Fake remedy and conspiracy also got significantly high percentages, but they got highest values in some small motives such as false hope, help, profit or “undermine target country institution”.

Conclusion:

In such a large scale and high impact humanitarian crisis like COVID-19 pandemic, people got flooded with numerous information messages from different sources. Lack of time to confirm and with motivation of trying to save themselves or others, they can quickly become victims of misinformation and its harmful effects, especially from the popularity and dynamic natures of social

media and social networks. This research used a publicly revealed dataset that is obtained day by day during this ongoing COVID 19 pandemic, and with various aspects or features. By considering fake news records grouped by certain features through the time, including primary language of the messages, main narratives (or the topics), motives, sources, misinformation types, and from those features toward a chosen feature Motive, this study revealed interesting findings about changes of fake news diffusions through time, as well as their proportions based on different motivations or motives of misinformation creators. This study can shed lights to help other future researchers define research directions, raise some suspected constructs for consideration, or reveal interesting unique patterns that this unusual era of COVID 19 pandemic is exposing toward worldwide communities. Seriously considering those points can help researchers, authorities, involved stakeholders, first responders and especially the potential victims figure out best guidelines or reaction plans in order to minimize or eliminate harms from such fake news or misinformation messages.