Routes of communication: understanding the information flows that shape public and official reaction to COVID-19

1 Project overview

During disease outbreaks of international concern, like SARS, pH1N1, WAEO **((jd: fill in better names with dates))** or the current COVID-19 outbreak, this process is both compressed and amplified. The stakes also become higher, because public behaviour directly affects the spread of infectious disease: people who avoid large gatherings may slow disease spread, while people who flee infected areas may accelerate it, for example. Excessive fear of disease spread can have severe economic effects, and may also lead to bias and discrimination against groups seen as linked to the disease, or merely seen as others.

The proposed research will study how information flows between forums, including scientific publications; governmental policies and agency recommendations; and mass and social media – and investigate how it affects public perceptions and behaviours. Our interdisciplinary group will analyze these flows by combining textual and contextual analysis; AI-assisted human-supervised data mining; time-series analysis; and mathematical modeling. We will study how good information competes with misinformation, and look for factors correlated with successful spread of good information. We will gather information on communication and surrogates for behaviour from a wide range of sources. Sources about communication will include agency websites; preprint servers and publicly available scientific journals; major newspaper websites; social-media platforms; and twitter data. Surrogates for public perceptions and behaviour will include twitter data (again); google trends; publicly available box-office information for movies and major sporting leagues; and information about cancellations and shortages (for example of face masks or pharmaceutical or pseudo-pharmaceutical products) from our textual analysis.

The project will be organized around three Research Questions:

RQ1 How does information (and misinformation) travel between scientists, public-health workers, mass media and social media? (Research sub-area "cultural dimensions of the epidemic")

RQ2 How does communication affect public behaviour and the course of the outbreak? (Research sub-area "public health response") **((jd: improve this quote))**

RQ3 How can scientists and policy-makers evaluate and improve the effectiveness of their communication? (Research sub-area "strategies to combat misinformation")

2 Background

Public-health communication is a balancing act. Officials are often caught between the need to be heard, and the danger of causing panic. This problem is particularly acute in the case of an infectious outbreak, since the presence of a novel pathogen increases both the importance of being heard and the danger that the public will over-react.

In the case of COVID-19, scientists are still scrambling to understand the pathogen's biology; public-health workers are scrambling to decide on the best recommendations and

policy decisions given current knowledge at any given time; and the mass media is scrambling to understand the situation and decide how best to communicate with the public.

There are other complicating factors. An outbreak of global concern represents an opportunity for mainstream and peripheral media, and for social-media actors to increase their "clicks" and "likes" and therefore prestige and/or profitability. These motivations work against the balancing act, and instead favor over-simplification and sensationalization.

It's known that traditional media can have strong influence on public perceptions, creating fear by overestimating risk during the SARS outbreak [**?**] and the influenza pandemic [**?**], or feeding into bias – e.g., anti-Chinese bias during the SARS crisis [**?**]. Media is also the tool public health authorities rely on to promote their concerns and recommendations during health risk outbreak. Understanding media effect on disease spread (e.g., media attention increases self-protection) can help enhance epidemic forecasting and preventive measures to slow the disease spread [**?**]. **((jd: Moved doi to auto.rmu. CHYUN delete this or lmk I can delete it.))**

While traditional news media (including online presence) remains influential, social media plays an increasing role in shaping how we communicate and understand information [**?**]. Social media can play a positive role spreading good information, [**?, ?, ?**], but may also spread misinformation and feed bias [**?, ?**]. Since the initial reports of cluster of acute severe respiratory disease (COVID-19) and the potential for global spread, there has been widespread discussion and dissemination of information through social media [**?**]. **((jd: The rest of this ¶ was long and not sharp. What do we need from [?]? Is [?] really from 2010? Is that who we should cite about how twitter works?))**

((cfs: we need a few setences about policy makers and google trend impact and roles to tie this all together))

((jd: We need to talk about how information spreads from scientists and policy makers to social media (both directly and indirectly). Not sure we need google trends here; we can talk about it nicely in metrics, I hope.))

((mli: maybe make a plot using google trend for the keyword search "coronavirus" and HuBei time series))

((jd: Shoudl we try to bring in factual themes? The key ones are risk of worldwide spread; case-fatality proportion; something about control strategy effectiveness. The fuzziness of the last one is why I haven't tried to do this yet.))

3   Methods and feasibility

3.1   Data and Sources

Science   We will develop systematic search and screening strategies to extract relevant peer-reviewed publications from Google Scholar and PubMed. To account for the strong influence of preprints early in the epidemic [**?**], we will will also include preliminary scientific findings posted on medRxiv through 31 March 20. We will index these papers and track their appearances in mass media and social media; we will also track which of the preprints are published after peer review.

Public health recommendations   We will collect and analyze reports, guidelines and recommendations available from the World Health Organization and from the central disease

control agency of each of our focal countries. It is worth noting that all of these agencies have launched special COVID-19 pages.

Mass media   We will use the Lexis-Nexis search engine (via McMaster University) and OriProbe Information Services to collect articles relevant to the outbreak, going back to the outbreak start in December, and continuing throughout the grant period. We will focus on the top English- and Mandarin-language newspapers (taking both circulation and online access into account) from Canada, China, England, Singapore, Taiwan, and USA. We will include the top Mandarin-language newspapers in both Canada and the USA.

Social media   We will efficiently collect data from twitter by purchasing API access. Twitter analyses will start in November 2020, to give a baseline for comparison. **((jd: Xingpeng: Need words on how to collect Weibo. Should we pay for access there?))**

Public response   Twitter (and Weibo) data will give us information not only on information flows,but also on public interest and attitudes. We will also probe public interest and concern using publicly available data from GoogleTrends, which tabulates frequency of searches (by search times and topics) in various regions across the world [**?**, **?**]. Economic data relating to the outcome of the COVID-19 outbreaks will be gathered as reference for contextual analysis (see below): for example, travel, movies box offices, cancellation of public events.

3.2   Analysis of Data
Data Mining

AI-assisted human-supervised data mining   Textual analysis of news media and social media (i.e., Twitter and Weibo) will mainly based based on a coding book and implemented by AI. The coding book should list themes and frames for the analysis, which will be categorized based on a pilot study by trained coders supervised by the team after the proposal is funded. The basic structure of the analysis will outline theme, information types (e.g., news, information and misinformation based on sources verification), who, where, theme tone and when. (see attached [Basic Coding frame](https://github.com/cfshi/coronavirus/blob/master/coding((need concrete ideas on how AI will analyze tweets and weibo))

Time-series analysis   Google trend, reported daily reports, patterns with policy change/case definitions. (CS: A couple of sentence on how we can hack it)

Contextual analysis   Contextual analysis will be studied by the researchers. Media content and information are often socially and politically cultivated []. We will implement contextual analyses of the quantitative findings and elaborate how information is framed and disseminated in this uncertain circumstances into context. Textual analysis using AI is reliable and efficient in terms of data mining. However, it can be insufficient in understanding what contents connote. For example, meanings signified in the public reactions to shortage of face masks in Toronto is likely different from the areas where are heavily impacted with

COVID-19 epidemically. Without referring to circumstance, we will miss what the results signify and its impact on the public. To amend this issue, we will conduct contextual analysis [] to elaborate the results of textual analysis of tweets and news and uncover its social and cultural perspectives of the text meanings.

Final stage of our data analysis, we will plot google trends data, the daily infected cases and fatality and frequency and main themes of media coverage and tweets, we expect to construct a proxy indicator for information dissemination and public reaction to the outbreak. This provides a platform for the contextual analysis and uncover the communication flow among the four forums.

Mathematical modeling Modelling: (Can we do this? or do we need this: incorporating media coverage, public and institutional reaction and report (which can be based on modelling results by scientific community and WHO etc), time travelling and spread network (geographical)

3.3 deliverables

- software

- Communication: Writing papers is certainly one, but we should be tweeting and etc. What are communication platforms we should be communicating in?

4 Research Setting & Personnel

The principal applicants have a long history of influential research on disease outbreak:

The research will principally take place at McMaster University. Nominated principal applicant Dr. David J.D. **Earn** (6 hours per week) led the creation of the International Infectious Disease Data Archive [33] and has expertise in gathering and curating infectious disease data, and in dynamical modelling. Principal applicant Dr. Jonathan **Dushoff** (5 hours per week) is an internationally recognized expert in infectious disease modelling, has extensive experience with statistical frameworks for fitting models to data, and has been involved in the Ebola challenge and other forecasting projects. Co-Applicant Dr. Benjamin **Bolker** (5 hours per week) is a highly accomplished ecological statistician with extensive experience in spatial-dynamical modeling, statistical modeling and statistical software. Co-Applicant Dr. Chyun-Fung **Shi** (40 hours per week), post-doctoral researcher, has ... Co-Applicant Dr. Michael **Li** (5 hours per week), post-doctoral researcher, has focused his research on epidemic forecasting and is experienced working with large databases.

((mli: fill in the rest later))

4.1 Collaborators and Knowledge Users

((mli: fill in later))

5   Research Time line

5.1   Year 1

5.2   Year 2

5.3   Potential Outcomes

Transfer findings to peer-reviewed publications. In addition to a grand one paper, we plan to transfer findings into subgroups, such as by affected country (North America, Asia-Taiwan, H.K. and Singapore), types of media (news media and social media (twitter)).

- Potential to contribute to the global response to COVID-19

- Social and policy countermeasures and Global Coordination Mechanism

6   Challenges and Mitigation Strategies

Social media in China is to included due to data availability. Social media such as twitter and Facebook are barred in China and WeChat does not share their database. Yet, there are information and coverage about the outbreak inside China in the media. We will categorize people's reaction and understanding of COVID-19 on the news content or the tweets to supplement our results.

Data curation   We will go back and document clearly all the policy changes and case defintions.

- data collected from National Health Commission

- Figure out how to use the data effectively (e.g. we are not using death, severity categorizations, number of tested, number of positive, and etc)

- Case definitions

- Media content

- Social media platforms

- Google trends

Analysis/ Pipelineing/ mainstreaming   Info delays, misinformation and miscommunications

Communication   how people are interpreting

7   Conclusion/summary

7.1   Note

We want to emphasize our bilingual (English and Chinese) background, which is essential in this study.

References