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 dynamical 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 [1] and the influenza pandemic [2], 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 [3]. **((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 [4]. Social media can play a positive role spreading good information, [5–7], but may also spread misinformation and feed bias [?, 8]. 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: What should we put back here? What do we want from [9]? Is there a more recent cite about twitter and pandemics than [10]?))**

((jd: More 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.))

((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

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 and Weibo by purchasing API access, using data going back to November 2020 – before the epidemic started – to give a baseline for comparison.

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 [11, 12]. 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

Textual analysis We will use both human coders and AI approaches for textual analysis of scientific papers, government recommendations, mass media and social media. Coders will develop codebooks based on scanning information, then refine these codebooks while systematically coding a randomly subset of articles from each stream. Codebooks will contain both themes and frames for the analysis. We will train AIs built using established approaches for textual analysis to these human-scored articles. Programmers will work back and forth with coders and subject-matter experts to reach a final codebook that is consistent with study aims and can be scored reliably by computer, allowing us to review a very large number of articles. This approach will allow us to scan for tweets (and Weibos) spreading misinformation in real time. Such information would be shared with public-health practicioners to assist with counter-messaging strategies.

Time-series analysis We will use cross-correlation analyses to look for indicators that information is moving from one communication forum to another; that events (like disease spread or public behavior) are affecting communication; or that communication is affecting events. Cross-correlation analysis is complicated and prone to false-positive results. Importantly, therefore, we will be able to use the cross-correlation analysis to generate hypotheses that can be checked by more detailed textual analysis. For example, if we hypothesize that tweets about fatalities are being driven at a certain time and place by mass media, or by government policies, we can sample from those tweets and examine them for detailed information or citations; if we hypothesize that a trend in self-isolation is driven by social media, we can search for mass media stories that interview people about their motivation. The ability to compare large-scale trends with detailed texts should amplify our pattern to detect and confirm patterns.

Dynamical modeling Dynamical modeling provides the link between individual events and emergent phenomena. We will make a range of simple dynamical models to further probe our time-series results by asking what mechanisms may underlie our observed connections, and what these connections might imply for the future. Dynamical models will allow us to explore hypotheses about what factors affect behaviour, and also to explore new hypotheses about how changes in behaviour are likely to loop back to disease transmission or to panic responses that might lead to shortages or to impacts on regional economies or the global economy.

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.

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** (4 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.

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

References

[1] T. R. Berry, J. Wharf-Higgins, and P. J. Naylor. SARS wars: an examination of the quantity and construction of health information in the news media. *Health Commun*, 21:35–44, 2007.

[2] J. M. Tchuenche, N. Dube, C. P. Bhunu, R. J. Smith, and C. T. Bauch. The impact of media coverage on the transmission dynamics of human influenza. *BMC Public Health*, 11 Suppl 1:S5, 2011 Feb 25.

[3] L. Kim, S. M. Fast, and N. Markuzon. Incorporating media data into a model of infectious disease transmission. *PLoS One*, 14:e0197646, 2019.

[4] J. Liu, L. Siegel, L. A. Gibson, Y. Kim, S. Binns, S. Emery, and R. C. Hornik. Toward an Aggregate, Implicit, and Dynamic Model of Norm Formation: Capturing Large-Scale Media Representations of Dynamic Descriptive Norms Through Automated and Crowdsourced Content Analysis. *J Commun*, 69:563–588, 2019 Dec.

[5] C. H. Basch and G. C. Hillyer. Skin cancer on Instagram: implications for adolescents and young adults. *Int J Adolesc Med Health*, 2020 Feb 7.

[6] M. Sun, L. Yang, W. Chen, H. Luo, K. Zheng, Y. Zhang, T. Lian, Y. Yang, and J. Ni. Current status of official WeChat accounts for public health education. *J Public Health (Oxf)*, 2020 Jan 23.

[7] N. Ahmed, S. C. Quinn, G. R. Hancock, V. S. Freimuth, and A. Jamison. Social media use and influenza vaccine uptake among White and African American adults. *Vaccine*, 36:7556–7561, 2018 Nov 26.

[8] W. Y. S. Chou, A. Oh, and W. M. P. Klein. Addressing Health-Related Misinformation on Social Media. *JAMA*, 320:2417–2418, 2018 Dec 18.

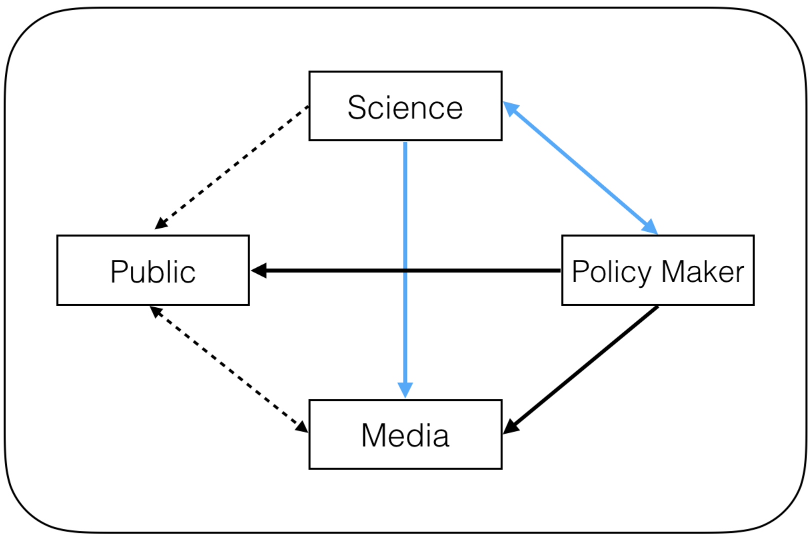
[9] S. Yousefinaghani, R. Dara, Z. Poljak, T. M. Bernardo, and S. Sharif. The Assessment of Twitter's Potential for Outbreak Detection: Avian Influenza Case Study. *Sci Rep*, 9:18147, 2019 Dec 3.

[10] C. Chew and G. Eysenbach. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. *PLoS One*, 5:e14118, 2010 Nov 29.

[11] J. Bousquet, I. Agache, J. M. Anto, K. C. Bergmann, C. Bachert, I. Annesi-Maesano, P. J. Bousquet, G. D'Amato, P. Demoly, G. De Vries, E. Eller, W. J. Fokkens, J. Fonseca, T. Haahtela, P. W. Hellings, J. Just, T. Keil, L. Klimek, P. Kuna, K. C. Lodrup Carlsen, R. Mosges, R. Murray, K. Nekam, G. Onorato, N. G. Papadopoulos, B. Samolinski, P. Schmid-Grendelmeier, M. Thibaudon, P. Tomazic, M. Triggiani, A. Valiulis, E. Valovirta, M. Van Eerd, M. Wickman, T. Zuberbier, and A. Sheikh. Google Trends terms reporting rhinitis and related topics differ in European countries. *Allergy*, 72:1261–1266, 2017 Aug.

[12] N. Mahroum, N. L. Bragazzi, F. Brigo, R. Waknin, K. Sharif, H. Mahagna, H. Amital, and A. Watad. Capturing public interest toward new tools for controlling human immunodeficiency virus (HIV) infection exploiting data from Google Trends. *Health Informatics J*, 25:1383–1397, 2019 Dec.

Appendix



- Blue arrows = evidence based info, but probably very limited, very little get through and hard to communicate/understand
- Black arrows = Policies and decisions (easy and straightforward to communicate)
- Dashed arrows: Opinions, and any form of communication. It does not even have to be evidence-based.

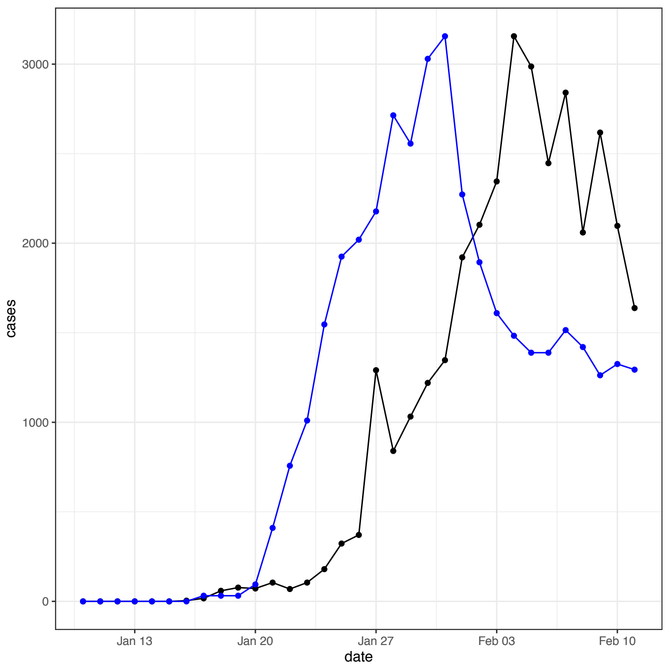


Figure 1: Black line represents COVID-19 incidence data from HuBei. Blue line represents google trend patterns of the search term "coronavirus" worldwide using the same time window as HuBei time series data. Note: google trend data are scaled relative to the peak to anonymize traffic flow (i.e. the peak will have a value of 1). To match the HuBei time series data, I scale the google trend data to the peak incidence of the time series.