Infodemiology: Computational Methodologies for quantifying and visualizing key characteristics of the COVID-19 infodemic

Authors

Dominic Vincent Ligot, Frances Claire Tayco, Mark Toledo, Carlos Nazareno, and Denise Brennan-Rieder.

Correspondence to Dominic Vincent Ligot (email: dominic.ligot@cirrolytix.com)

Abstract

Objectives. Infodemics of false information on social media is a growing societal problem, aggravated by the occurrence of the COVID-19 pandemic. The development of infodemics has characteristic resemblances to epidemics of infectious diseases. This paper presents several methodologies which aim to measure the extent and development of infodemics through the lens of epidemiology.

Methods. Time varying R was used as a measure for the infectiousness of the infodemic, topic modeling was used to create topic clouds and topic similarity heat maps, while network analysis was used to create directed and undirected graphs to identify super-spreader and multiple carrier communities on social media.

Results. Forty-two (42) latent topics were discovered. Reproductive trends for a specific topic were observed to have significantly higher peaks (R_t 4-5) than general misinformation (R_t 1-3). From a sample of social media misinformation posts, a total of 385 groups and 804 connections were found within the network, with the largest group having 1,643 shares and 1,063,579 interactions over a 12 month period.

Conclusions. These approaches enable the measurement of the infectiousness of an infodemic, comparative analysis of infodemic topics, and identification of likely super-spreaders and multiple carriers on social media. The results of these analyses can form the basis for taking action to stem an ongoing spread of misinformation on social media and mitigate against future infodemics. The methods are not confined to health misinformation and may be applied to other infodemics, such as conspiracy theories, political disinformation, and climate change denial.

Keywords

COVID-19, Infodemiology, Network Analysis, Topic Modeling, Reproductive Number, Misinformation, Disinformation, Infodemic, Fake News, Crowdtangle, Facebook, Reddit

Introduction

Misinformation reduces individual capacity to discern fact from fiction and causes harm (2,27,16). With 4.66 billion people having access to the internet and 53% of the global population using social media (17), significant numbers of people separated by space and time

are affected (2) and the damage caused by inaccurate or fabricated information can be profound. Misinformation has caused widespread lack of understanding of major public health issues, sometimes resulting in life-or-death impacts (28).

An infodemic on social media bears a resemblance to an epidemic of a viral disease in that it has similar properties: a population “infected” with a piece of inaccurate information and transmission of that piece of information from one “infected” individual to others (7). The propagation of social media information in an infodemic is not limited by boundaries (2), crossing countries, continents and languages (20).

The WHO recognizes inaccurate information as a threat and recommends systematic monitoring and control measures (18). This paper describes three methodologies to analyze the content and spread of misinformation in the coronavirus disease 2019 (COVID-19) pandemic that was caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and it aims to analyze key characteristics such as morbidity and rate of spread of information and the social media groups and pages that are spreading the false information.

The methodologies outlined in this paper are applicable not just applicable to health misinformation but also other domains such as conspiracy theories, political and election disinformation, and climate change denial.

Methodology

Topic Modelling

Topic modeling is a dimensionality reduction technique, which links words of original texts with identical context, by examining repeating patterns of co-occurrences of words in a combined set of documents (also known as a corpus), and discovers themes that run through them, called topics (1).

Topic modeling can be performed on any text corpus, or any dataset with a field containing free-form text values to build inputs or features from. To demonstrate this method, COVID19Misinfo.Org's daily COVID-19-related claims data from January 1 to October 20, 2020 (13) was used, containing fields for the texts of claims in English and fact-checker ratings (e.g. True, False, Misleading, Unproven).

Data Preparation

As a data cleansing and preparation step, the texts from documents in the corpus were tokenized, stop words were removed, and the tokens were lemmatized. A document-term matrix representation of the processed data was derived, with term-frequency inverse document frequency (TF-IDF) statistics as values (30).

Non-negative matrix factorization (NMF) (6) was employed to mine the topics. To select the optimal number of topics, different numbers of topics were tested. Topic coherence score, which measures whether pairs of terms used to describe a topic tend to co-occur together relative to the

whole corpus was computed at each iteration. C_v was the variant of coherence score used, as it is the one most aligned with human interpretability (26). The NMF model with the highest C_v , obtained with an elbow (the point of inflection on the plot of scores) method, was chosen as the best one (29).

Time-varying reproductive number of infodemic time-series

For infodemics, instances of false information mentioned in social media are treated as analogous to *cases* of infections. A reproductive number, which epidemiologists use to describe infectiousness of a disease, can be estimated from social media posts with false information to characterize the information diffusion process among users of those networks (21, 31).

During an actual course of an epidemic, an effective reproductive number or time-varying reproductive number, R_t , is calculated as a way of monitoring the trend of an epidemic. R_t estimates the average number of new infections caused by a single case at time t . An epidemic that maintains R_t above a threshold of 1.0 is considered growing, while R_t below 1.0 denotes a weakening trend (8).

An R_t figure was generated from a time-series of number of posts about a particular false information topic to provide a measure of the rate of its spread. Counts of engagements and interactions on the social media posts were considered a secondary measure of the *virality* of the topic, and were also used to generate an infodemic R_t (8, 21, 31).

The results of the topic modeling of debunked COVID-19 claims were used to score daily Facebook page post data on reliability. The dataset posts were from January 16 to October 23, 2020 and contained the keywords *Coronavirus* or *COVID* (9, 32). The dataset fields included

date created, post texts (e.g. message, link text, image text, description) and interactions (e.g. likes, comments, shares, reactions). This was compared with John Hopkins University Center for Systems Science and Engineering (CSSE) COVID-19 daily cases data from January 22 to October 20, 2020 (10). R_t was then estimated for both time-series (8).

Topic similarity

Latent topics automatically generated using NMF were evaluated for the similarity of their content using *cosine similarity*, a popular measure for detecting similarity between documents of varying sizes (15). Using the cosine values, a similarity matrix between pairs of topics, based on their top 20 terms (3) was constructed. Cosine values range from 0 to 1. The closer the values are to 1, the more similar the topics.

The matrix was visualized as a heat map, which showed high similarity values in green, low values in yellow, and zeros in red, for ease of interpretation. This type of visualization enables one to easily identify topic pairs that are quite similar or related.

Network Analysis

A network graph is an easily interpretable format for representing multi-relational data. A graph G consists of vertices or nodes denoted as $V(G)$ and edges that link the nodes denoted as $E(G)$. Nodes represent entities while edges describe the relationships between the nodes. Graphs that have directional edges are called *directed* graphs otherwise they are referred to as *undirected*.

Graph Characteristics and Metrics

Once a graph is constructed, key measures can be calculated for each node. In directed graphs, *in-degree* is the number of edges pointing into a node $k_i^{in} = \sum_j a_{ij}$, while *out-degree* is the number of edges pointing out of a node $k_i^{out} = \sum_j a_{ij}$. Adding *in-degree* and *out-degree* results in *degree* $k_i^{tot} = k_i^{in} + k_i^{out}$, which is also the total number of edges in undirected graphs. (23)

Important nodes in a graph are highlighted using *centrality* measures such as: *page rank* (24)

$PR(u) = \sum_{v \in B_u} \frac{PR(v)}{L(v)}$ which ranks nodes based on the structure of incoming links; *betweenness*

centrality (5) $c_B(v) = \sum_{s,t \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)}$ which is based on the shortest path between nodes; and

closeness centrality (12) $c(u) = \frac{n-1}{\sum_{v=1}^{n-1} d(v,u)}$ which is the reciprocal of the average shortest path

distance to the node over all other reachable nodes.

Contact Tracing of False Information

Social media groups that spread false information were considered infected hosts and formed the nodes in the networks. Infodemic spreaders were identified through topics and URLs that were being shared in Facebook and Reddit through *undirected* graphs. Super-spreaders and multiple carriers were identified through *directed* graphs by considering the period topics were shared.

Dataset and generating the networks

NewsGuard, an entity which uses journalists to rate news sites on reliability, released a list of Red-Rated Sites with False Claims about the Coronavirus (22). This list of sites was used to sample URLs spreading misinformation. Crowdtangle was then used to extract social engagement data from these Facebook and Reddit communities, e.g. (i) Red-listed URL Link - <https://www.infowars.com/>, (ii) Source - /r/conspiracy, (iii) Followers - 1,189,707, (iv) Date - Wed Apr 29 2020 17:02:37 GMT+0000, (v) Interactions - 343, (vi) Post Type - Reddit, (vi) Link - <https://redd.it/gad5vc>

The data also included community size, posting frequency, and member interactions for each platform. The CrowdTangle free browser extension was used and up to 500 rows of data per URL was gathered. CrowdTangle is a public insights tool owned and operated by Facebook (9).

Undirected graph - grouping cases by strain of URLs

Data was exported within a date range of November 2019 to November 2020 and the following metrics were generated through aggregation: (i) Share Rate (SR) derived by counting the times the URL is shared in a specific group, (ii) Interaction Rate (IR) derived from the total number of user interactions of a red-listed site.

In the graphs, the red-listed site was represented as red nodes with a fixed size while the social communities were represented as the blue nodes with sizes reflected by either SR or IR. SR and

IR per URL were then represented as colored edges, with all edge values normalized to 1-50. Color assignment is as follows: blue (1-10), yellow (11-20), orange (21-30), red orange (31-40), and red (41-50).

Directed graph - grouping cases by sequence

Using the same dataset, cases were grouped by time and edges were created from paths generated between the nodes in temporal sequence $P = (v_t, v_{t+1}, \dots, v_{t+n})$ to visualize the infodemic spread over a time window.

Network Metrics and Visualization

Degree, betweenness, closeness, and page rank centrality values were generated for each node. Network visualizations were generated using the Force Atlas 2 algorithm (19) with a random seed of 300.

Results

Topic Modeling

Forty-two (42) latent topics were discovered and visualized as word clouds, which display clusters of words with varying sizes that represent the weight of the relevant terms in the topics. The combination of words and their sizes give us an idea of the contents of the posts under each topic. For labeling, top posts under each topic were examined.

Figure 1 shows some of the topic word clouds generated. For example, word cloud (a) highlights the terms *throat*, *drink*, *hot*, *water*, *prevent*, and *kill*. Some top posts under this topic include “*Drinking water or hot teas kills the coronavirus, which is weak and cannot resist heat*”, “*Drinking water and gargling helps against the coronavirus*”, and “*The coronavirus, before it reaches the lungs, remain in the throat for four days...*” This topic was labeled *Virus in the throat + water, gargling and tea as COVID treatment*.

Figure 1: Example of Latent Topics' Word Clouds:

(a) Virus in the throat + water, gargling and tea as COVID treatment; (b) Anti-masks; (c) COVID studies + hydroxychloroquine; (d) Bat + 5G Origin + tracing; (e) President / VP related news; and (f) COVID cases numbers.

Different topic categories also surfaced. Categories include health-related misinformation (e.g. diagnostics, preventions, treatments, symptoms, transmission), as well as false information related to governments and public figures, false claims about virus severity, and speculations about the origin of COVID-19 and conspiracy theories, to name a few.

Infodemic and Pandemic R_t

Time varying reproductive numbers, R_t , were derived based on the number of posts with misinformation, number of interactions with such posts, number of posts by topic, and number of COVID-19 cases.

Figure 2: Brazil's Infodemic Effective Reproductive Number

Figure 2 presents comparisons of the R_t figures using Brazil as an example. The first graph compares the R_t based on the number of posts and the COVID-19 case counts. In the example, both series fluctuated mostly above the threshold of 1.0 within the covered period, which suggests that an infodemic is happening simultaneously with the epidemic.

The second graph compares R_t based on all misinformation against posts about a specific topic. In this example the R_t on the topic "President and Vice President" was observed to have

significantly higher peaks at 4-5 than the general R_t of misinformation with peaks at 1-3. Recurring peaks in the R_t of the “President and Vice President” topic were observed from May 2020.

Topic Similarity Study

The topic heat map shows similarity in the content in several of the 42 misinformation themes identified. Cosine values range from 0 to 0.31, which indicate some shared content between some pairs of topics.

Figure 3 shows an excerpt of the heatmap. In the example, topic *Coronavirus + claim + evidence + share* shares similar features with *COVID Pre-2019 and Early News About It*, with cosine value of 0.25. Likewise, the topic *Coronavirus / COVID transmission + treatment + survival in high temperature* appeared quite related with the topic *Virus in the throat + water, gargling, and tea as COVID-19 treatment* (cosine similarity = 0.29).

Figure 3: Topic Similarity Heat Map

Network Analysis

Figure 4: Undirected Network Graph from Reddit Groups (left)
and Directed Network Graph from Facebook Groups (right)

The left portion of Figure 4 is an undirected network graph that shows how 20 red-listed sites (red nodes) are being shared across Reddit subgroup communities known as Subreddits (blue nodes) and how much users in each community interact with the site link posting. There are a total of 385 groups and a total of 804 connections found within the network. Examining the graph shows *r/Conspiracy* and *r/Conservative* as the largest nodes which are central to the network, and that these groups are major hotspots for misinformation. It also mirrors the findings that conservatives are more likely to share articles from fake news domains (14). *r/Conservative*

also has the most share count totaling to 1,643 shares and most interaction counts totaling 1,063,579 interactions and this data was normalized to the ranges 1-50.

This *undirected* graph resulted in a clustering of hosts based on the specific URL. Each URL is a source node (virus) and each social media group is a target node (host), an edge is created by sharing or interacting with a red-listed site within the group (viral transmission), and a collection of nodes and edges are clusters of *digital hotspots* (33) within the network (infected population).

The graph indicates the existence of super-host communities, illustrated by large interaction rates and shared site postings which further suggests that red-listed site operators or subreddit members are sharing questionable content to target communities.

The *directed* graph visualizes the sharing of a red-listed site over time, showing which groups are more susceptible to infection, what links are driving the infodemic, and which central nodes are possible superspreaders. Groups with themes such as *vaccine*, *anti-aging*, and *conservative* were observed to spread and react to content shared by the community more often than other groups.

node	page_rank	betweenness	closeness	degrees	Share Count	Share Rate Normalized	Interaction Count	Interaction Rate Normalized
Conservative	0.006051	0.046317	0.46528	13	1643	1500.00	1063579	1500.00
conspiracy	0.008142	0.0790124	0.51145	18	509	464.76	177794	251.58
Politics	0.001688	0.0004847	0.2847	3	37	33.86	79219	112.65
r/Libertarian: For a Free Society	0.006008	0.0365303	0.43982	13	36	32.95	31900	45.96
World News	0.0032	0.0117508	0.35702	6	14	12.87	30202	43.57
AskThe_Donald	0.005015	0.0177537	0.39105	11	53	48.47	21877	31.83
Today I Learned (TIL)	0.001483	0.0021038	0.31604	2	2	1.91	19362	28.29
News item submissions for TWIT's Tech News Today	0.000956	0	0.26172	1	5	4.65	18961	27.72
ConservativesOnly -- If you are a conservative... welcome!	0.004209	0.0130663	0.36612	9	314	286.74	17542	25.72
Popcorn tastes good.	0.001007	0	0.29646	1	1	1.00	17496	25.66
Novel Coronavirus (COVID-19)	0.003142	0.0143173	0.39412	6	6	5.56	15859	23.35
All posts must be related to gun politics; civil discussion is	0.002938	0.0035413	0.31455	6	13	11.95	14823	21.89
wallstreetbets	0.00144	0.0002452	0.26693	2	5	4.65	13912	20.61
conservatives	0.005338	0.0170177	0.39105	12	154	140.68	12896	19.17
...	...	...	...	...	...	...	...	...

Figure 5: Network Analysis Metrics Output

Figure 5 shows the network analysis metrics for the reddit network containing the 20 unique red-listed URLs with the highest interaction rates in the entire dataset. Interactions (likes, comment, shares) correlate with degree (number of node connections) which indicates that users were interacting heavily with the red-listed content being shared in the group. Higher share rates also indicate that more members are actively and deliberately sharing red-listed content to the group. Nodes with higher closeness centrality are spreading information more efficiently relative to other nodes.

Discussion

This paper presents methodologies to measure the extent and development of infodemics: Time varying R to measure the infectiousness of an infodemic, topic clouds and topic similarity heat maps to analyze content of false information, and directed and undirected graphs to identify super spreader and multiple carrier accounts on social media. Posts from Facebook and Reddit extracted through CrowdTangle were used to illustrate the approaches, but other data sets such as from Twitter, Instagram, and other platforms may be substituted.

Topic modelling was used to describe the major misinformation topics and categories and encapsulate the major misinformation narratives of the infodemic. Carried out over smaller windows of time, it may be possible to use this approach to demonstrate evolution of misinformation (35). In addition, topic modelling provides a means of monitoring the social media landscape for the appearance of new topics/categories which could possibly enable early detection of disinformation.

The R_t provides a means to illustrate both the relative popularity of a given piece of misinformation over time. Additionally, having the ability to estimate topic-level R_t and monitor the ones exceeding the overall R_t helps in identifying the primary viral themes that are driving the misinformation spreading during the infodemic. Furthermore, having the ability to visualize and compare both infodemic and epidemic R_t can serve as a foundation for future research on their relationship.

Identification of topic similarities can aid in development of communication campaigns that can counter multiple misinformation themes. Moreover, since cosine values suggest the level of similarity of the most relevant terms between topics, and automatically generated topics are assumed to be distinct, identifying topics that are highly related with numerous other topics might indicate potential cases of deliberate disinformation being engineered towards the unsuspecting public. The similarity heat map analysis can serve as an alert system if subsequent investigation would be necessary.

The network analysis identifies the propagators of misinformation, major influencers, the extent of their networks and the particular misinformation they are pushing. Ranking nodes by graph metrics and user shares and interactions might identify digital hotspots that tend to be more susceptible to misinformation, which are candidates for proactive management strategies. Pareto principle or power law dynamic might apply to infodemics where 80% of new transmissions are caused by fewer than 20% of the carriers, with individuals or members who are aggressively spreading false information (25, 34).

The approaches described in the paper are not confined to health misinformation, and may be applied towards understanding other infodemic themes such as political disinformation, climate change denial, and conspiracy theories. These approaches empower the analysis and management of infodemics through an epidemiological lens, and provide bases for strategies to stem an ongoing spread and mitigate against future infodemics.

Authors

The CirroLytix Research Team includes Dominic Vincent Ligot, Frances Claire Tayco, Mark Toledo, Carlos Nazareno, and Denise Brennan-Rieder. All team members jointly conceptualized the study, analyzed and interpreted the data, wrote and revised the manuscript and decided for submission.

Corresponding author: Dominic Vincent Ligot, dominic.ligot@cirrolytix.com.

References

1. Aggarwal, C., & Zhai, C. (2012). *Mining Text Data*. Springer.
2. Ahmed, S. M., Mushtaq, K., & Al Soub, H. (2020;103(2):). Social Media Misinformation”—An Epidemic within the COVID-19 Pandemic. In *The American Journal of Tropical Medicine and Hygiene* (pp. 920-921).
3. Aletras, N., & Stevenson, M. (2014). Measuring the Similarity between Automatically Generated Topics.

4. Bernard, R., Bowsher, G., Sullivan, R., & Gibson-Fall, F. (2020). *Disinformation And Epidemics: Anticipating The Next Phase Of Biowarfare*. Mary Ann Liebert, Inc.
<https://doi.org/10.1089/hs.2020.0038>
5. Brandes, U. (2001). Page 1 A Faster Algorithm for Betweenness Centrality.
6. Chen, Y., & Bordes, J. (2016). An experimental comparison between NMF and LDA for active cross-situational object-word learning. In *Sixth Joint IEEE International Conference Developmental Learning and Epigenetic Robotics (ICDL-EPIROB)*. Cergy-Pontoise.
7. Cinelli, M., Quattrocioni, W., Galeazzi, A., & Valensise, C. (2020). The COVID-19 social media infodemic. In *Scientific Reports* (Vol. 10(1), p. 16598).
8. Cori, A., Ferguson, N., Fraser, C., & Cauchemez, S. (2013). *A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics, Volume 178*(Issue 9), Pages 1505–1512. <https://doi.org/10.1093/aje/kwt133>
9. *Crowdtangle Team*. (2020). Facebook. Menlo Park, California, United States.
10. Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *20*(5), P533-534.
[https://doi.org/10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1)
11. DQ, N. (n.d.). *The degree distribution of a network*. The degree distribution of a network. Nykamp DQ, “The degree distribution of a network.” From Math Insight.
http://mathinsight.org/degree_distribution
12. Freeman, L. (1978). Centrality in Social Networks Conceptual Clarification. *Centrality in Social Networks Conceptual Clarification*.

13. Gruzdz, A., & Mai, P. (2020). *Misinformation Portal – A Rapid Response Project from the Ryerson University Social Media Lab*. Retrieved November 30, 2020, from <https://covid19misinfo.org/>
14. Guess, A., Nagler, J., & Tucker, J. (2019, January 9). *Less than you think: Prevalence and predictors of fake news dissemination on Facebook*. Less than you think: Prevalence and predictors of fake news dissemination on Facebook | Science Advances. Retrieved November 30, 2020, from <https://advances.sciencemag.org/content/5/1/eaau4586>
15. Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques*, third edition (3rd ed.).
16. Hartley, K., & Vu, M. (2020;53(4)). Fighting fake news in the COVID-19 era: policy insights from an equilibrium model. In *Policy Sciences* (pp. 735-758).
17. *Hootsuite & We Are Social (2020)*. (n.d.). Digital 2020: October Global Statshot. <https://datareportal.com/reports/digital-2020-october-global-statshot>
18. Islam, M., Sarkar, T., & Khan, S. (2020). COVID-19–Related Infodemic and Its Impact on Public Health: A Global Social Media Analysis. In (pp. 1621-1629). *The American Journal of Tropical Medicine and Hygiene*.
19. Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014, June 10). *ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software*. ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. Retrieved November 30, 2020, from <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0098679>
20. Johnson, N., Velásquez, N., & Restrepo, N. (2020). The online competition between pro- and anti-vaccination views. In *Nature* (Vol. 582(7811), pp. 230-233).

21. Li, M., Wang, X., Gao, K., & Zhang, S. (2017). *A Survey on Information Diffusion in Online Social Networks: Models and Methods*, *Information* 2017(8), 118.
22. NewsGuard. (2020, November 25). *Red-Rated Sites With False Claims About The Coronavirus: 366 And Counting*. Coronavirus Misinformation Tracking Center. Retrieved November 30, 2020, from <https://www.newsguardtech.com/coronavirus-misinformation-tracking-center/>
23. Nycamp, D. (n.d.). *Node degree definition*. Math Insight. Retrieved November 30, 2020, from http://mathinsight.org/definition/node_degree
24. Page, L., Brin, S., Motwani, R., & Winograd, T. (1998). The PageRank Citation Ranking: Bringing Order to the Web. *The PageRank Citation Ranking: Bringing Order to the Web*.
25. Patel, N. (2020, November 20). *What's a coronavirus superspreader?* What's a coronavirus superspreader? Retrieved November 28, 2020, from <https://www.technologyreview.com/2020/06/15/1003576/whats-a-coronavirus-superspreader/>
26. Röder, M., Both, A., & Hinneburg, A. (2015). *Exploring the Space of Topic Coherence Measures*. ACM.
27. Spies, S. (2020). *How Misinformation Spreads [Internet]*. Mediawell. Retrieved November 30, 2020, from [Mediawell.ssrc.org](https://mediawell.ssrc.org)
28. Starbird, K., Spiro, E., & Koltai, K. (2020). *Misinformation, Crisis, And Public Health—Reviewing The Literature - Version V1.0. [online]*. MediaWell. <https://mediawell.ssrc.org/literature-reviews/misinformation-crisis-and-public-health/versions/v1-0/>

29. Syed, S., & Spruit, M. (2017). *Full-text or abstract? Examining topic coherence scores using latent dirichlet allocation*, 2017 IEEE International conference on data science and advanced analytics (DSAA), P165-174.
30. Weiss, S., Indurkha, N., & Zang, T. (2015). *Fundamentals of Predictive Text Mining*. Springer-Verlag.
31. Zhang, Z., Liu, C., Zhan, X., Zhang, C., & Zhang, Y. (2016). *Dynamics of information diffusion and its applications on complex networks*, *Physics Reports*(651), 1-34.
32. Zhou, X., & Zafarani, R. (2020). A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. <https://doi.org/10.1145/3395046>
33. Lessler, Justin, et al. *What is a Hotspot Anyway?*, 2017, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5462559/>
34. *How to fight the COVID-19 infodemic: lessons from 3 Asian countries*. 2020. World Economic Forum. Retrieved November 30, 2020 from <https://www.weforum.org/agenda/2020/05/how-to-fight-the-covid-19-infodemic-lessons-from-3-asian-countries/>
35. Jang S, Geng T, Li, Xia R, Huang, Kim et al. 2018. *A computational approach for examining the roots and spreading patterns of fake news: evolution tree analysis*. *Computers in human behaviour*. 2018;84:103-113.