Behavior Analysis of Pandemic Source Media Communications

1st Yohn Jairo Parra Bautista, PhD
Computer and Information Sciences
Florida A&M University
Tallahassee, USA
yohn.parrabautista@famu.edu

2nd Jinwei Liu, PhD
Computer and Information Sciences
Florida A&M University
Tallahassee, USA
jinwei.liu@famu.edu

3rd Richard Aló, PhD
Dean of College of Science and Technology
Florida A&M University
Tallahassee, USA
richard.alo@famu.edu

2021 IEEE International Conference on Big Data and Smart Computing (BigComp) | 978-1-7281-8924-6/20/$31.00 ©2021 IEEE | DOI: 10.1109/BigComp51126.2021.00018

***Abstract*—We aim to collect and analyze source media top stories and investigate the patterns of health communication change in this ongoing COVID-19 worldwide crisis. We explore the influence of source messages in mass media communications. The pandemic is exposing dramatic behavioral changes requiring immediate reflection. With various measures such as: lockdowns, mask wearing and social distancing in place, it becomes essential to focus on the impact and effect of mass media. We use SiloBreaker dashboard to collect top stories since the pandemic's early outbreak. Techniques like tf-idf, topic modeling and LDA Emotion Analysis, assist in producing useful information about the significant impact of mass media on source messaging change.**

***Index Terms*—text analysis, pandemic, behavior change, topic modeling, mass media**

I. Introduction

Mass media influences society in specific ways that impact behavior. Attitude and behavior change involve many cognitive processes as to how we perceive the world [1]. For instance, during these pandemic times, mass media campaigns and their impact increase considerably, accentuated by five communication factors: source, message, channel, received, and destination [2]. It is crucial to identify the emotional level of acceptance that most affects the communication factors.

Mass media communications work diligently to deliver top stories of the latest COVID-19 information to the public audience. Important outcomes come from the ability to break out of silos in getting the top stories from different worldwide media sources. We collect and analyze media top story reports [2]. We classify massive vs non-massive stories to bring attention to the exposure and awareness to the message. We present lexical diversity by periods in order to look at the change in knowledge using unique words and memory of the message content during these periods. With this training dataset, we analyze emotion level, important TF-IDF words, and Topic Modeling [3].

II. Background

Covid-19 pandemic cases currently expose more knowledge on the spread and advance of contagious and infectious disease. We suspect that the only main approach available now to reduce transmission is behavioral[4]. For instance, hand washing, cough and sneeze protection by wearing a mask, and above all, social distancing can mitigate spread. The rapid spread of the virus receives much attention from governments, companies, and pharmaceuticals. Countries like Australia, Germany, Spain, and the USA are devising protocols to protect the most vulnerable. Sweden is an example of a country primarily relying on voluntary social distancing guidelines ever since the pandemic began, including working from home where possible and avoiding public transport[5]. However, we know that the strategy taken by the Swedish authorities was to introduce sustainable, long-term measures to behavior change[5].

Company actions like Google Doodle ask people to wear masks and save lives. For instance, the circles on the grounds of San Francisco's Dolores Park are a clear example of the Health Action Process Approach(HAPA)[6]. Policymakers have a set of tools to encourage and recommend these actions [7]. An essential step to identify the interventions that most reduce transmission at the lowest economic and psychological cost is randomized controlled trials with conjunction of quasi-arbitrary cutoffs and observational studies[8]. Interventions could range from messaging campaigns to promote social distancing to laws and regularization.

Policymakers' interventions deal with uncertainty. Interventions' benefits, in terms of reducing virus transmission, exceed its economic and psychological costs [9]. For instance, actions in drugs, treatments, and vaccines can improve The Health Belief Model (HBM) [6]. Perceived susceptibility and severity is an indicator of people who want to volunteer in a coronavirus vaccine trial. However, systemic racism and chronic health inequalities can bias the advancement of these actions as a digit in the HBM.

III. Methods

A. Data Collection

We collect 2194 worldwide news from SiloBreaker database related to COVID-19 from April 1 to July 31, 2020. SiloBreaker database is one of the most robust, intelligent platforms for the big data era. We organize data in top stories that contain actions from governments, companies, drugs, treatments, vaccines, societal and geopolitical impact and cybersecurity[10]. We then apply pre-processing techniques to clean and prepare the data for further analysis.

DOI 10.1109/BigComp51126.2021.00018

We gather insights from the early period of April to July. We classify more than 10 documents related to one topic as massive stories. We create a dashboard with some queries in SiloBreaker to collect daily top stories using keywords like: ”coronavirus”, ”pandemic”, ”vaccines” , ”treatments”, ”government actions” and ”company actions”.

Given the desire to know whether mass media influence behavior change, we undertook feature selection using topic modeling to investigate the change of topics over the periods. Data processing combines multiple information from different media sources in the USA and from the rest of the world. We use topic modeling to compare different models by grouping similar words of the five communication factors[11,12]. Previous research focused on the implications of health communication during early outbreaks[14]. There is a need for an evaluation of health communication data during past months.

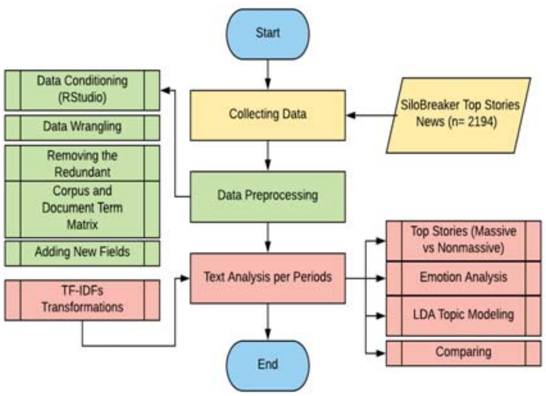

Fig. 1. Data processing flowchart.TF-IDF: term frequency-inverse document;LDA:latent Dirichlet allocation

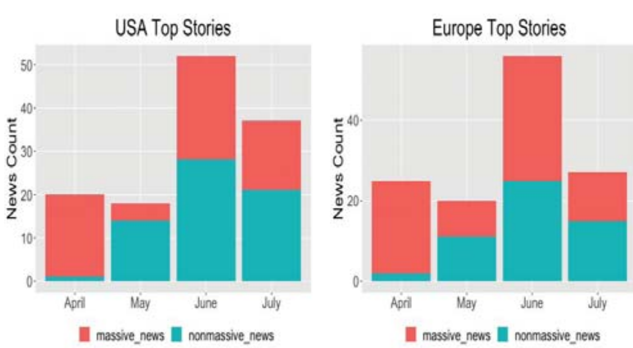

Fig. 2. Massive vs Non-massive USA Top Stories Covid-19 related and Europe .

In figures 2 and 3, we tabulate the distribution of massive vs non-massive stories released to the world from April 1, to July 31. All the countries show a lot of activity in June; however, we can see an interesting pattern of few massive stories in USA during month of May. In contrast, Australia

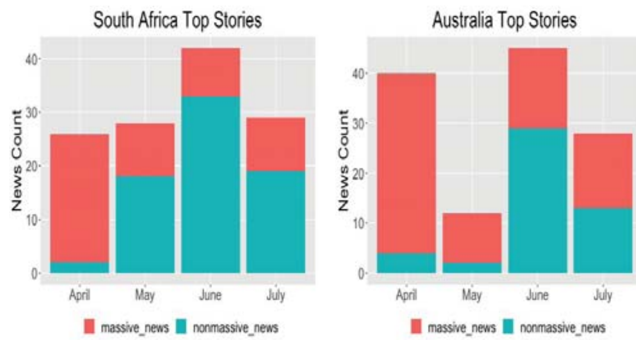

Fig. 3. Massive vs Non-massive South Africa Top Stories Covid-19 related and Australia .

has a significant amount of massive news activity during the month of April.

B. Topic Modeling

Topic modeling is well known for text mining researchers in Natural Language Processing (NLP), which aims to extract and interpret knowledge from data [13]. A popular method, such as Latent Dirichlet Allocation (LDA) has been proven beneficial in finding useful words that are associated with topics[13].

LDA is a mathematical method for estimating a mixture of topics with a mixture of words. The method is astonishingly useful in finding the mixture of words that are associated with each topic, while also determining the mixture of topics that describe each document[10].

IV. Results

We design a model to identify themes with top words by source media between countries. We extract the top terms for each topic and then we visualize six top terms in this case. Unsupervised learning required human interpretation and these topics reflect the mass media sources of April, May, June and July. We have 1000 total media sources and we select 125 top stories per region or country to have a tidy balance training data.

In figures 4 and 5, we have a variance in emotion level through the periods. We use NRC lexicon to match our dataset. NRC lexicon is a collection of lexicons that was entirely created by the experts of the National Research Council of Canada [17]. USA shows more top stories with surprise and anger emotion levels in April compared with Europe. We can observe that Europe’s stories with fear impact have slight upward trend in May. South Africa shows more stories with anticipation, fear and surprise in May; whereas Australia has major emotion impact in April.

In figure 6, themes of death, updates and nurses in USA were having a lot of attention in April. In the second topic, expectations about the virus spreading and people suffering showed decrease in May. However mass media was covering a significant amount from June to July.

In figure 7, we have Europe with themes of record, report and data that increase since May. In contrast, we have a lot of media covering topics about global issues related to the pandemic in May.

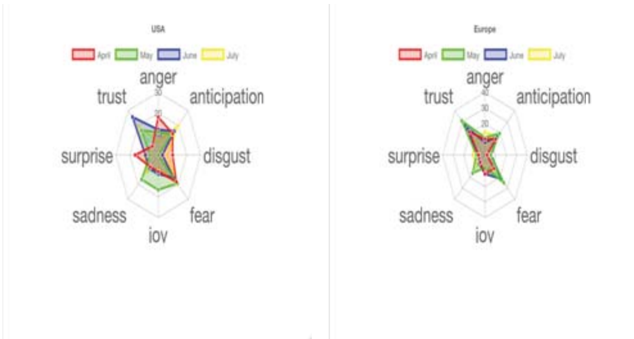

Fig. 4. Emotion Level Analysis with Top Words by Source Media Between USA and EUROPE

Fig. 5. Emotion Level Analysis with Top Words by Source Media Between SOUTH AFRICA and AUSTRALIA

In figure 8, we have themes of death, testing and exposed losing interest since May in South Africa. Surprisingly, themes of spread in some countries like Kenya was also not in the intention of massive stories.

In figure 9, Australia shows decreasing interest in topics related to the virus in states like Victoria. Interest with neighboring states was synchronized with topic one.

V. Discussion

We gathered data since early April with a focus on top stories that perceived the threat of the health risk to be dangerous. The cost-benefit to adopt a new behavior is low when the benefits of engaging in protective action cover the five communication factors. Mass media campaigns are rapid and severe. In terms of promoting preventive actions, they try to change the beliefs of people. However, the specific beliefs: perceived susceptibility and severity to the disease in conjunction with perceived benefits and barriers are actions strongly linked to the behavioral questions[6, 15].

Perceived susceptibility and severity is the primary indicator of resistance to behavior change. Mass media campaigns carry a confident attitude that impacts resistance to behavior change. For instance, a person's perception of the disease in terms of emotional arousal is a measure only at the resistance level [17]. We label news data to be explored in case of mass media news campaigns. We investigate emotion labels in a document collection of top stories, main content, travel ban, and vaccines related to COVID-19 from April 1, 2020 to July 31, 2020.

Moreover societal and economic impact in behavior change due to Covid-19 is showing to the world embarrassing corruption activities[17]. As an example, the AFC (African National Congress) is outraged and deeply embarrassed by recent allegations that some, including its own leaders and members, have sought to benefit unlawfully from the devastating suffering and impoverishment caused by the Covid-19 pandemic. Surprisingly, we found that few documents in non-massive stories were delivered to the world with this type of behavior.

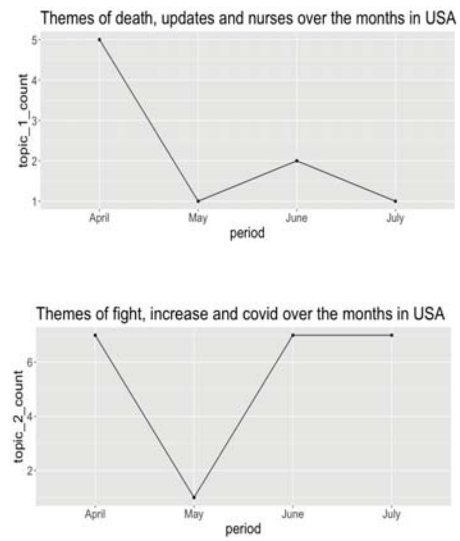

Fig. 6. Themes Overtime Between Topic 1 and 2 in USA.

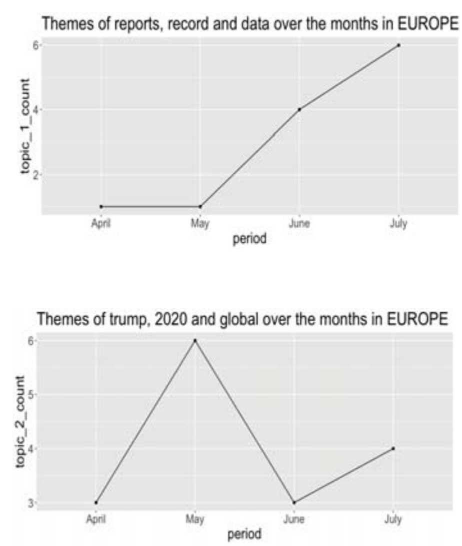

Fig. 7. Themes Overtime Between Topic 1 and 2 in EUROPE.

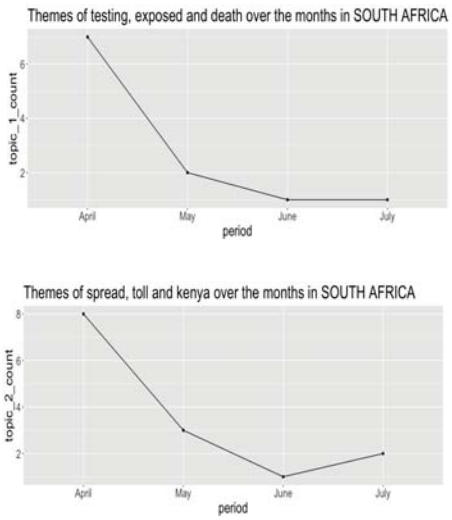

Fig. 8. Themes Overtime Between Topic 1 and 2 in SOUTH AFRICA.

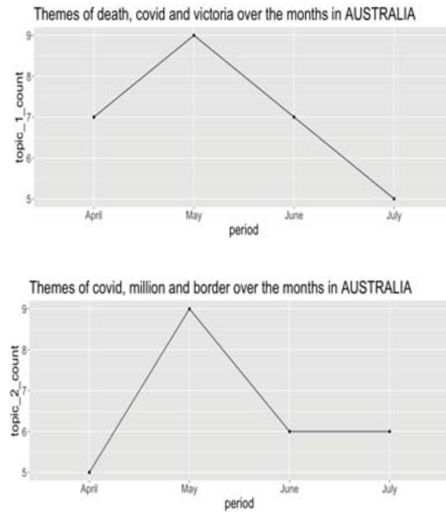

Fig. 9. Themes Overtime Between Topic 1 and 2 in AUSTRALIA.

VI. Acknowledgement

Research is partially funded by GOOGLE LLC COVID19 AI and Data Science for the Social Good Fund [Richard Aló, PhD, PI]

VII. Conclusion

Collecting and analyzing top stories on the novel coronavirus shed light on how the worldwide media have delivered health information during the COVID-19 crisis. Our study provides insights of mass media sources and their message impact. Our model identified source media top words for the countries under study during the period from April 1 through July 31, 2020. With 1,000 total media sources, we selected 125 top stories per region or country to yield tidy balance training data. The National Research Council of Canada's Word- Emotion Lexicon [word lists associated with anger, fear, anticipation, trust, surprise, sadness, joy and disgust] provided a tool to match our dataset. For example, it demonstrated how top stories involving word-emotion variances evolve over this time period in the countries listed. Unsupervised learning required human interpretation and these topics reflect the mass media sources. The term frequency-inverse document frequency tool provided a means to measure the importance of word in a top story leading to topic modeling allowing evaluation of words impacting behavior change. From these followed indications that perceived susceptibility and severity as the primary indicator of resistance to behavior change. We note that mass media campaigns carry a confident attitude that impacts and can increase resistance to behavior change.

References

[1] . G., Yang, C. H., Gutierrez, B., Wu, C. H., Klein, B., Pigott, D. M., … Scarpino, S. V. (2020). The effect of human mobility and control measures on the COVID-19 epidemic in China. Science, 368(6490), 493–497.

[2] Dehning, J., Zierenberg, J., Spitzner, F. P., Wibral, M., Neto, J. P., Wilczek, M., Priesemann, V. (2020). Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. Science, 369(6500), 0–10.

[3] Haushofer, J., & Metcalf, C. J. (2020). Which interventions work best in a pandemic? Science, 368(6495), 1063–1065.

[4] Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., … Vespignani, A. (2020). The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science, 368(6489), 395–400.

[5] World Health Organization. Coronavirus disease (COVID-19) outbreak URL:https://www.who.int/emergencies/diseases/novel-coronavirus-2019

[6] Cane, J., O'Connor, D., Michie, S. (2012). Validation of the theoretical domains framework for use in behaviour change and implementation research. Implementation Science, 7

[7] Rourke, B. M., Eccleston-turner, M., Phelan, A., Gostin, L. (2020). Policy opportunities to enhance sharing for pandemic research. 368(6492).

[8] Walker, P. G. T., Whittaker, C., Watson, O. J., Baguelin, M., Winskill, P., Hamlet, A., … Ghani, A. C. (2020). The impact of COVID-19 and strategies for mitigation and suppression in low- And middle-income countries. Science, 369(6502), 413–422.

[9] Zhang, J., Litvinova, M., Liang, Y., Wang, Y., Wang, W., Zhao, S., … Yu, H. (2020). Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. Science, 368(6498), 1481–1486.

[10] Yang, P., Li, W., Zhao, G. (2019). Language model-driven topic clustering and summarization for news articles. IEEE Access, 7, 185506–185519.

[11] Zou, X., Zhu, Y., Feng, J., Lu, J., Li, X. (2019). A novel hierarchical topic model for horizontal topic expansion with observed label information. IEEE Access, 7, 184242–184253.

[12] Gerlach, M., Peixoto, T. P., Altmann, E. G. (2018). A network approach to topic models. Science Advances, 4(7)

[13] Z. Liu, T. Qin, K. Chen and Y. Li, (2020). "Collaboratively Modeling and Embedding of Latent Topics for Short Texts," in IEEE Access, vol. 8, pp. 99141-99153.

[14] H. Jelodar, Y. Wang, R. Orji, and H. Huang, (2020) "Deep sentiment classi-fication and topic discovery on novel coronavirus or covid-19 onlinediscussions: Nlp using lstm recurrent neural network approach,"arXivpreprint arXiv:2004.11695.

[15] C. Ordun, S. Purushotham, and E. Raff,(2020) "Exploratory analysis of covid-19 tweets using topic modeling, umap, and digraphs,"arXiv preprintarXiv:2005.03082.

[16] K. Sharma, S. Seo, C. Meng, S. Rambhatla, and Y. Liu, (2020) "Coronavirus onsocial media: Analyzing misinformation in twitter conversations,"arXivpreprint arXiv:2003.12309.

[17] Mohammad, S. M., Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. Computational Intelligence, 29(3), 436–465. https://doi.org/10.1111/j.1467-8640.2012.00460.x