**COVID 19 Media Coverage and its Influence on Twitter Hashtag Participation**

1. **Introduction**

News about COViD-19 Pandemic has dominated the world’s attention in the first half of 2020. The Pandemic has almost secured the traditional media,despite its decline, a monopoly because of the high entry to reporting on the issue — grass-root media are not expected to gather data from a plethora of channels or to produce original content as efficiently or credibly as established news outlets.News about COViD-19 before May is almost always of negative tone and on similar subjects. All factors above constituted a natural experiment in which the origin, the topic importance, and the general psychological feature of news are controlled, all of which are factors that may alter consumer’s behaviour. The aim of the research is to understand what types of news content have grappled social media users’ attention and have influenced their narrative during the COViD-19 Pandemic and to provide insight into the general preference for news consumption and sharing in the event of a major social crisis.[[1]](#footnote-0)

1. **Data Acquisition and Explanation**

The data acquisition for this study is divided into two parts: first, we created a list of significant events; then we calculated the percentage of Tweets containing hashtags of interest using a Kaggle database.

For the significant event list, we used Cable News Networks’ (CNN) archive of daily coronavirus-related news, in which the “What You Need to Know” section listed 2 to 5 pieces of news the editors consider the most significant. We then categorised them into socio-economic news, policy news, public health news, and opinions of notable figures. All of the categories are further split by the subject on which they report.

As for the tweets, we generated the ratio of mentions of specific hashtags to the number of all coronavirus-related Tweets from a Kaggle database [[1]](https://www.kaggle.com/smid80/coronavirus-covid19-tweets-early-april) [[2]](https://www.kaggle.com/smid80/coronavirus-covid19-tweets). The database includes all Tweets that contain #coronavirus, #coronavirusoutbreak, #coronavirusPandemic, #covid19, #covid\_19 from March 12th to April 30th. [[2]](#footnote-1)Since two new hashtags were added to the database on March 17th, daily raw counts of “hashtags of interest” might be inflated since that date, hence we used daily “counts per one million Tweets” to quantify the popularity of a hashtag. We converted all hashtags to lowercase, dropped all special characters, and merged hashtags of similar meaning (e.g. #stayhome #stayathome #stayhomestaysafe) to extract the general idea of a tag.

1. **Data Analysis**

We constructed mathematical models to capture the relationship between the increase in hashtag participation and different types of events. First, we constructed a multiple regression model of #stayhome on all types of events, from which we obtained an R-squared value of 0.7093 and 7 significant factors. We then conducted backward BIC and backward AIC and arrived at our final model: regression on pubhealth\_world, pubhealth\_us, opinion\_other, opinion\_sci and socioecon\_world with a R-squared of 0.5632. The following is its corresponding diagnostic plots, from which we can confirm the validity of our model.

1. **Conclusion and Insight**

The study has made the following contributions: First, we have developed a robust tool which measures the relative popularity of a hashtag/hashtag group in a Tweet database.Second, we found that Twitter hashtag participation mainly responds to US domestic news, general world news regarding public health and opinion pieces. News about specific countries don’t yield a significant influence on hashtag participation.This is even the case for the #china hashtag. Statistically significant variables for the participation of #China are public health events in the US and scientific opinion pieces. These findings led us to conclude that Anglophone audiences are still predominantly concerned about US domestic public health news and socioeconomic news about the world in general, but cared little about situations of specific countries other than the US. This ironically mirrors the initial apathy towards COViD19 situation in China before it became a global pandemic, and further showed that even in an age of the internet, we remained isolated.

1. Unlike conventional methods that track likes, shares, and comments, we track the number of specific Twitter hashtags, since we believe hashtags are concise labels for a set of specific ideas. [↑](#footnote-ref-0)
2. We assumed the ratio of our “hashtags of interest” in the database was close to that in the general Tweet population. We have also limited our scope to Anglophone audiences, hence Tweets labelled in other languages were dropped. [↑](#footnote-ref-1)