

Business Problem

Hong Kong Protests: Dive in to the Unstable Time

The 2019 Hong Kong Protests sparked off by the Extradition Law Amendment Bill introduced by the Special Administrative Region (SAR) government in March this year. This issue has gone viral online and led to an increasing number of protestors and a series of planned demonstrations across the island.

Due to the newsworthy and contentious nature of the issue, a detailed analysis integrating social science research methodology and data analytics methods will benefit many people in the society, ranging from news media agency to sociologist researching on political issues.

As more and more people are using social media to express their views, it becomes more and more valuable and newsworthy to capture, analyse and report the public sentiments online as a whole.

Newsagents who want to stay relevant and competitive in today's world will have to leverage computational social science analytics to understand the public sentiment and get insights from it. For researchers, computational social science analytics provide accurate statistical and scientific analysis on the topic at hand.

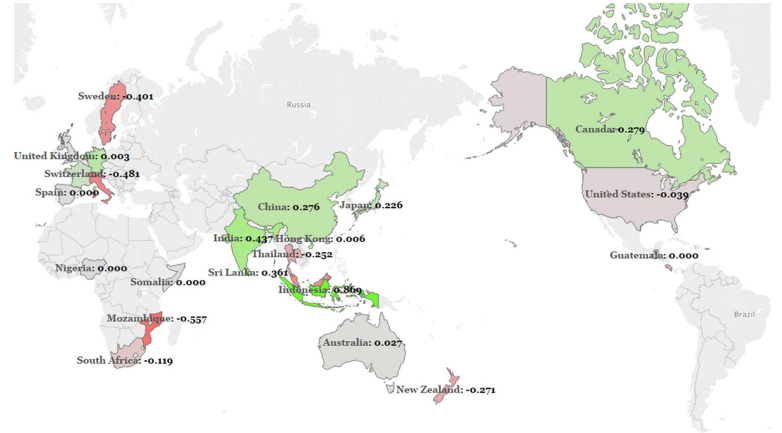


Figure 1: Tableau Information Dashboard - Sentiment Score by Country/Region

Data Analysis Methods

Data Extractions to Insights Generation

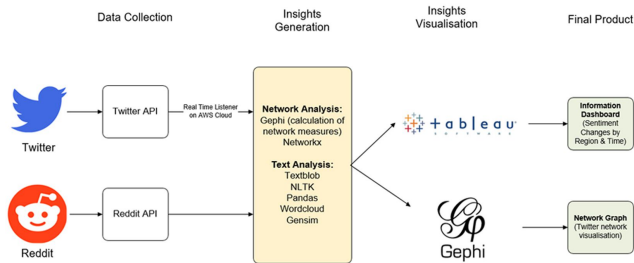


Figure 2.1: Data Analysis Flow Diagram (Illustration of Insights Generation Process)

We have scraped data from social media platforms like Twitter and Reddit using real time listener on AWS cloud to maximise the amount of data collected. After data manipulation and data cleansing, we have generated data sets to perform network analysis and text & sentiment analysis.

Tableau and Gephi are the main tools for visualising insights gathered. The final products are Tableau Information Dashboard (Shown in Figure 1) and network graph generated using Gephi. Various network measures are calculated for the team to provide more accurate interpretations of the graphs.



Figure 2.1: Data Analysis Flow Diagram (Illustration of Insights Generation Process)

Network Analysis

For Network Analysis, we have visualised the interactions between twitter users using Gephi and calculated the network measures at both macro and micro level. A series of network graphs of different dates have been generated.

At the micro level, we are able to identify different types of influencers in the network based on the centrality measures. For example, looking at the network graph of



Figure 3.3: Editorial Board CityU Twitter Profile

data extracted on 1 Oct 2019, the National Day of China, the twitter user that has the highest degree centrality is Editorial Board of City University of Hong Kong (@cityusuedb).

At the macro level, we are able to observe the changes of network structure and characteristics over time. Identify how time (different major events) can impact the network structure on Twitter.

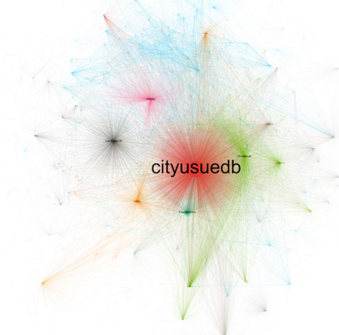


Figure 3.4: Network Graph of data extracted on National Day of China

Text Analysis

Our text analysis consists of both **sentiment analysis** and **topic extractions**.

For sentiment analysis, we have used vader polarity score (from NLTK) to observe overall sentiment of twitter users in different regions over time. To further analyse users' sentiment towards each topic, we have first extracted a list popular topics, and under each topic, we have conducted sentiment analysis as well as common words related to the topic which can be representative of a user's opinion.

In Figure 3.1, we have generated a WordCloud to visualise a list of popular words and phrases used by twitter users. Sentiment scores are assigned to all the phrases in WordCloud, green colour represents positivity in users' sentiment towards the topic and red represents negativity. In the WordCloud, we can see that the most green word is "human right", this suggests that users' general sentiments towards the topic "human right" are



Figure 3.1: Hong Kong Map-style WordCloud with Sentiment Score Assigned

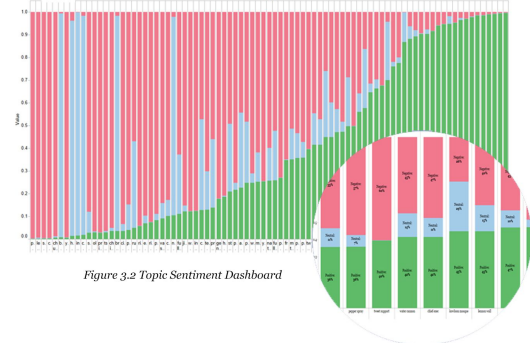


Figure 3.2: Topic Sentiment Dashboard

In Figure 3.2, a Topic Sentiment Dashboard is used to illustrate the list of popular phrases we have generated, and for each phrase, the proportion of tweets by sentiment are shown in a stacked bar chart. This dashboard allows people to understand the public sentiment of different specific topics related to the Hong Kong protests.

Conclusion

Looking into the Future

Coming to an end to the project and consolidates all the findings we have generated so far ...

Due to the limitations of the project faced, our team could only show the "tip of the iceberg" for this issue that is full of insights. We believed that this computational social science analytics project can be more valuable in the future as it has the scalability and wide application to different topics of social sciences. With more data from different sources, our project can be scaled up to provide more comprehensive insights. This systematic nature of this project allow us to apply the project into any social science related issues, providing accurate scientific & statistical findings to our client