DALHOUSIE UNIVERITY

Global Student Impact on Canadian Tuition

Project Report

Pratik Sakaria
pratik.sakaria@dal.ca
B00954261

Advisor: Prof. Stephen Brooks

Introduction:

In an era where social media platforms have become the central forums for public discourse, analyzing sentiments expressed on these platforms provides invaluable insights into public opinion on global events. This report focuses on the intricate dynamics of public sentiment during a significant geopolitical event—the United States' ban on Huawei, a leading Chinese telecommunications giant. The ban, rooted in national security concerns, has sparked widespread debate and discussion across the globe, particularly on social media platforms like ‘X’, formerly known as Twitter. Inspired by a foundational research paper [1], which proposed a novel framework for sentiment analysis using a lexicon-based approach coupled with geographic visualization, this report seeks to delve deeper into the public sentiment through enhanced and meticulously applied visual analytics.

The essence of this study lies in its strategic use of visualization techniques to dissect and understand the layers of public sentiment expressed on Twitter regarding the Huawei ban. By leveraging and building upon the methodologies outlined in the referenced research, we endeavor to create a multidimensional analysis through the following visualizations:

1. **Dot Distribution Map**: This visualization serves as the geographical backdrop of our analysis, showing the global distribution of tweets related to the Huawei ban. By mapping the origins of these discussions, we gain preliminary insights into the geographical nuances of the sentiment.
2. **Sentiment Dot Plot Map**: Elevating the analysis, this map integrates the geographical data with sentiment analysis, offering a vivid spatial representation of public opinion. This implementation allows us to visualize not just where the conversations are happening but also the predominant sentiments in those locations.
3. **Time Series Graph**: The temporal dimension of sentiment is captured in this visualization, which tracks the change in public opinion over the duration of the Huawei ban. This graph highlights the fluctuations in sentiment, providing clues to key events or announcements that might have influenced public opinion.
4. **Word Cloud Visualization**: By identifying and emphasizing the most frequently used words in the Twitter discussions, this enhanced visualization brings to light the central themes and concerns driving the public discourse. The word cloud acts as a lens, focusing on the essence of the collective sentiment.

Collectively, these visual tools not only offer a granular view of the public's reaction to the US ban on Huawei but also underscore the broader implications of such geopolitical decisions on international relations, technology markets, and public perception. Through a detailed examination of these visualizations, this report aims to contribute to the understanding of how digital platforms reflect and influence public sentiment on critical global issues.

Background:

The US-China trade war, characterized by escalating tariffs and trade restrictions, significantly impacted global economic and technological landscapes. A notable event in this conflict was the United States' decision to ban Huawei, a leading Chinese technology company, on national security grounds. This move, preventing Huawei from engaging with US companies, became a focal point of contention and discussion worldwide, particularly on Twitter. The platform served as a digital arena where individuals from various backgrounds voiced their opinions, concerns, and sentiments about the ban's implications for global trade, technology, and geopolitical relations.

Analyzing sentiments from Twitter discussions offers insights into public perception of the Huawei ban within the broader context of the US-China trade war. This approach enables us to understand the multifaceted views of the global community on this issue, reflecting concerns ranging from the future of international trade to the implications for technological innovation and security. By examining these online conversations, this report aims to provide a nuanced analysis of the public sentiment surrounding one of the most consequential developments in the US-China trade dynamics.

Data Acquisition and Preprocessing:

Obtaining relevant data for social media analysis is often a challenging task. For this project, limitations on direct access to Twitter data necessitated exploring alternative methods. The work by Stieglitz S. et al. [2] highlights these challenges and explores strategies for data acquisition from social media platforms. I evaluated and ultimately selected "twscraper," a third-party Python library, for its effectiveness in extracting pertinent tweet data relevant to the US ban on Huawei. Once acquired, the data underwent meticulous cleaning and pre-processing steps to ensure its relevance and authenticity for the visualization objectives. Techniques similar to those outlined by Pradha S. et al [3] were employed, focusing on removing irrelevant information, standardizing text formatting, and handling inconsistencies in user-provided data. More specifically the data was cleaned for special characters, stop words and data was filtered to only include English tweets. The locations were cleaned and the timestamp was processed to have the date and time stored in two separate columns in the desired format.

Visualization Methodology:

This project utilized a combination of informative visualizations to analyze public sentiment on Twitter regarding the US ban on Huawei.

- **Dot Distribution Map Visualization**
  The dot distribution map depicts the geographic distribution of tweets related to the US ban. Leveraging the OpenCageData API, I performed forward geocoding to translate tweet-derived locations into precise geographical coordinates. This process aligns with geocoding techniques implemented in research by Antoniou et al. (2018) [4]. The cleaned locations were then plotted against the co-ordinates obtained from geocoding and the resulting map provides a visual illustration of the global reach of the discussion on the Huawei ban.

Figure 1: Dot Distribution Map

- **Sentiment Dot Distribution Map Visualization**
  Building upon the dot distribution map, the sentiment dot distribution map incorporates sentiment analysis. Tweets were classified as positive, negative, or neutral using a lexicon-based approach. This approach is similar to the methodology employed by N. H. Khun et al. (2019) [1] in their research on visualizing Twitter sentiment during the US ban on Huawei. Existing sentiment lexicons, such as those described by Mohammad and Turney (2010) [5], could be leveraged to categorize tweets based on their sentiment. The map displays dots colored according to the sentiment of the corresponding tweet, allowing for a quick visual assessment of how sentiment is distributed across different regions.

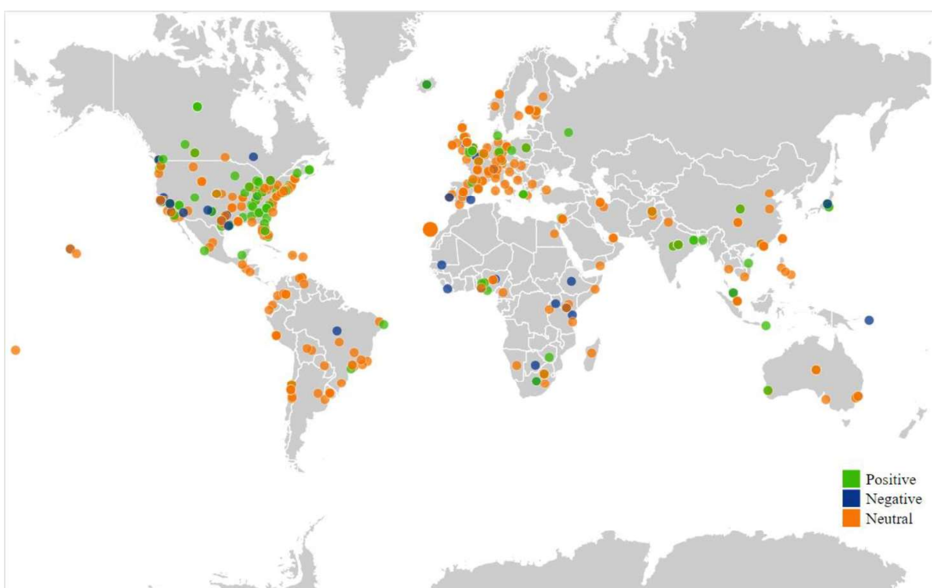

Figure 2: Sentiment Dot Distribution Map

- **Time Series Graph Visualization**
  The time series graph tracks how sentiment evolved over time during the period surrounding the US ban on Huawei. This approach is similar to the work by Hoang et al. (2019) [6] who utilized time series analysis to understand sentiment variations in social media data. By analyzing this graph, we can identify if public sentiment shifted significantly throughout this event. The data points on the graph represent the sentiment of each tweet, analyzed by using established sentiment analysis techniques, plotted against the day/date on which the tweet was created.

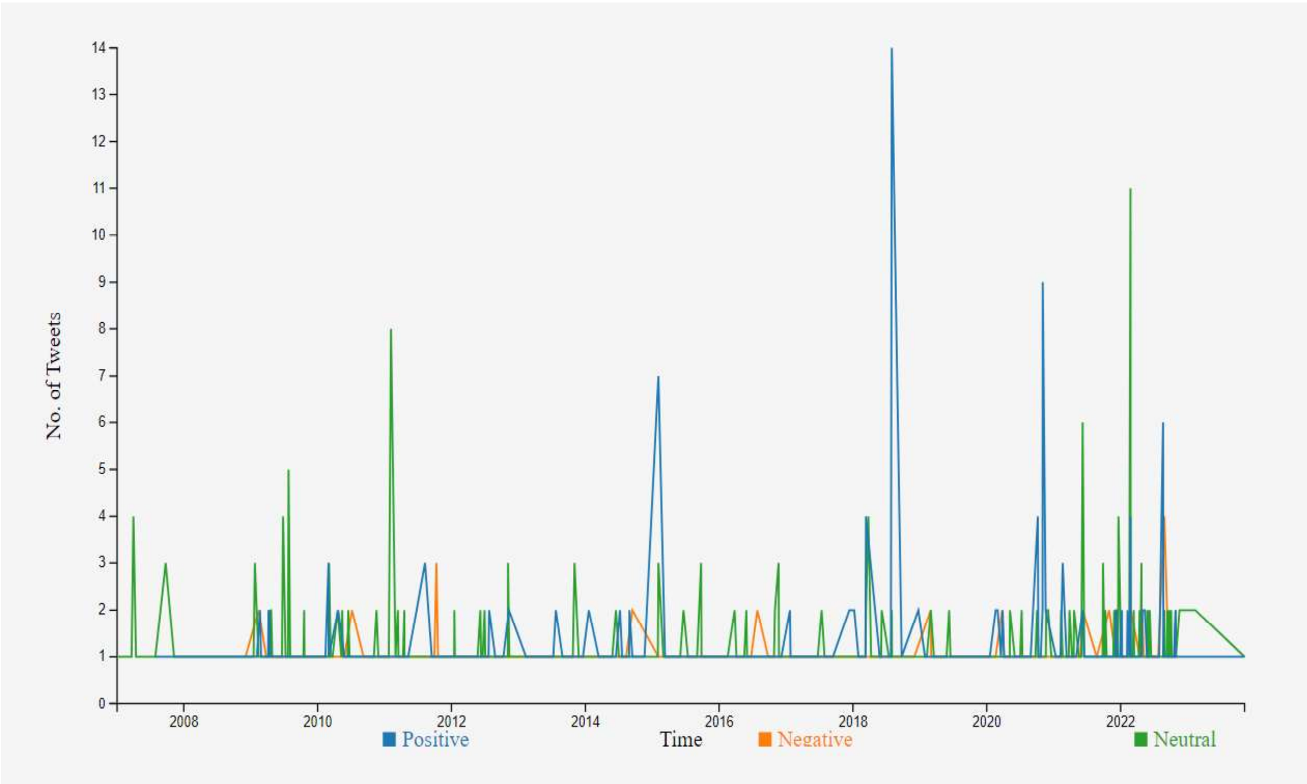

Figure 3: Sentiment Time-Series Plot

- **Word Cloud Visualization**
  The word cloud visualization captures the essence of the public dialogue on Twitter regarding the US ban on Huawei. The creation process involved techniques outlined in the research by J. Xu et al. (2016) [7] on word cloud generation. This included data extraction, data cleaning to remove non-essential elements, and frequency analysis to identify prominent themes and topics. The word cloud displays words with a larger font size based on their frequency of occurrence, highlighting the most prevalent discussion points.

Figure 4: Word Cloud Visualization

Discussion and Analysis:

By analyzing the implemented visualizations together, we gain valuable insights into public sentiment on Twitter regarding the US ban on Huawei.

Dot Distribution Map: The dot distribution map as seen in Figure 1, reveals the geographic spread of tweets, allowing us to identify areas with high tweet volume.

- ***Global Engagement***: Based on the distribution of dots in Figure 1, it can be inferred that the US ban on Huawei is a subject of global discourse, not just confined to the US or China. This reflects the international implications of the ban, touching on issues like global trade, technology markets, and geopolitical strategy.

- ***Concentration of Discussion***: The dots seem to be more densely clustered in North America, Europe, and parts of Asia. These regions likely have a higher population of Twitter users, and they might be more directly affected by or interested in the ramifications of the ban due to economic ties with both the US and China.
- ***Areas of Limited Discussion***: Some areas appear to have fewer or no dots, which might suggest less engagement with the topic on Twitter. This could be due to several factors, such as lower Twitter usage, less relevance of the topic to the region, or even limited internet access.
- ***Limitations***: The map may not capture the full scope of the global conversation if Twitter is censored or less popular in certain regions, like China, where the platform is blocked, and domestic platforms dominate social media interactions.

Sentiment Dot Distribution Map: The sentiment dot plot map as seen in Figure 2, depicts the sentiment distribution across different regions, helping us visualize areas with predominantly positive, negative, or neutral sentiment towards the ban.

- ***Concentration of Negative Sentiment***: Blue dots, representing negative sentiment, appear to be prevalent, particularly in North America and parts of Asia. This could imply a significant level of concern or disagreement with the US ban among Twitter users in these areas.
- ***Neutral Sentiment***: Orange dots, showing neutral sentiment, are scattered across various continents but are less concentrated than negative or positive sentiments. This indicates that while there is a discussion on the topic, a portion of the conversation may be informational or unbiased in nature.
- ***Positive Sentiment***: Green dots signifying positive sentiment seem to be less widespread than negative ones. The distribution of positive sentiment in specific regions could suggest support for the US ban on Huawei, possibly due to national security concerns or geopolitical alliances.

Time Series Graph:

- ***Pre-2018 Activity***: Prior to 2018, there is minimal activity across all sentiments. The green (neutral) line suggests a consistent, low level of neutral discussion about Huawei, which could be related to general news or conversations about the company without strong sentiment.
- ***Spike in Positive Sentiment in 2018***: The blue line shows a dramatic spike in positive sentiment in 2018. This could correspond with the period when the US announced its actions against Huawei, indicating a surge of tweets that viewed the ban positively, possibly reflecting approval of the US government's decision on the grounds of national security or competition in the tech industry.
- ***Post-2018 Trends***: After the 2018 peak, the graph shows continued engagement with the topic, with noticeable fluctuations. Positive sentiment remains higher than neutral and negative, suggesting sustained support or agreement with the ban over time.
- ***Negative Sentiment Trends***: The orange line, indicating negative sentiment, shows less frequency and intensity compared to positive sentiment. There are smaller spikes of negative sentiment, which may correspond with specific developments in the US-Huawei saga, such as corporate responses or legal challenges.
- ***Neutral Sentiment***: Green (neutral) sentiment maintains a steady presence with occasional peaks, suggesting ongoing discussions that may be reporting facts or updates without expressing strong opinions.

- ***Recent Activity***: Towards the end of the timeline, there's an uptick in neutral discussion, which might indicate recent developments in the situation that are being discussed without a clear consensus on sentiment.

Word Cloud Visualization: The improved word cloud in Figure 4, highlights prominent topics of discussion, identifying the key issues and concerns surrounding the US ban on Huawei. Following is the brief analysis of the word cloud visualization:

- ***Key Political Figures***: The prominent display of "China," "President," "Trump," and "Biden" suggests that conversations are heavily focused on the political figures involved in the decision-making process or affected by the ban. This indicates that the issue is seen within a political context, potentially associated with the policies of different US administrations.
- ***Technological Aspect***: Words like "tech," "phones," and "race" hint at discussions around the technological implications of the ban, possibly reflecting concerns over the competition in the tech industry and the future of telecommunications.
- ***Legal and Economic Dimensions***: Terms such as "case," "trade," and "market" indicate that the legal and economic ramifications of the ban are also significant points of discussion. This could relate to the impact on global markets and international trade laws.
- ***Emotional and Reactionary Language***: The presence of more emotionally charged or casual swear words (though it could be a part of different phrases) could point to strong opinions and emotional reactions within the public discourse.

Conclusion

This report focused on implementing and modifying visualizations from a research paper by N. H. Khun et al. (2019) [1] to analyze public sentiment on the US ban of Huawei using Twitter data. The implemented visualizations, including dot distribution maps, sentiment dot plot maps, time series graphs, and word clouds, provide a valuable means to explore and understand public opinion on social media. This approach aligns with the growing body of research that utilizes social media data analysis and visualizations to understand public sentiment surrounding significant events [8, 9, 10].

By analyzing the sentiment distribution, geographic spread, and evolution of public discourse over time, we gain a more comprehensive picture of public reaction to the US ban on Huawei. Future work, as discussed in the next section, can involve expanding the analysis and incorporating more advanced techniques to gain even deeper insights.

Future Work:

Future work can involve several avenues to further explore public sentiment on the US ban of Huawei:

Expanding the Analysis to Multiple Languages: Sentiment analysis can be applied to tweets in other languages to gain a more global perspective on public opinion [10].

Utilizing Machine Learning for Sentiment Analysis: Machine learning models can offer more nuanced sentiment classification compared to lexicon-based approaches [8].
Correlating Sentiment with Other Factors: Analyzing correlations between sentiment and user demographics or tweet source (e.g., news outlets vs. individual users) can provide deeper insights [9].

References:

[1] N. H. Khun and H. A. Thant, "Visualization of Twitter Sentiment during the Period of US Banned Huawei," *2019 International Conference on Advanced Information Technologies (ICAIT)*, Yangon, Myanmar, 2019, pp. 274-279, doi: 10.1109/AITC.2019.8921014. Available Online: https://ieeexplore-ieee-org.ezproxy.library.dal.ca/document/8921014 [Accessed April 4th, 2024]

[2] Stieglitz S., Mirbabaie M., Ross B., Neuberger C.,"Social media analytics – Challenges in topic discovery, data collection, and data preparation", International Journal of Information Management, Volume 39, 2018, Pages 156-168, ISSN 0268-4012, Available Online: https://doi.org/10.1016/j.ijinfomgt.2017.12.002. [Accessed April 4th, 2024].

[3] S. Pradha, M. N. Halgamuge and N. Tran Quoc Vinh, "Effective Text Data Preprocessing Technique for Sentiment Analysis in Social Media Data," *2019 11th International Conference on Knowledge and Systems Engineering (KSE)*, Da Nang, Vietnam, 2019, pp. 1-8, doi: Available Online: https://ieeexplore-ieee-org.ezproxy.library.dal.ca/document/8919368 [Accessed April 4th, 2024].

[4] Melo, Fernando & Martins, Bruno. (2016). "Automated Geocoding of Textual Documents: A Survey of Current Approaches: Automated Geocoding of Textual Documents." *Transactions in GIS*. Available Online: https://onlinelibrary.wiley.com/doi/epdf/10.1111/tgis.12212 [Accessed April 4th, 2024].

[5] Mohammad S. and Turney P.,"Crowdsourcing a Word-Emotion Association Lexicon", *Computational Intelligence*, Available Online: https://arxiv.org/pdf/1308.6297.pdf [Accessed April 4th, 2024].

[6] Ibrahim N.F., Wang X., "Decoding the sentiment dynamics of online retailing customers: Time series analysis of social media". *Computers in Human Behavior, Volume 96, 2019, Pages 32-45, ISSN 0747-5632,* doi: https://doi.org/10.1016/j.chb.2019.02.004 . [Accessed April 4th, 2024].

[7] J. Xu, Y. Tao and H. Lin, "Semantic word cloud generation based on word embeddings," *2016 IEEE Pacific Visualization Symposium (PacificVis)*, Taipei, Taiwan, 2016, pp. 239-243, Available Online: https://ieeexplore-ieee-org.ezproxy.library.dal.ca/document/7465278 [Accessed April 4th, 2024]

[8] R. Sharma, D. Lowe, K. Somani and B. Galhotra, "A Study on Lexicon Based Techniques of Twitter Sentiment Analysis*," 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, 2022, pp. 562-566, Available Online: https://ieeexplore-ieee-org.ezproxy.library.dal.ca/document/9785231 [Accessed April 4th, 2024]

[9] V. Sathya, A. Venkataramanan, A. Tiwari and D. D. P.S., "Ascertaining Public Opinion Through Sentiment Analysis," *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, India, 2019, pp. 1139-1143, Available Online: https://ieeexplore-ieee-org.ezproxy.library.dal.ca/document/8819738 [Accessed April 4th, 2024]

[10] Chen, W., Wang, Y., & Yang, S. (2014). "Empirical comparison of sentiment analysis techniques for social media". *International Journal of Advanced and Applied Sciences, 5(4) 2018,* Pages: 115-123. Available Online: http://science-gate.com/IJAAS/Articles/2018/2018-5-4/15%202018-5-4-pp.115-123.pdf [Accessed April 4th, 2024]