Exploring Agenda Setting Through Social Media by Text Mining and Data Analysis: A Case Study of Roe V. Wade Overturn and Discussions on Twitter

Sihui Zheng[1,a,*], Yiling Zhang[2,b], Leran Wang[3,c], Sitian Wu[4,d], Boyin Chen[5,e]

[1]*Stony Brook University, Stony Brook, 11790, USA*
[2]*Jiangsu Tianyi High School, Wuxi, 214000, China*
[3]*Macau University of Science and Technology, Macau, 999078, China*
[4]*HuaerZizhu Academy, ShangHai, 201100, China*
[5]*Sanda University, ShangHai, 200120, China*
*a. sophiazheng777@gmail.com, b. elainezhang1024@163.com, c. 1756229035@qq.com,
d. 29628883705@qq.com, e. 1248190348@qq.com*
*corresponding author

Abstract: Traditional theories of agenda setting address that mainstream media (news reports) can influence the public's perceived importance of agenda. However, as the rise of social media, such as Twitter, penetrated people's lives, they became one of the critical tools for receiving and exchanging information. The study examines whether social media, such as Twitter, can influence the public agenda. On June 24, 2022, Roe V. Wade was announced to be officially overturned, along with uproar on the Internet. However, a month ago, on May 2, 2022, a few news agencies reported that the decision was leaked from the Supreme Court, and the discussion on social media was not yet significantly heated. This study attempts to understand the overturn of Roe V. Wade through Twitter, which yields an agenda-setting effect by raising the public's attention. The authors used Twitter's streaming media and RStudio to collect and select a total of 10,000 tweets in two time periods, from May 3, 2022, to May 17, 2022, and June 24, 2022, to July 17, 2022, respectively, with 5,000 tweets in each period for data analysis. The results showed that the total number of retweets from the 5,000 Twitter users during the official version was 23.2 times higher than during the leaked version. In addition, after the release of the official version, people showed significantly higher levels of emotional expression and anger than during the leaked version.

Keywords: Roe V. Wade, Abortion, Feminism, Twitter, data mining

1.   Introduction

With the surge of the Internet and technology, social media has skyrocketed and dominated our life. It serves various purposes, such as personal entertainment, commercial advertising, political advocacy, etc. The number of people who receive news from social media is also growing. According to a Pew Research report in 2016, roughly 64% of U.S. adults get information on social media, and 18% do so often [1]. As people pay more attention to social media, researchers have investigated its possible impacts on the public [2]. Apart from changing the way people communicate, social media

transforms public administration towards a new format that is more transparent, more actively participated by citizens, and with closer collaboration [3]. In this way, the public's engagement becomes more important in political events and may be influenced by information on the network.

For example, Roe V Wade might present how the media influenced the participants. During the event, a leaked version was exposed on the internet on May 2, 2022, and only a few news agencies brought it up. The public attention was not heated significantly. After the release of the official version on June 24, 2022, significant media and newspapers competed to report [4]. Discussions on social media, such as Twitter, became highly intense. Many celebrities commended and protested online, such as Michelle Obama. There has been research on the development of the Roe V Wade movement and its impacts [5]. However, whether the Twitter discussion impacts the public's attention is still unknown. This research conducts a comparative analysis between the critical actors in the two versions to determine the changes in people's engagement levels and emotions. The research question of this paper is whether the intensity of social media discussion will impact people's perceived importance of a social event.

2. Literature Review

2.1. Agenda Setting Theory

Agenda Setting was first observed by a journalist, Walter Lippmann, in the 1920s. He pointed out that the media dominates over the creation of pictures in our heads and that the public reacts not to actual events but to the pictures in our heads. Researchers Maxwell McCombs and Donald Shaw have advanced this concept. For example, they observed the capacity of mass media to influence the voters' opinions on the presidential campaign of 1968. In their book *Emergence of American Political Issues,* they stated that it is abundantly evident that editors and broadcasters play an important part as they go through their day-to-day tasks in deciding and publicizing news [6]. The concept of agenda setting is that the press selectively chooses what we see or hear in the media.

2.2. Roe V. Wade History

Roe v. Wade, 410 U.S. 113 (1973), is a great example. It was a landmark decision of the U.S. Supreme Court in which it ruled that the Constitution of the United States conferred the right to have an abortion [7]. By a vote of seven to two, the court justices ruled that governments lacked the power to prohibit abortions. Instead, they judged that the U.S. constitution protected a woman's right to terminate her pregnancy.

However, on May 2, 2022, the liberal-leaning POLITICO published a story confirming that Roe V. Wade would be overturned, which immediately drew attention from all sides. Many people attach great importance to the follow-up development of this case and have published their views. Later, Justice Alito's 98-page draft majority opinion for Dobbs, overturning Roe V. Wade, was posted online by POLITICO. Finally, on June 24, 2022, the U.S. Supreme Court decided by a 5-4 vote to formally overturn Roe v. Wade, a case that has existed for half a century.

Following the ruling, thousands of protestors defending abortion rights thronged the streets outside the U.S. Supreme Court. Politicians and groups on both sides of the debate are preparing for challenges to the ruling. The court's three liberal justices issued a jointly authored dissent in light of the verdict." Whatever the scope of the coming laws, one result of today's decision is inevitable: The curtailment of women's rights and their status as free and equal citizens," they wrote [8].

Outside the U.S., some world leaders have slammed the ruling. Canadian Prime Minister Justin Trudeau called the verdict "horrific," while British Prime Minister Boris Johnson called the decision "a big step backward", reported AFP. French President Emmanuel Macron also condemned the ruling, saying it challenged women's freedoms. And our research will be focused on how these social media

and news media attention alter people's engagement with such an event.

2.3. Twitter as a Medium for Social Movements

Twitter is a prevalent platform for people to share their experiences and perspectives. Due to the exploding number of users and the powerful social penetration of social media, Twitter has gradually penetrated every corner of social life. It can escalate a regional hotspot into an international one. Because Twitter is global and international, the public can be the first to find out what is happening around the world through Twitter. The emergence and development of social movements cannot be achieved without the intervention and shaping of social media such as it. Twitter enables the public to express their personal opinions, and the function of retweets and replies allows users to connect, interact, and initiate movements. Twitter is influential in people's views and can form a form of a march and influence the government's decision. The impact of Twitter on social movements is continuous and far-reaching. It can either push a social movement in an increasingly positive direction or become a tool to exacerbate social tensions and oppose state policies. Therefore, the research uses Twitter to explore how media affects users' engagement in a social event, such as Roe V. Wade's overturn.

2.4. Machine Driven Classification of Open-ended Responses (MDCOR)

The research will apply Machine Drive classification of Open-ended Responses (MDCOR) for data analysis. It's a software application to classify text responses in academic studies and analyze the qualitative data quantitatively. MDCOR enables researchers to "access the fully classified responses" in traditional quantitative analyses [9]. In addition, it analyzes open-ended responses, such as post content on Twitter on this research.

3. Methodology

This study compares the engagement and emotion of Twitter users after the leaked version and after the official version. In addition, the research compares the number of tweets, comments, and likes numbers of tweets during two different time frames. The authors utilize Twitter's Streaming and R studio to collect data on people's tweets on Twitter. The authors also use MDCOR to analyze the qualitative contents.

3.1. Data Collection

The research utilized Twitter's Streaming Application Programming Interface to extract data. The authors used imported the Twitter dataset to R Studio, using R Language to collect tweets containing the five hashtags: "abortion rights", "my body my choice", "reproductive justice", "reproductive rights", and "Roe Vs Wade" from two-time frames corresponding to the leaked version and the official version. The two periods are: May 2, 2022 to May 9, 2022, and June 24,2022 to July 1, 2022. As it is shown in graphic 1, The author started as "2022-05-03T00:00:00Z", and the end as "2022-05-17T10:00:00Z". Then, data for this first period is stored. The second data set comes out after repeating the method again and changing the time (start as "2022-6-24T00:00:00Z, end as "2022-07-01T00:00:00Z). Then, data analysis methods in Excel are used for randomly picking out 5000 posts for each time period. The 5000 posts during May 2, 2022 to May 9, 2022 is Data1, while the 5000 posts during June 24,2022 to July 1, 2022 is Data2.
Team 1 Friday: Search one
#Livestreamshopping team
library（academic twitter）

```
RoeVWade < -
    get_all_tweets(
        query = c("abortionrights", "mybodymychoice", "reproductivejustice", "RoeVsWade") ,
        start_tweets = "2022-05-03T00:00:00Z",
        end_tweets = "2022-5-17T00:00:00Z"
        file = "RoeVWade1",
bind_tweets = TRUE ,
Data_path = "data/RoeVWade1"
bearer_token
```
Graph1: Code for Fetching Twitter Data

3.2.   Data Analysis with Excel

After sorting out data, general statistical calculations help reflect the engagement level of the viewers. Based on 5000 tweets extracted from each period, the excel sheet calculates the total retweet, replies, like, and quote numbers. Then, divide the real numbers by 5000 and generate average numbers of retweets, replies, likes, and quotes. Comparison between the average during the leaked version and the ones during the official version then shows the change—-if the standard gets more extensive during the official version, it shows that agenda setting brings up people's attention. If the two data sets are roughly the same, there may be no change in engagement. If the average is lower during the leaked version, the release of the official version may negatively affect citizens' meetings.

3.3.   Data Analysis with MDCOR

As basic calculations in the excel sheet reflect engagement, the Machine Learning Program Machine Driven Classification of Open-ended Responses (MDCOR) helps provide further interpretation of people's emotions, such as outrageous level. Using a quantitative method, this program helps to change qualitative data (the contents of tweets) into qualitative data.

The program is performed in the following steps: 1. Load data1 and data2 separately, which are all the open-ended responses. 2. Conduct text cleaning/mining; some open-ended responses might be excluded if their content contains unique or outlier words that can not be classified. The program also yields ten keywords that appear most frequently, listing out along with their frequency. 3. Remove the most common words (optional). 4. Select machine learning sampling parameters. 5. Execute metrics for an optimal number of code selections to generate graphs of correlations. 6. Select the optimal number of codes. And 7. Execute MDCOR. After running the program, MDCOR provides researchers with a visual and interactive representation of the most common and frequently used words in the open-ended responses.

4.   Results

4.1.   Twitter Users' Engagement Level

After collecting 5000 users' posts during two different time frames, the authors calculated the sum and average of retweets, replies, likes, and quotes. According to Table 2, the average Retweets for the leaked version was 417.3524, while that for the official version was 8508.8002. When compared, the official version has 20.39 times the amount of retweets as the leaked version. Though retweet numbers in data 2 are highly above that of data 1, the number of likes, quotes, and replies varied in the opposite direction: the average responses, preferences, and sections of tweets during the official version are all slightly lower than the ones during the leaked version.

Table 1: Sums of Retweets, Replies, Likes, and Quotes for Tweets About Both Versions.

	Sum of Total Retweets	Sum of Replies	Sum of Likes	Sum of Quotes
Data 1 (Leaked Version)	20,86,762	1,657	17,460	394
Data 2 (Official Version)	42,544,001	1,267	11,732	258

Table 2: Averages of Retweets, Replies, Likes, and Quotes for Tweets About Both Versions.

	Average of Total Retweets	Average of Replies	Average of Likes	Average of Quotes
Data 1 (Leaked Version)	417.3524	0.3314	3.4292	0.0788
Data 2 (Official Version)	8508.8002	0.2534	2.3646	0.054190296

4.2. Twitter Users' Emotions

After executing MDCOR on Data 1 and 2, authors obtained frequently used words and intertropical maps. Across the five topics in data 1, "woman", "my body my choice", "abortion right", "abortion s health care", "roe vs. wade", "reproductive justice", and "women's rights" were common keywords produced by the hashtags. After ruling them out, authors found that the most frequently appeared words are "reproductive rights", "prochoice", "rapist", etc. (Figure 6). Topic 1, 2, and 3 focus on justice and rights, while topic 4 emphasizes abortion healthcare and claims that abortion rights are human rights. Topic 5 refers to the supreme court. Across the five issues during the official version (data 2), effective keywords involved "woman", "life", "abortion", "my body my choice", "rights", and "roe vs. wade." Discussion on the official version also involved discussion on specific regions/states (Figure 7), such as "Ohio (Figure 9)." Moreover, as shown in the intertropical distance maps, the distances between circles in the two datasets are similar, indicating a roughly same level of relationship between major topics. In Data 2, words like "must" frequently appear (Figure 10), meaning people are more determined. Additionally, both datasets appear the profanity "fuck." However, the word appears more frequently during the official version (Figure 8) than the leaked version (Figure 2).
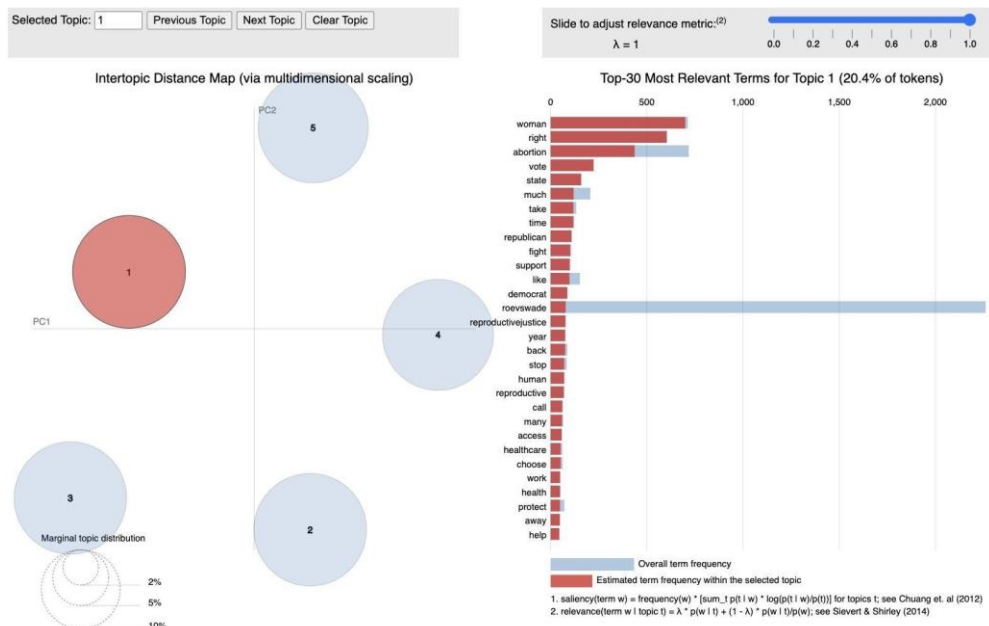
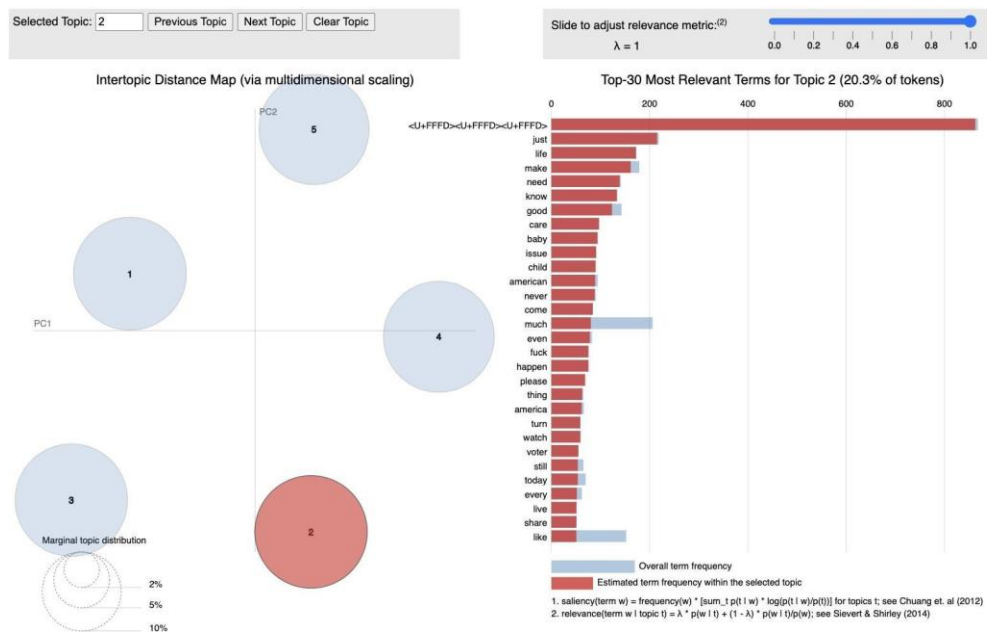Figure 1: Intertopic Distance Map and Top Keywords for Topic 1 in Data 1.

Figure 2: Intertopic Distance Map and Top Keywords for Topic 2 in Data 1.
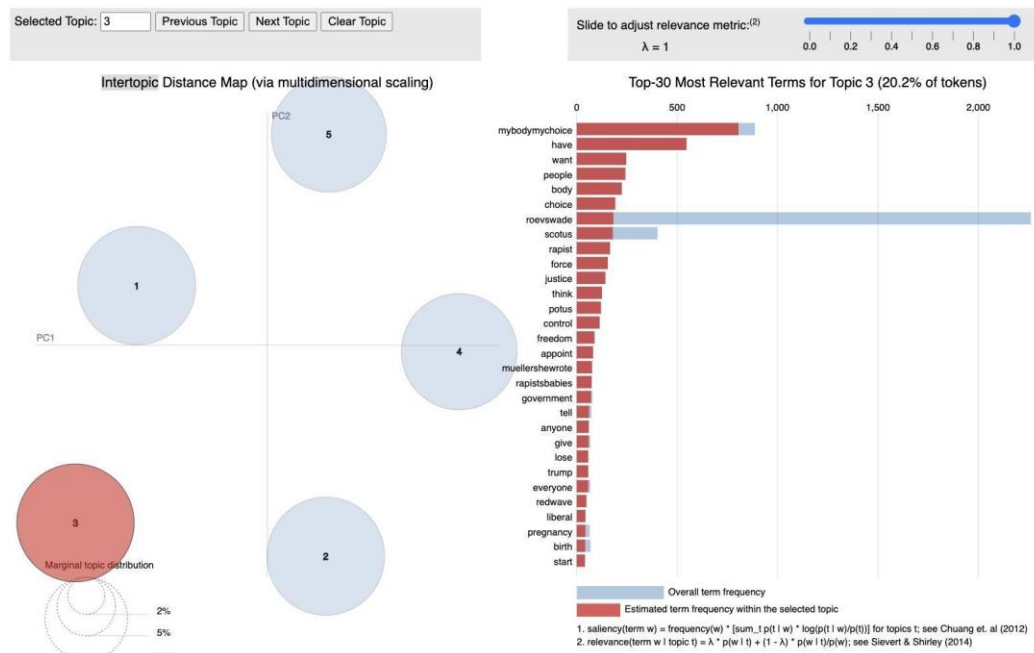
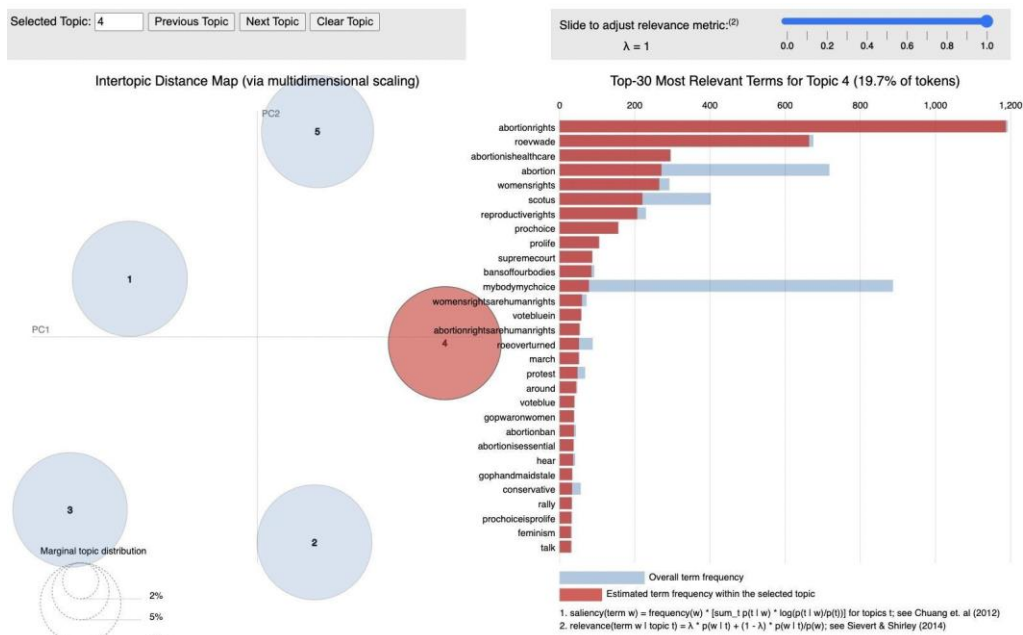Figure 3: Intertopic Distance Map and Top Keywords for Topic 3 in Data 1.

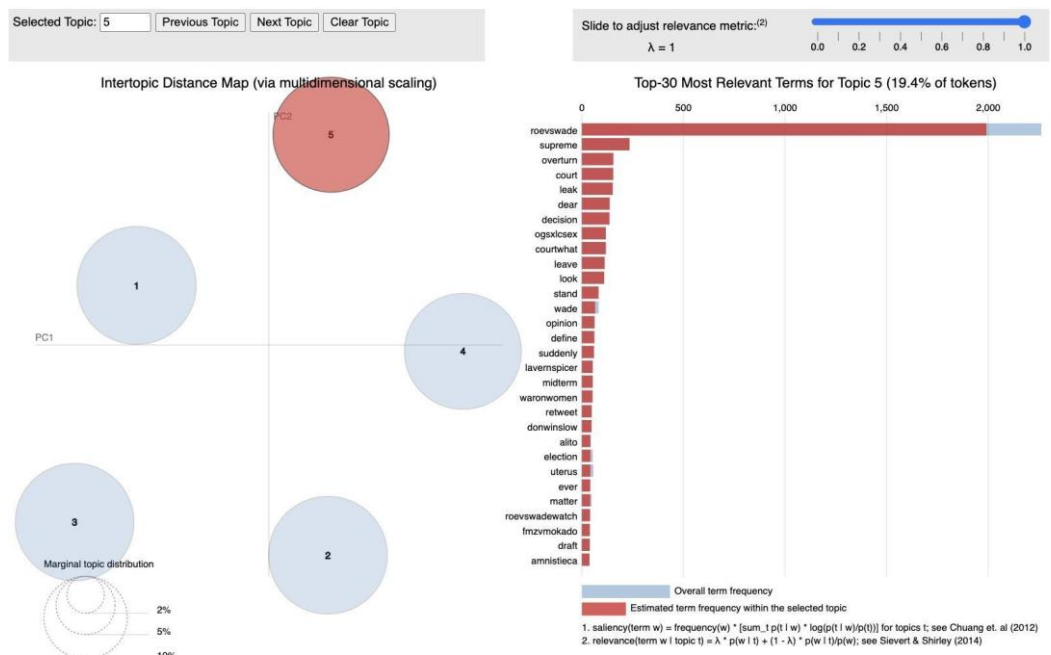Figure 4: Intertopic Distance Map and Top Keywords for Topic 4 in Data 1.

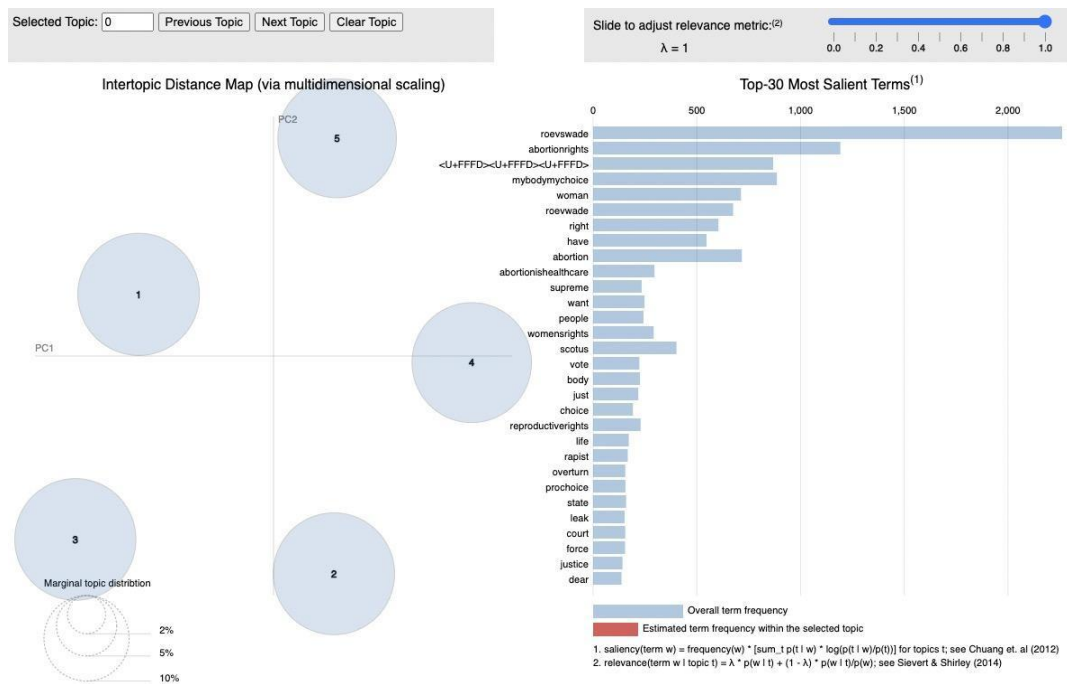Figure 5: Intertopic Distance Map and Top Keywords for Topic 5 in Data 1.

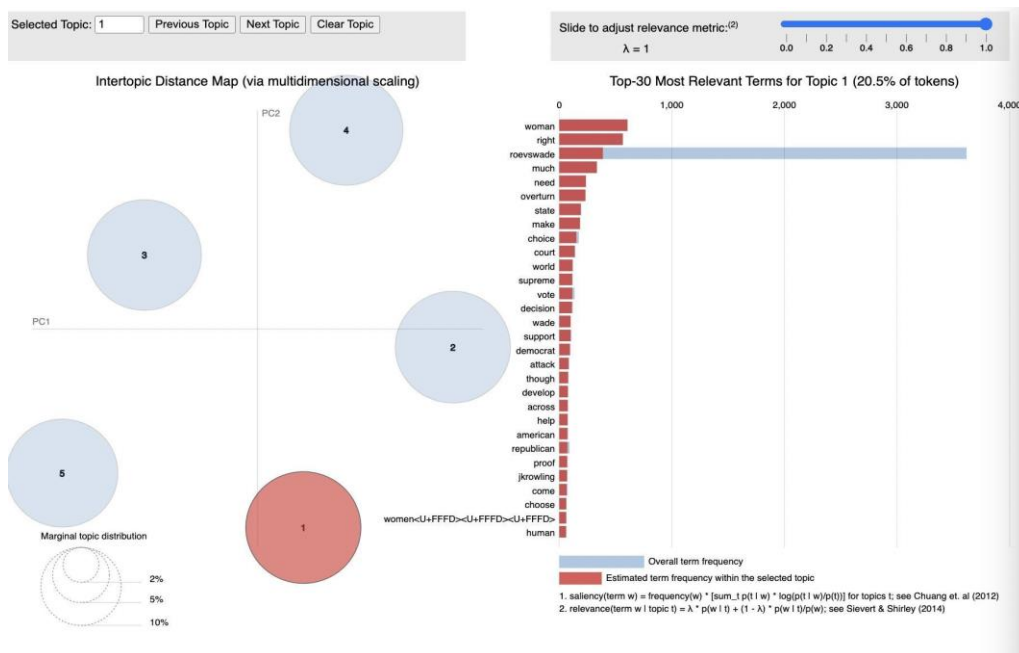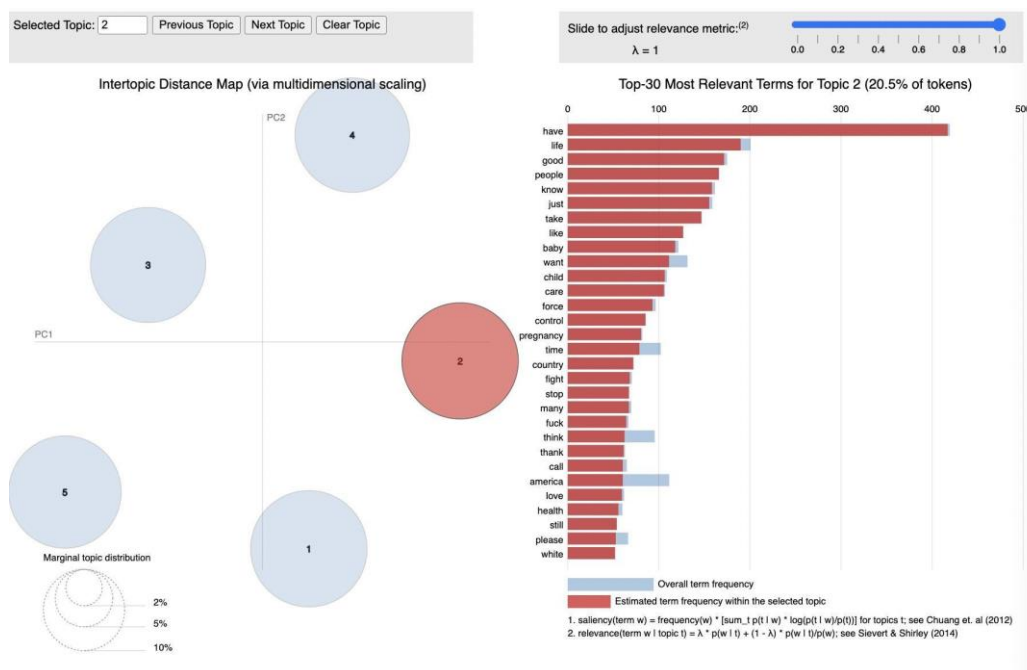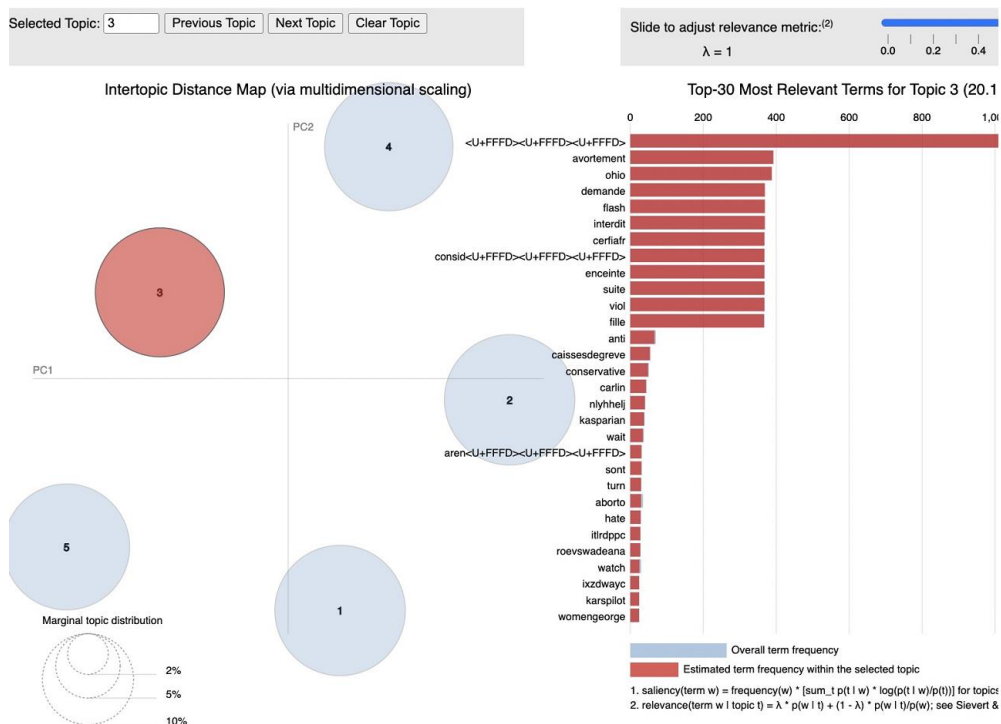Figure 6: Intertopic Distance Map and Overall Top Keywords in Data 1.

Figure 7: Intertopic Distance Map and Top Keywords for Topic 1 in Data 2.

Figure 8: Intertopic Distance Map and Top Keywords for Topic 2 in Data 2.

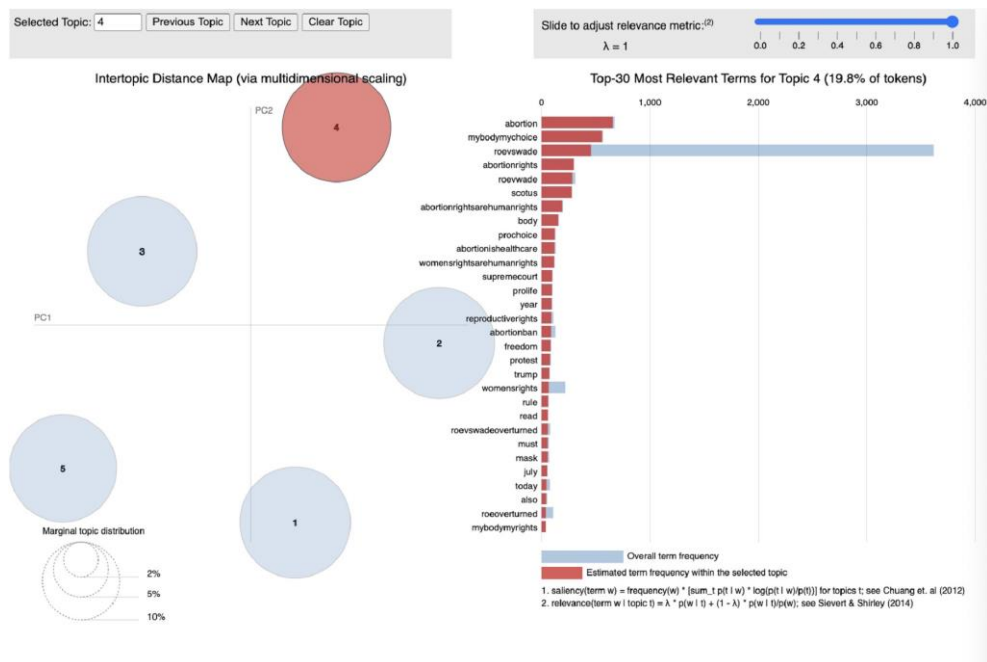Figure 9: Intertopic Distance Map and Top Keywords for Topic 3 in Data 2.

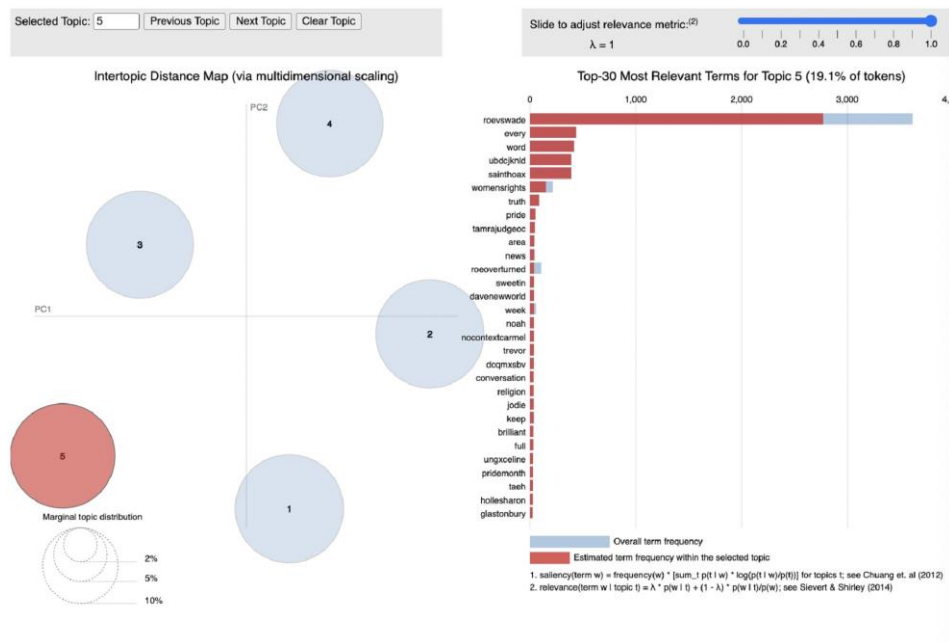Figure 10: Intertopic Distance Map and Top Keywords for Topic 4 in Data 2.

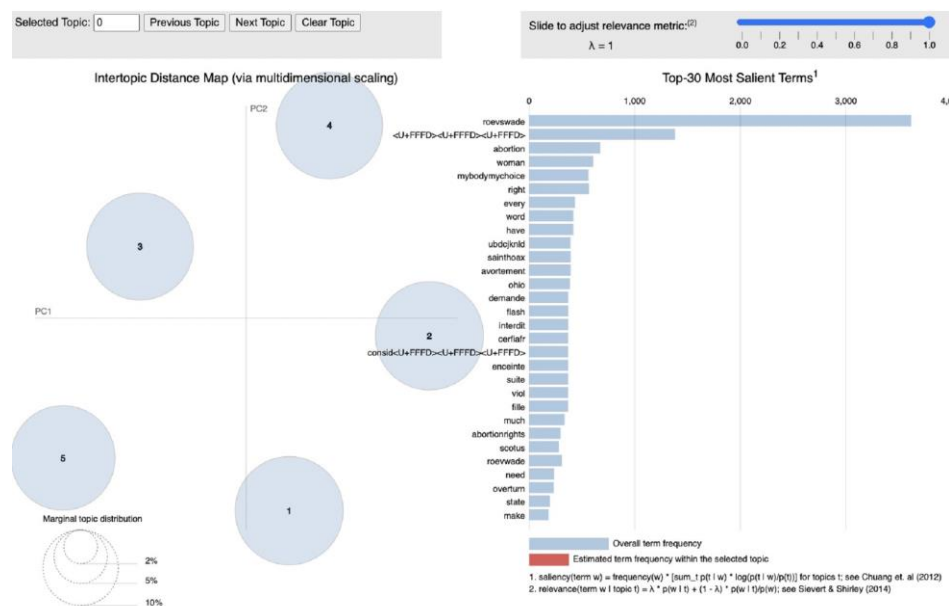Figure 11: Intertopic Distance Map and Top Keywords for Topic 5 in Data 2.

Figure 12: Intertopic Distance Map and Overall Top Keywords in Data 2.

5. Discussion

During the official version, the number of retweets is 20.39 times that of the leaked version. Twitter users were more engaged and followed the official version than the leaked version. However, the number of replies, likes, and quotes counts were lower but not significantly. At the time of the leak, the story didn't have that much influence on Twitter, and there was also the possibility of rumors. But when the authorities published the incident on social media, users' comments were critical and attitudinal with facts. Additionally, the profanity "fuck" appears more often during the official version. As the word could demonstrate users' emotions, such as anger, it indicates that people might engage more emotionally with the discussion. As a result, user engagement and emotions are much stronger

after an official announcement than before the leak. Social media can influence users' attitudes and engagement toward this issue to some extent.

6. Conclusion

This paper mainly uses Twitter streaming media and R Studio to collect data and use multidimensional analysis methods to analyze qualitative content. By comparing the engagement and emotion of social media users before and after Roe v Wade's official overturn was reported by official media, this paper analyzes whether the intensity of social media coverage can affect the attention and participation of users in a social event. This study concluded that social media could influence people's attention to a social event. The finding is significant as social media plays a critical role in people's everyday life.

Acknowledgements

References

[1] Gottfried, J., & Shearer, E. (2020). News use across social media platforms 2016. Pew Research Center's Journalism Project. Retrieved October 15, 2022, from https://www.pewresearch.org/journalism/2016/05/26/news-use-across-social-media-platforms-2016/

[2] Hong, H. (2013). Government websites and social media's influence on government-public relationships. Public relations review, 39(4),346-356. https://doi.org/10.1016/j.pubrev.2013.07.007

[3] Karakiza, M. (2015). The impact of social media in the Public Sector. Procedia – Social and Behavioral Sciences. Retrieved September 2, 2022, from https://www.sciencedirect.com/science/article/pii/S1877042815012744?via%3Dihub

[4] History.com Editors. (2018). Roe v. Wade. History.com. https://www.history.com/topics/womens-rights/roe-v-wade

[5] Regan, D. H. (1979). Rewriting Roe v. Wade. Michigan Law Review, 77(7), 1569-1646.https://doi.org/10.2307/1288037

[6] Shaw, Donald Lewis. (1977). The emergence of American political issues : the agenda-setting function of the press. St. Paul :West Pub. Co.,

[7] BBC News. (2022). Roe v Wade: What is US Supreme Court ruling on abortion? BBC News. from https://www-bbc-com.translate.goog/news/world-us-canada-54513499?_x_tr_sl=en&_x_tr_tl=zh-CN&_x_tr_hl=zh-CN&_x_tr_pto=sc

[8] Yeoh @GraceYeohCNA, G. (n.d.). 2022. CNA explains: Why is the overturning of the Roe v Wade abortion ruling in the US so significant? CNA. https://www.channelnewsasia.com/world/cna-explains-abortion-roe-v-wade-us-supreme-court-2773416

[9] Canché, Manuel S. González (2015). Social media usage. Pew research,125,52-68. http://www.pewinternet.org/2015/10/08/social-networking-usage-2005-2015/