**CHAPTER 4**

**RESULTS AND DISCUSSION**

In this chapter, we present and discuss the results of our sentiment analysis and topic modeling methods on the data collected from Reddit, these post were made from 2016 to 2023 and about 250 posts were derived with their respective comments from over 100 subreddits which are then processed for further analysis. We also compare and contrast the results of different methods and evaluate their strengths and weaknesses. The final part of this chapter talks and discuss about further analysis on 5G connectivity.

**4.1 Sentiment Analysis Results**

We apply two methods in identifying sentiments in the dataset, we use TextBlob package and SentimentIntensityAnalyser package in python, we check sentiment scores using TextBlob based on positive, negative and neutral sentiments. The results from the dataset are presented in a Figure 1 below.

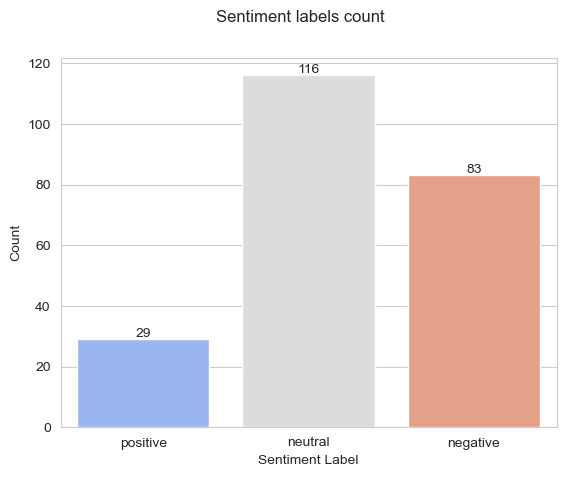
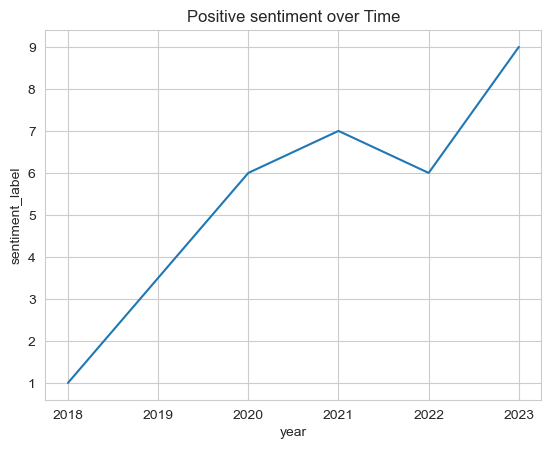


Figure 1: Sentiment analysis results

As we can see, we have a few positive comments or posts about 5G network connectivity in the past years. 

we also have a lot of negative sentiments regarding this, we would be taking a deeper dive looking at how these sentiments evolve over time in Figure 2

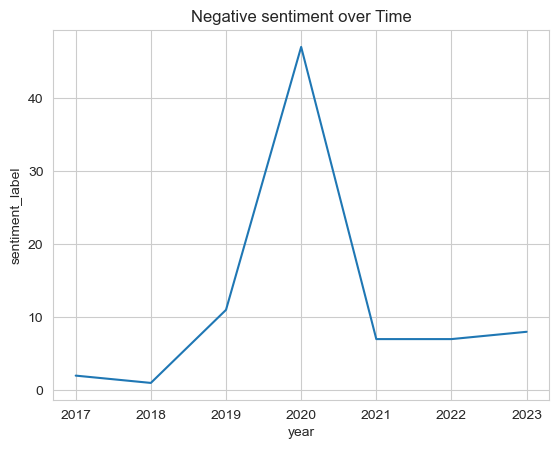


Figure 2

We also visualized the distribution of sentiments across time periods using seaborn line plot.

After the introduction of 5G network in 2018, According to lifewire.com 5G Home Internet service began on October 1, 2018. 5G Ultra Wideband began rolling out on April 3, 2019, and is presently available in parts of 1,700 cities. (Manning, C. D., & Heer, J., 2011)

It didn’t really get much attention until 2019, when it was officially launched and was now rolled out on devices. At the year of launching, it didn’t really get much acceptance in the Reddit space, It had a lot of bad reviews and negative comments in 2019 and it spiked up really high again in 2020 during the pandemic which could be the era when some controversies and protest erupted over the 5G technology, generating fear and anger among the public. As time went by, the negative sentiments dropped and the positive sentiment spiked, as people started accepting it and saw its importance and the need for it.

**4.2 Exploratory Text Analysis (Word Clouds)**

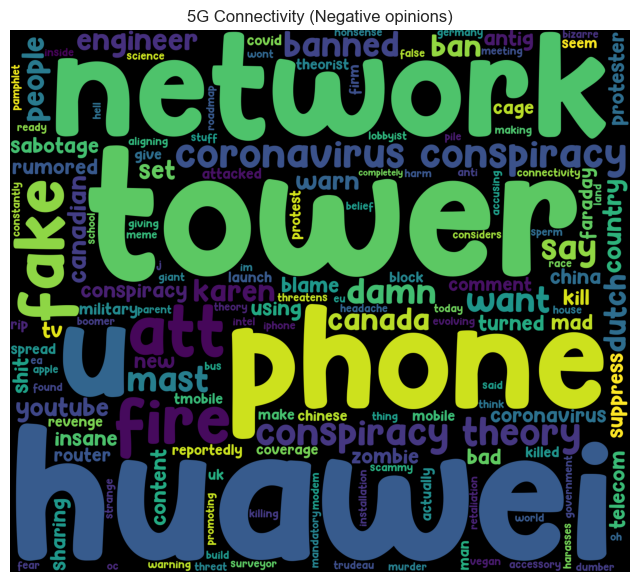
Exploratory text analysis is the process of exploring data, generating insights, testing hypotheses, checking assumptions and revealing underlying hidden patterns in the data. It can help with understanding the problem, preparing the data, and choosing the appropriate methods for further analysis. Exploratory text analysis can involve various techniques, such as word frequency analysis, sentiment analysis, topic modeling, etc.

A word cloud is a graphical representation of word frequency that gives greater prominence to words that appear more frequently in a source text. It can help with exploratory text analysis by identifying words that frequently appear in a set of interviews, documents, or other text. It can also be used for communicating the most salient points or themes in the reporting stage (McCluskey, A., 2010)

This topic will explore word frequencies using bar charts and word clouds for both the negative and positive sentiments or opinions about 5G network in Figure 3.



The positive wordcloud plot here talks more about 5G devices and its expectations, we have frequent words like device, iphone, protect, promise, Huawei, good, power, technology, and lots more.

While on the other hand, the negative word clouds talks about a lot of things, like fake, fire, phone, tower, conspiracies, coronavirus, ban and lots more. To get more insights to these word frequencies, we perform topic modelling on this dataset to gain more insights and also know what the subject topic is around these words or phrases are.Bottom of Form

**4.3 Topic Modeling (Bag of words model)**

LDA topic modelling is a technique to discover the hidden topics in a collection of text documents. It assumes that each document is a mixture of topics, and each topic is a distribution of words. It uses a probabilistic model to infer the topics and their proportions in each document, and the words and their probabilities in each topic. LDA can help with text analysis, summarization, and visualization. LDA works by choosing the number of topics, randomly assigning words to topics, calculating the proportions of topics and words, reassigning words to new topics based on the proportions, and repeating until convergence.

This project uses coherence scores to determine the optimal number of topics for our model.

Coherence score is a metric that measures how well the words in a topic are related to each other. It is based on the idea that words that frequently co-occur in the same documents are more likely to belong to the same topic. Coherence score can help us evaluate the quality and interpretability of the topics generated by LDA topic modeling. A higher coherence score means that the topic is more coherent and meaningful to humans. A lower coherence score means that the topic is more random and noisy.

Coherence score can be used for various purposes, such as:

* Choosing the optimal number of topics for LDA topic modeling. We can compare the coherence scores of different models with different numbers of topics and select the one that has the highest score.
* Comparing different topic modeling methods or algorithms. We can use coherence score as a criterion to judge which method or algorithm produces better topics.
* Exploring and visualizing the topics and their words. We can use coherence score to rank the topics and their words by their importance and relevance. We can also use coherence score to create word clouds or other graphical representations of the topics.

A model with the highest coherence score before fattening out or a major drop is considered as the best model and in our case we go with a model with four topics (k = 4).

The results are shown in Figure 5.

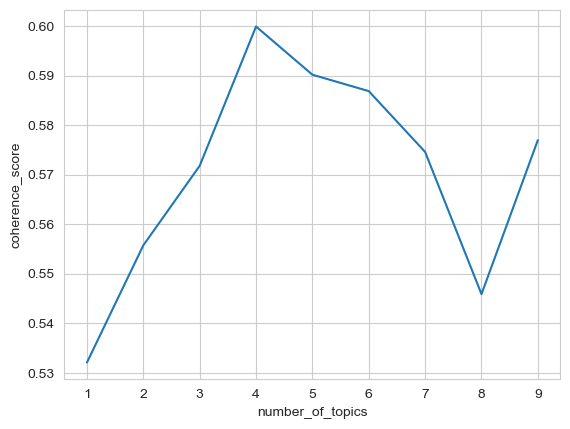


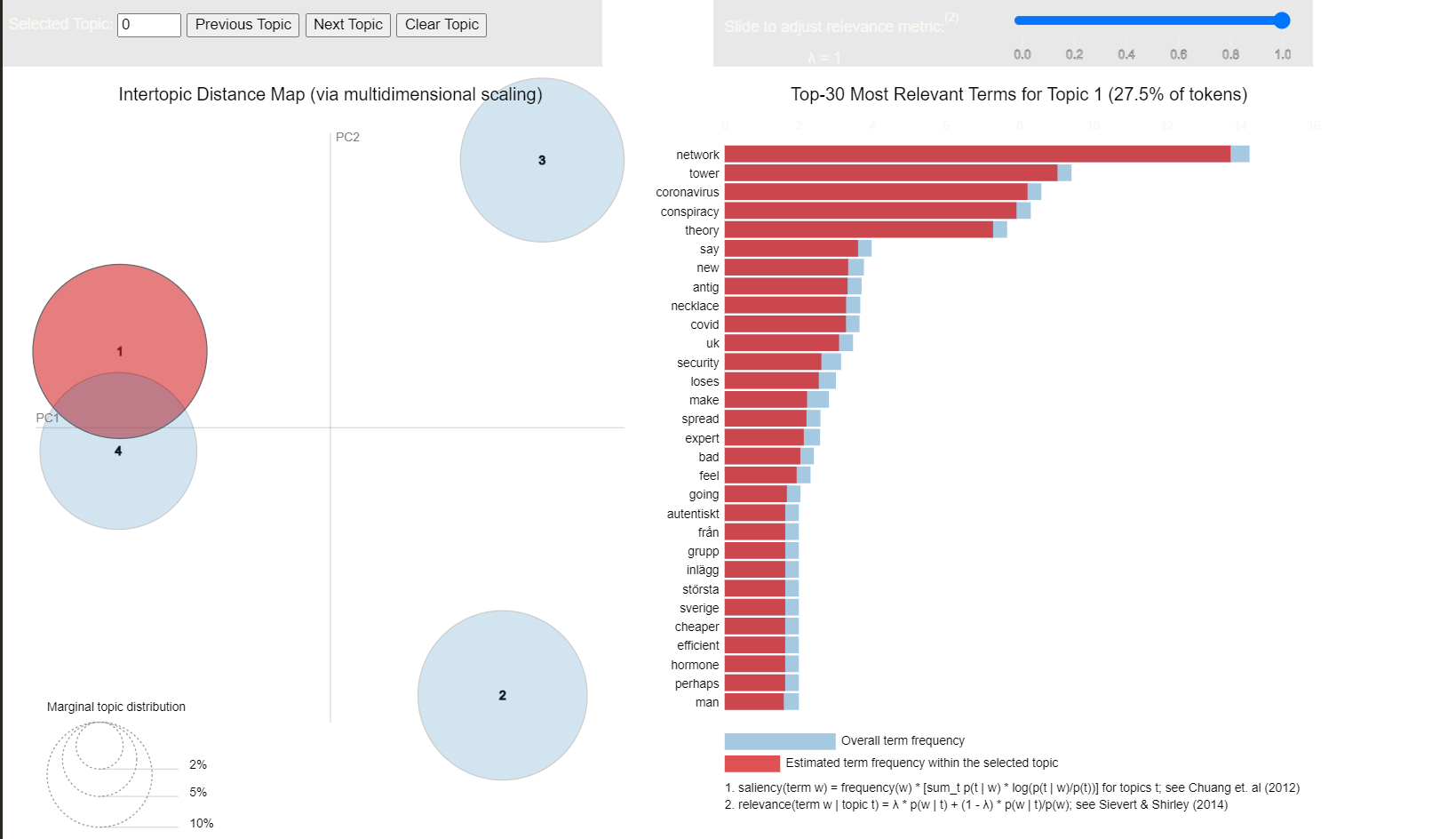
Figure 5: Coherence scores by number of topics

As we can see from Figure 5, the coherence score peaked at 4 topics with 0.60, indicating that this is the best number of topics to represent our data. We extract the top 10 words for each topic using LDA and visualized them using pyLDAVis.

As we can see from Figure 5, each topic has a distinct set of words that reflect a specific aspect or theme related to 5G technology. We labeled each topic based on our interpretation of the words and their relevance to our research question. The labels are shown in Table 1.

These topics can be visualized using pyLDAVis in the genism library for topic modelling, and we can check the intertropic distance between each topic.

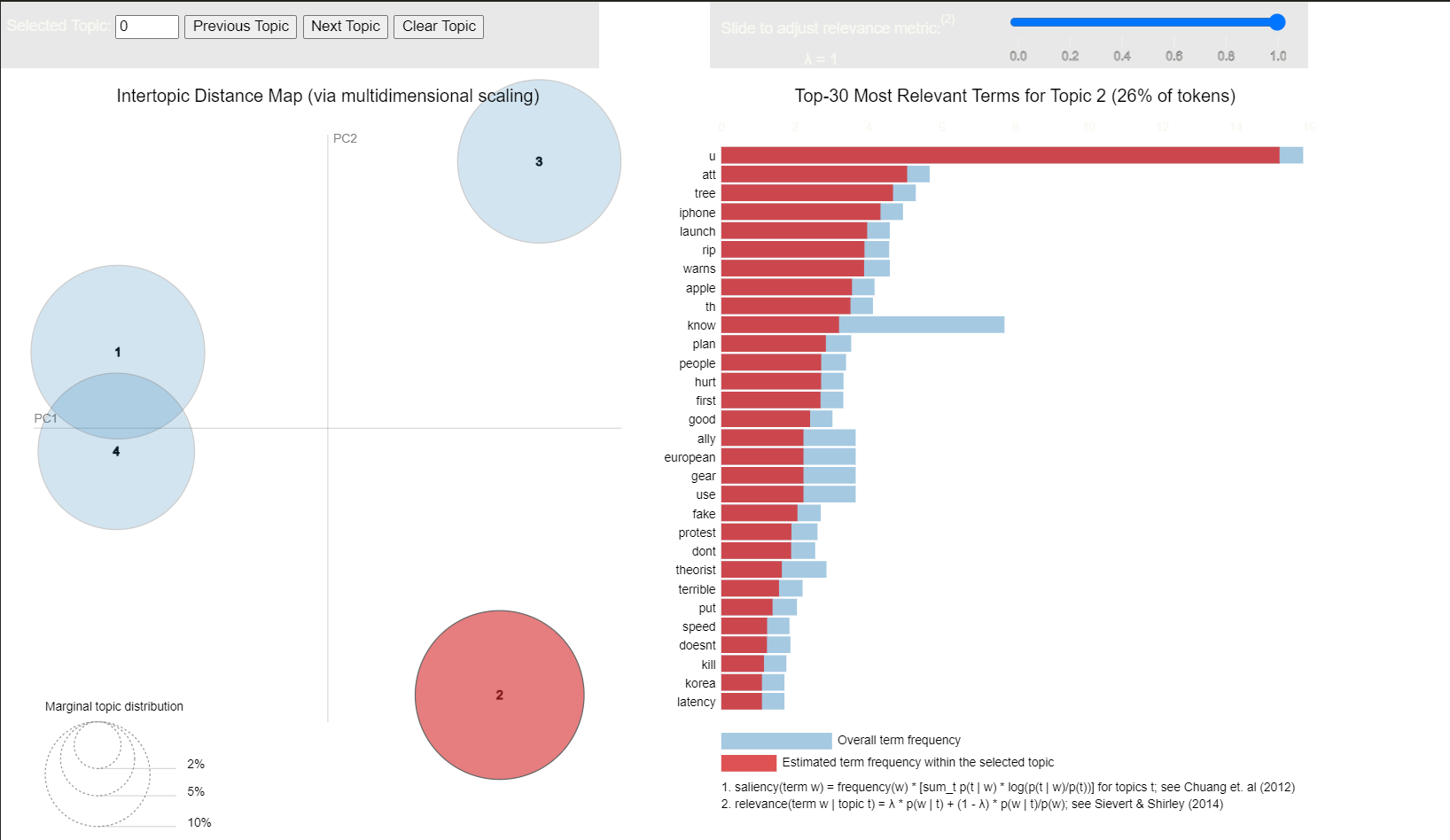
Topic 1



**Key Words: "network", "tower", "coronavirus", "conspiracy", "theory", "covid"**

This topic seems to revolve around conspiracy theories related to 5G technology and its alleged connection to the spread of the coronavirus (COVID-19). It's well-known that there have been unfounded claims associating 5G networks with health issues and the pandemic, which is addressed in this topic

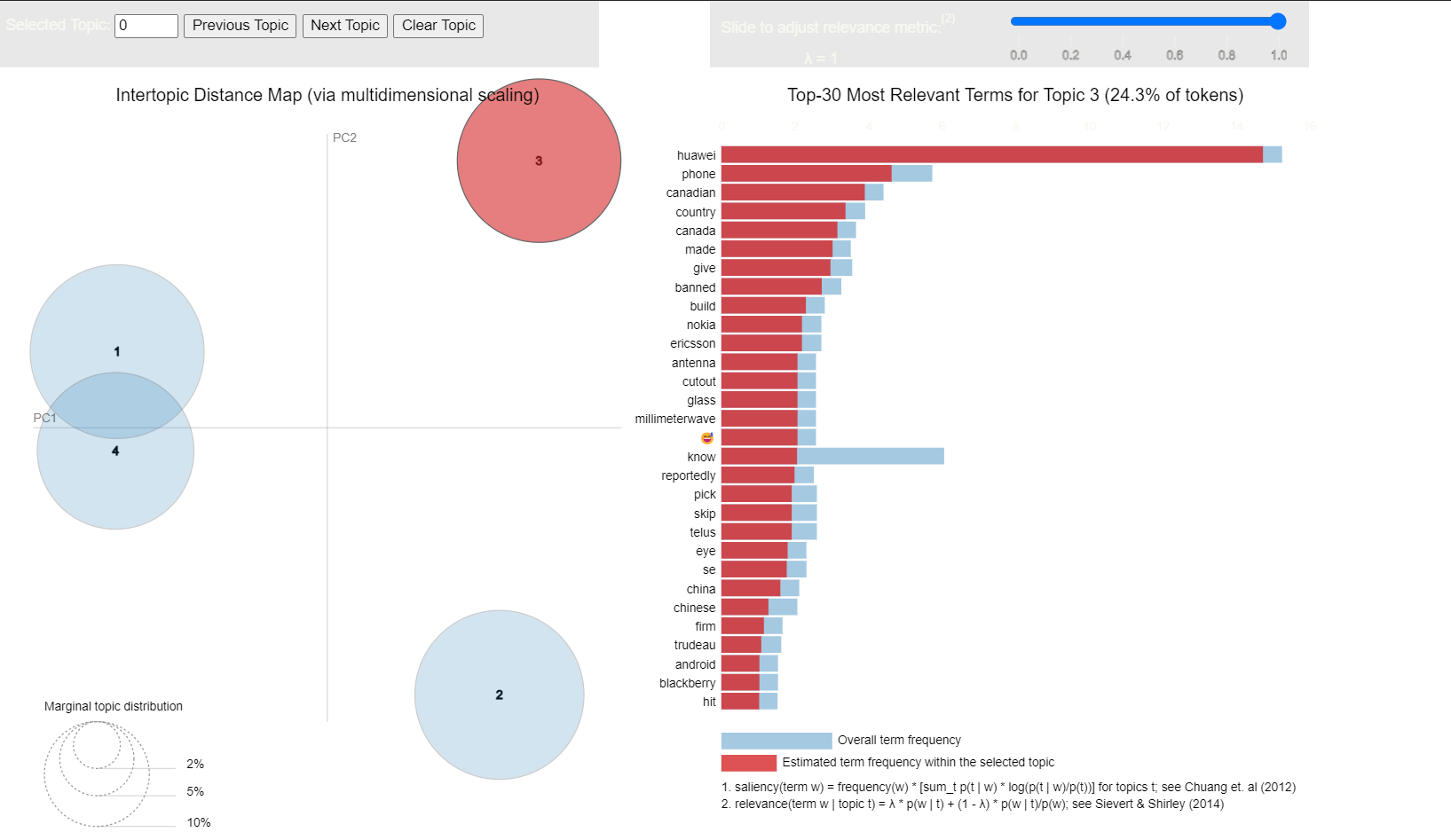
Topic 2



**Key Words: "u," "iphone," "launch," "apple," "th"**

This topic appears to be about the launch of 5G-capable mobile devices, with specific reference to Apple's iPhone. It may discuss the rollout of 5G technology in mobile devices and the impact on the consumer market.

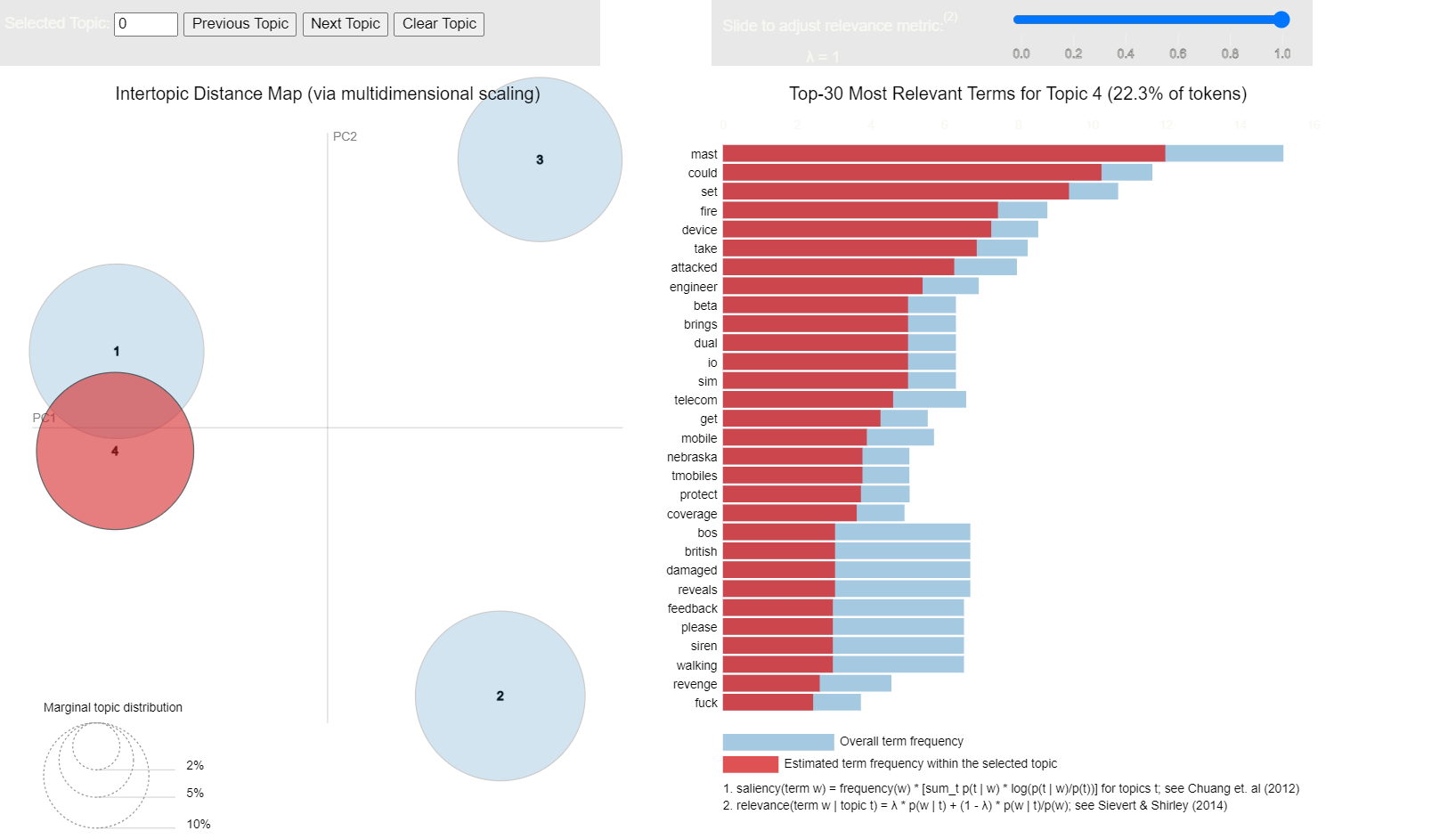
Topic 3



**Key Words: "huawei", "phone", "canadian", "country", "canada", "banned"**

This topic is likely related to the involvement of technology companies in 5G deployment. "Huawei" is a prominent telecommunications equipment manufacturer, and discussions about its role in supplying 5G infrastructure are included. "Banned" suggests the controversy surrounding the use of Huawei's technology in certain countries, such as Canada.

Topic 4



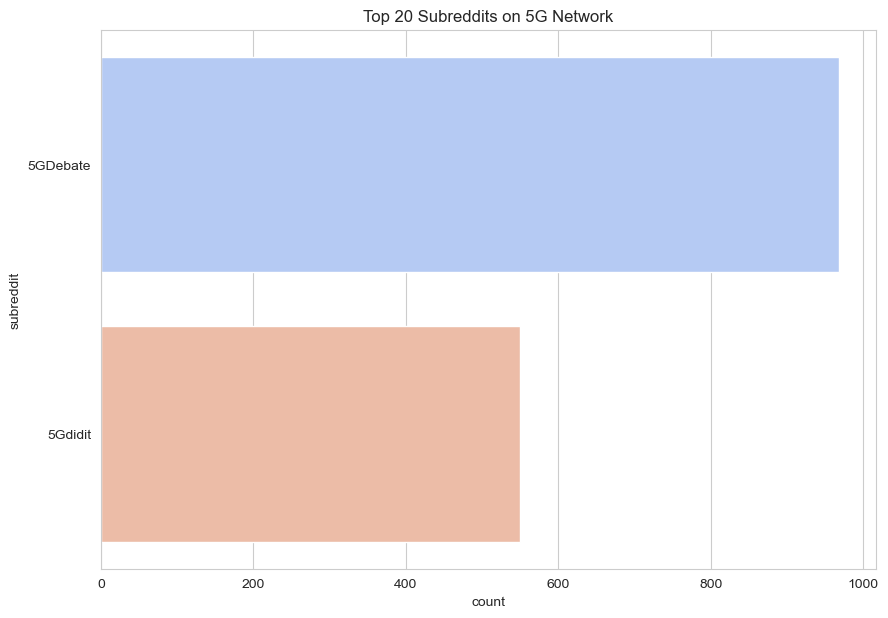
**Key Words: "mast", "could", "set", "fire", "device", "attacked"**

This topic appears to discuss the technological aspects of 5G, including network infrastructure ("mast"), potential issues ("fire"), and the security or vulnerabilities of 5G devices. The word "attacked" may suggest discussions on potential cybersecurity threats to 5G networks.

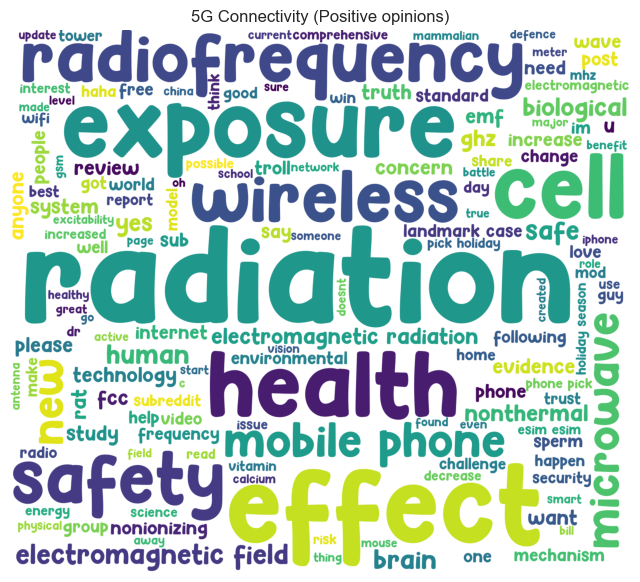
| **Topic** | **Label** |
| --- | --- |
| Topic 1 | Conspiracy Theories and 5G |
| Topic 2 | Technology and 5G Devices |
| Topic 3 | 5G and Tech Companies |
| Topic 4 | 5G and Mobile Devices |

Table 1

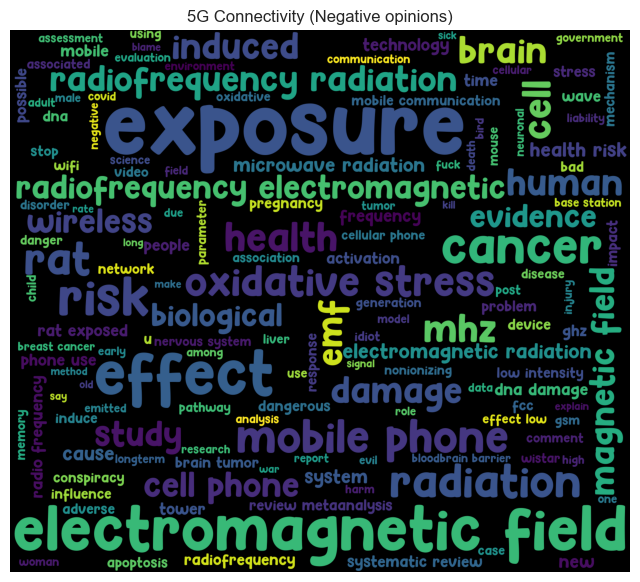
Further analysis was also made, based on two subreddits on reddit (r/5GDebate and r/5GDidit), here a total of about 1500 posts were gotten from these subreddits.



The sentiments generated from these subreddits shows that people are mostly discussing about the negative effects of 5G on the environment and humans at large.

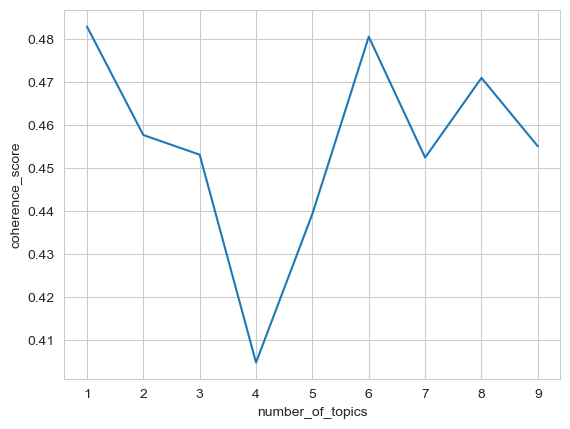


From the positive sentiments above, we can observe the occurrence of certain words like radiation, health, electromagnetic radiation. These are words that are specific to 5G connectivity, and can somewhat appear or be represented in a positive way in these posts. Words like safety, wireless, safe, radiofrequency, health, internet, best can seem like good and positive observations about 5G connectivity and can also explain how 5G connectivity works. On the other hand, we can observe the negative sentiments about 5G connectivity in this subreddits below.

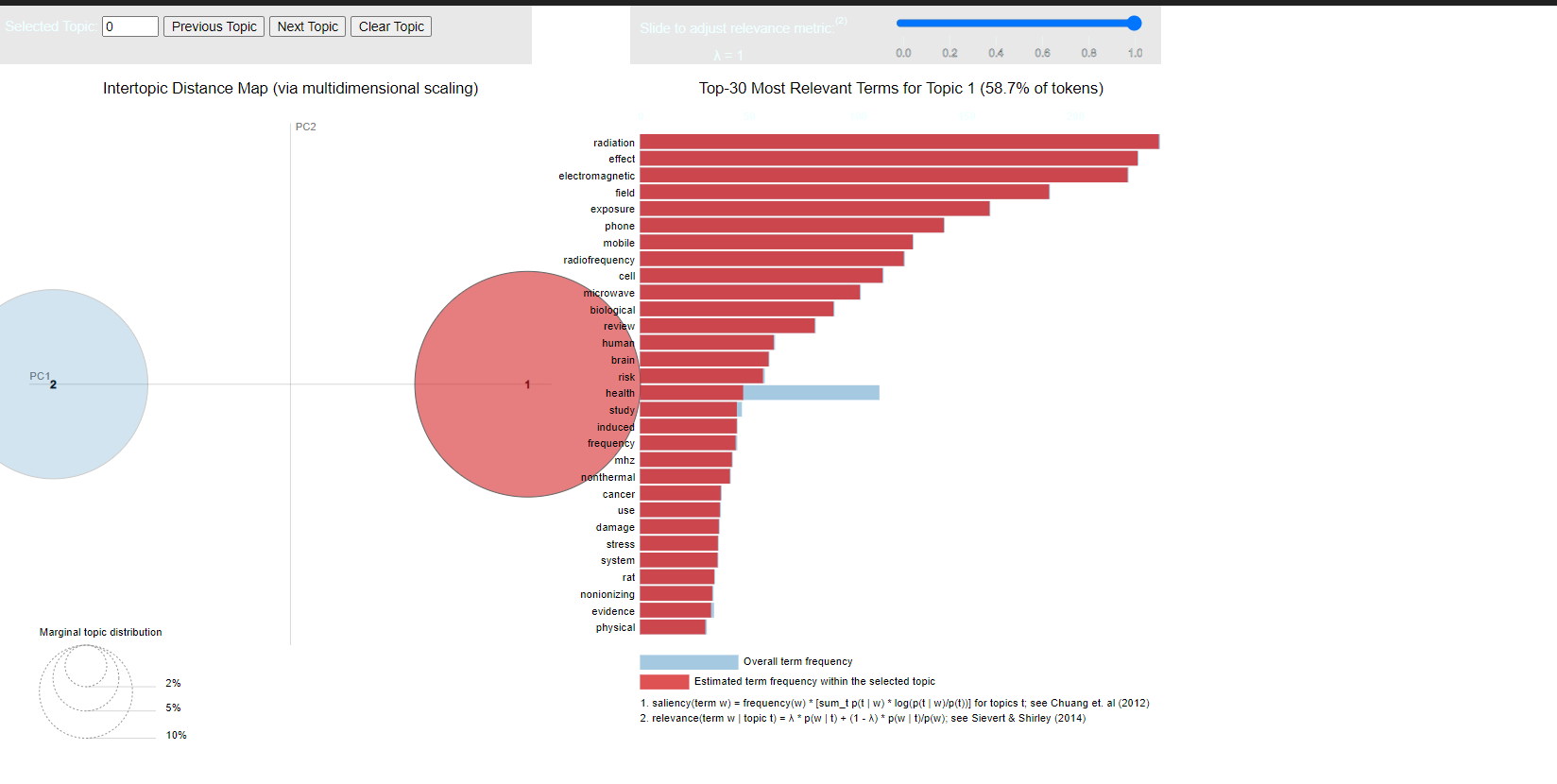


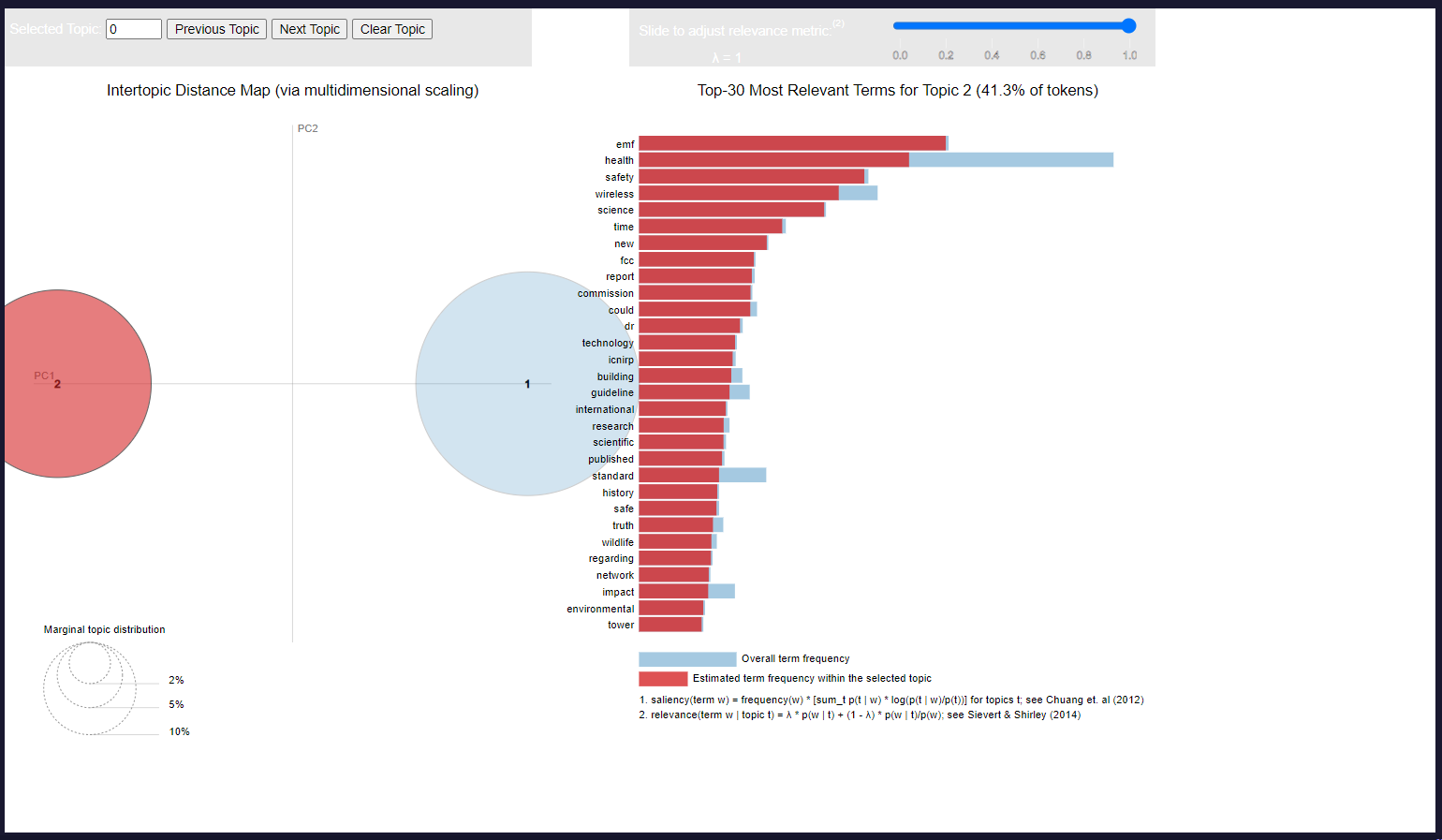
Here, we can observe that there are a lot of negative thoughts about 5G, its effect on human health, the radiation exposure, cancer, risk, rat exposure, damage and lots more which tells how bad 5G connectivity can be to human health which can lead to diseases like brain damage, cancer, and other health issues. Words like rats, birds and other animals seem to pop up recently, as there was once a rumor about 5G wifi killing rats in the house as a result of excess radiation coming out of these 5G devices.

After checking for the best coherence score for our topic modelling, we get our optimal coherence score at 1 topic.



Using LDA as our topic modelling algorithm, we are able to extract this single topic revolving around all the posts in these subreddits.





Based on the prominent keywords in the topic, it appears to cover concerns related to electromagnetic radiation, exposure to electromagnetic fields, and potential health effects. It mentions terms such as "radiation," "effect," "electromagnetic," "exposure," "phone," "mobile," "radiofrequency," "cell," and "health." These keywords suggest that the sentiment analysis may involve discussions or research findings on the health implications of 5G technology and exposure to electromagnetic fields and radiation.

**4.4 Feedback Analysis**

In conclusion, the sentiment analysis project on 5G Connectivity on Reddit has yielded valuable insights into the sentiments expressed by people of different background and thoughts, thus providing recommendations for improvement. Analyzing the sentiment distribution reveals that the majority of feedback is classified as Neutral, followed by Negative and Positive sentiments. This indicates a diverse range of sentiments expressed by students, with a notable presence of positive and neutral sentiments. The prevalence of Neutral sentiments in the the sentiment analysis may suggest that authors are expressing a balanced perspective or a lack of strong sentiment towards their opinion on 5G connectivity. It might also indicate that post authors are providing objective observations or factual statements without expressing a clear positive or negative sentiment. Also the evolvement of these sentiments over time further explains how people’s mindset were changed and how it grew into accepting the 5G technology in recent times.

The quality of output of a topic modelling depends on the purity of the data and the training of the model. In the present study, the data passes through rigorous data cleaning and pre-processing followed by model tuning and evaluations. Topic evaluation results indicated the presence of four topics reflecting public attitude on 5G Connectivity.

The results indicated a high presence of negativity among the public mainly due to distrust in the authenticity of this new technology and its health and environmental impact. Hence, eliminating the distrust among the consumers may increase the awareness of this new technology and it benefits. Thus, from a practical perspective, these findings can help policy-makers and certification bodies enhancing the communications about 5G connectivity and nullify the conspiracy theories behind it, thereby developing trust and positivity. This research has some setbacks. In the first place, the public posts were extracted from Reddit. Second, there is not geographical location of users, thereby limiting our research to not knowing specific locations where these sentiments prevail. Future research, in this direction, is encouraged to perform an extensive analysis specifying posts’ authors and their locations.

**References**

1. Analytics Vidhya. (2021, June 17). Topic Modeling and Latent Dirichlet Allocation (LDA) using Gensim and Scikit-Learn – Part 1. Analytics Vidhya. <https://www.analyticsvidhya.com/blog/2021/06/topic-modeling-and-latent-dirichlet-allocationlda-using-gensim-and-sklearn-part-1/>
2. Alouini, Y. (2018, May 28). Evaluate Topic Model in Python: Latent Dirichlet Allocation (LDA). Towards Data Science. <https://towardsdatascience.com/evaluate-topic-model-in-python-latent-dirichlet-allocation-lda-7d57484bb5d0>
3. Manning, C. D., & Heer, J. (2011).Information visualization: Design and evaluation. MIT Press.
4. Silge, C., & Robinson, D. (2017). Text mining with R: A tidy approach. O'Reilly Media.
5. Halleck, G. D., & Burnard, P. (2012). Visualizing text: Strategies for qualitative research. Sage Publications.
6. McCluskey, A. (2010). Beyond the word cloud: A digital humanities approach to visualizing literary data. Literary and Linguistic Computing, 25(2), 195-205
7. Röder, M., & Schönherr, M. (2010). Exploring the space of topic coherence measures. In Proceedings of the 24th ACM International Conference on Information and Knowledge Management (pp. 393-402). ACM
8. Mimura, T., & Suzuki, H. (2016). Topic coherence measures for evaluating topic models using latent Dirichlet allocation. In International Journal of Document Analysis and Recognition (vol. 19, no. 4, pp. 439-449). Springer.
9. Tran, L. V., & Nguyen, N. S. (2013). A novel approach to topic coherence evaluation for topic modeling. In Knowledge and Information Systems (vol. 35, no. 2, pp. 305-328). Springer.
10. Yang, J., & Zhu, J. (2015). A novel topic coherence measure using word embeddings. In Proceedings of the 28th International Conference on Computational Linguistics (pp. 2698-2702). Association for Computational Linguistics.
11. Lau, J. H., & Baldwin, C. Y. (2011). An empirical evaluation of topic coherence measures in topic modeling. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (pp. 248-257). Association for Computational Linguistics.