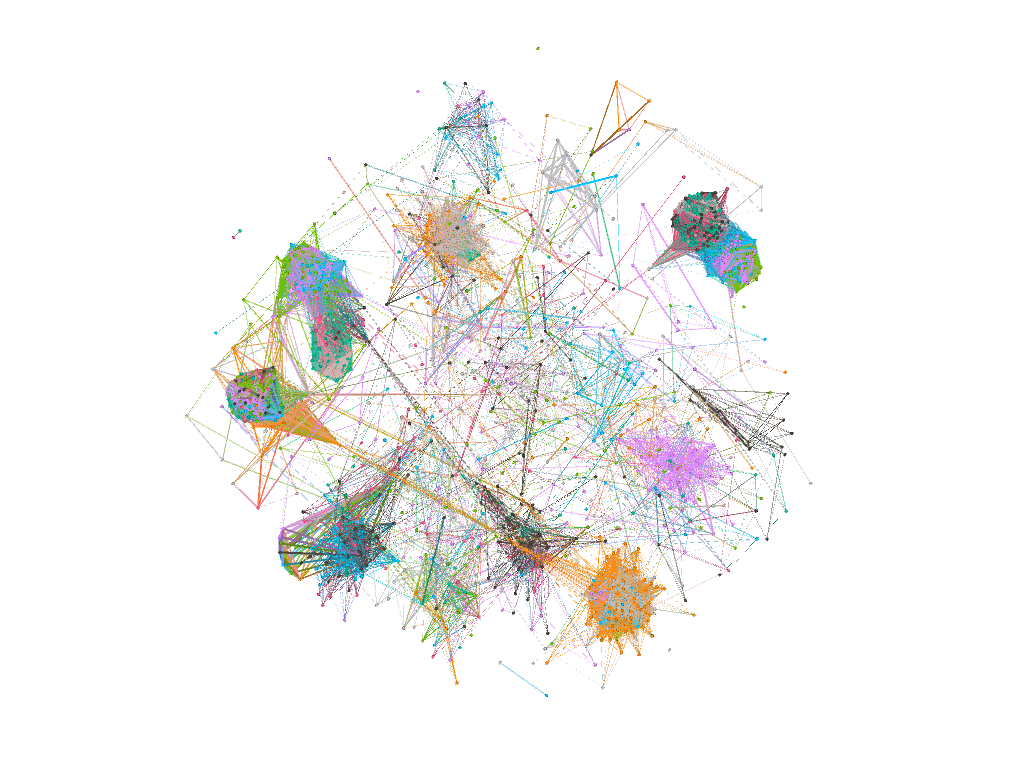
In this analysis, I explore the textual patterns and thematic structures within the Al-Jazeera news corpus related to the Gaza Conflict (2023–2024), using Term Frequency-Inverse Document Frequency (TF-IDF) as a core method for identifying significant keywords and article similarities. To achieve this, I used tools such as Gephi for network visualisation and Plotly for temporal data plotting, drawing insights from a CSV file generated through TF-IDF scoring. This approach enabled the mapping of connections between articles, revealing the frequency and thematic overlap across the dataset.

the tfidf which includes the use of Gephi and Plotly to visualise the data in the csv constructed from the Gaza Conflict (2023-2024) in the Al-Jazeera news corpus. The purpose of tfdif was to see frequency and similarities between articles and find bunches of articles which talked about similar topics based on text frequency.

Exploration:

The main purpose of my exploration was initially to map the articles in Gephi to see the bunches of articles which formed clusters linking to each other with similar topics, this was accomplished by first creating nodes and edges from the csv file provided under tfidf folder titled ‘tfidf-over-0.3’. Ploting the Nodes and Edges results in this visual:



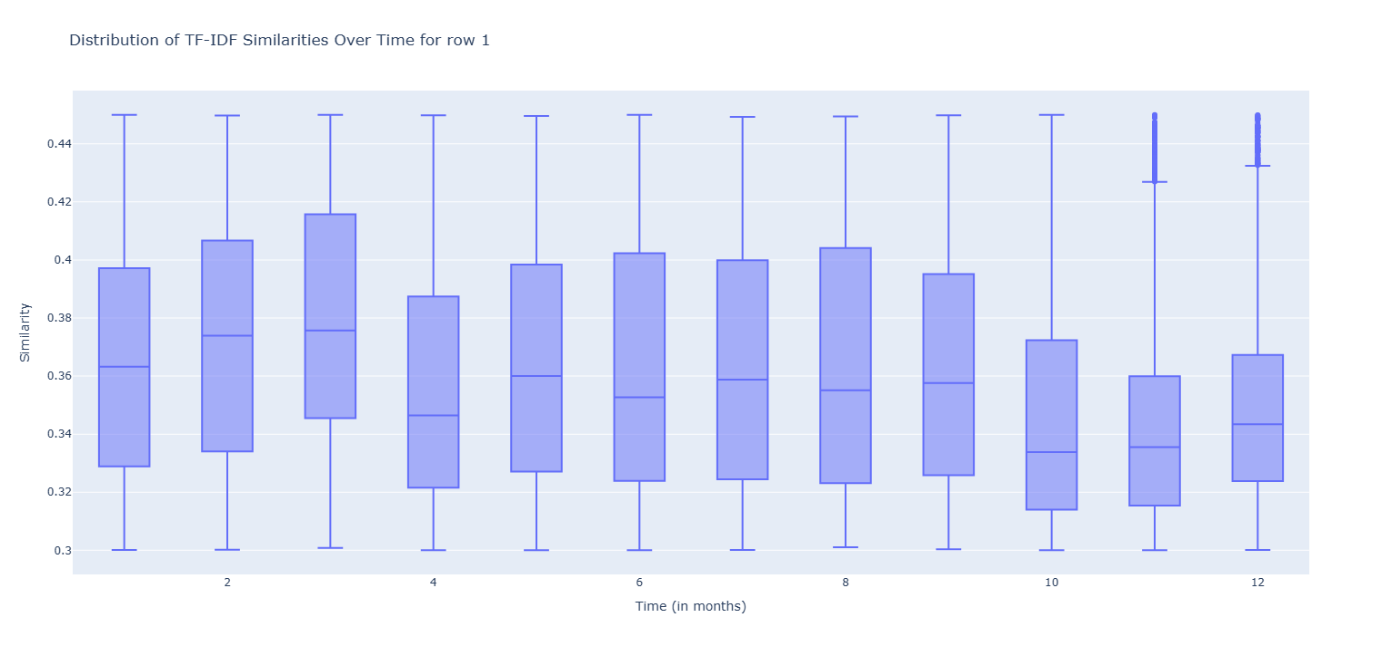
We can see the plots connected by lines, the closer the plots the more similar topics they have and the lines connecting them link them to the next related articles which fall under a different topic within the context of the Gaza conflict.

Visualisation and plots:

From the exploration I thought it best to produce a visualisation of the most number of similarities which lay between the 0.3 and 0.45 mark. It would show the average similarities and with outliers over a period of months from both the datasets.

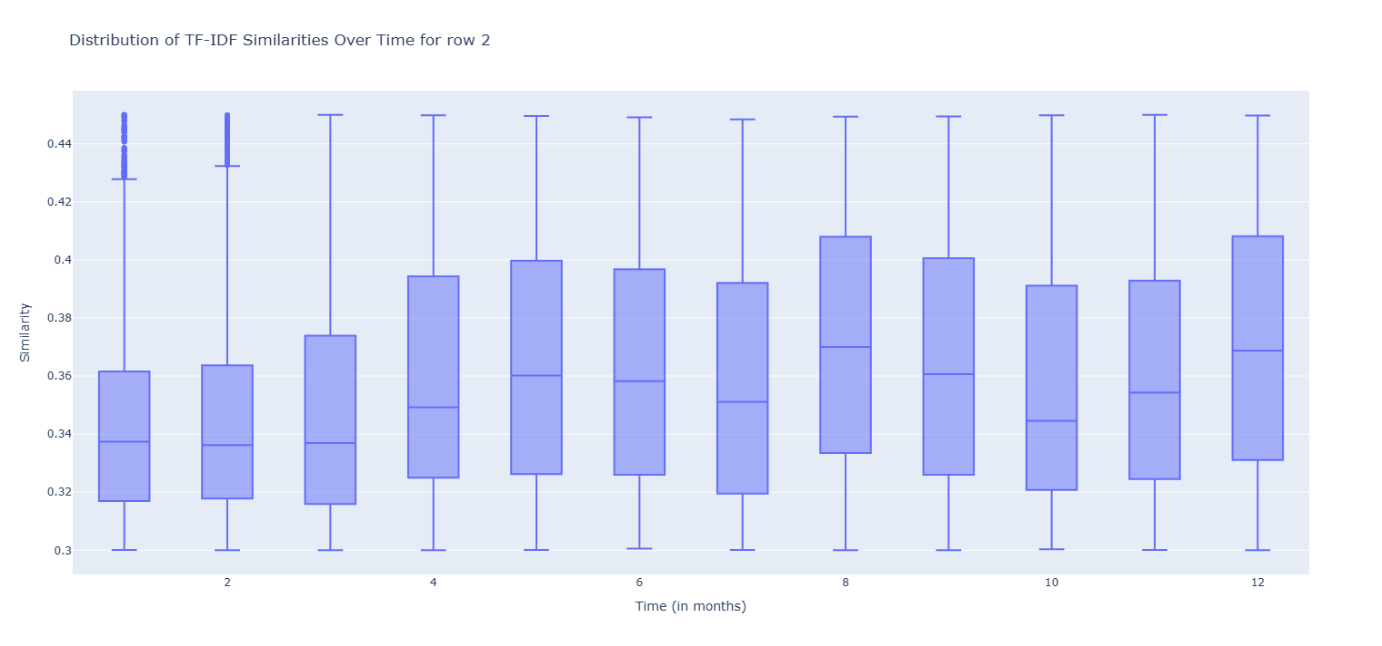
It would be more beneficial to link this with topic modelling to get an greater understanding of the data and to visualize it in an easier manner to see the main topics which linked to each other, mainly UN, resolution, Gaza and Human Rights.

Plot from the first visualization:



The plot was pretty wide spread and ranged at a varying mean from 0.34-0.38. Hovering on the box plots gives more information about the plots. We can see that there are some outliers in months 11 and 12.

Plot from the second visualization:



This plot had outliers in the first and second month with an increasing mean and the spread as we move through the months towards the end. Comparing these two graphs can help is understand the majority of the linking topics being talked about during the Gaza conflict (2023-2024) from the Al-Jazeera corpus.

The TF-IDF analysis, supported by Gephi and Plotly visualisations, provided valuable insights into the thematic clustering of news articles covering the Gaza Conflict. The visualisations highlighted meaningful patterns across time, including outliers and shifts in article similarity, pointing to evolving narratives in the news cycle. Future work could integrate topic modelling to deepen these insights and offer a more refined understanding of discourse trends throughout the conflict period.