Characterise And Understand How Twitter Is Used To Discuss Economic Topics

Harshita Tanwar
Department of Computer Science
University of Exeter
Exeter, UK
ht477@exeter.ac.uk

Diogo Pacheco
Department of Computer Science
University of Exeter
Exeter, UK
d.pacheco@exeter.ac.uk

***Abstract*—In the realm of modern communication, social media platforms have become essential spaces for discussing various topics, including economics. This study dives into the intricate landscape of economic discussions on Twitter, with a specific focus on the sub-network defined by the hashtag #Econtwitter. By understanding how this group interacts, community detection, and shedding light on the popular themes, we aim to gain insights into economic conversations. Using natural language processing, text is refined by removing hashtags, links, and non-ASCII characters. Tokenization generates word embedding via Word2Vec for visual representation of economic discussions. Cluster interpretation via word clouds and Latent Dirichlet Allocation (LDA) reveals dominant themes in each community, adding qualitative insight. Assigning topics unveils core economic discourse themes, enriching understanding. Network analysis using Gephi, based on retweet relationships, illustrates user interactions visually. To meet the main objective, a comprehensive methodology while using a dataset of tweets is used where each tweet's text becomes the main feature, and its community cluster serves as the label. Employing a Support Vector Machine (SVM) classifier, this study yields predictive insights by analyzing tweet text to forecast clusters with an accuracy of approximately 60%. Moreover, the insights derived from this research possess the capacity to enlighten public discourse, steer policy deliberations, and cultivate a more knowledgeable societal dialogue concerning economic affairs.**

***Index Terms*—Gephi, community cluster, Word2Vec, SVM**

I. Introduction

The fusion of social media and intellectual exchanges has ushered in a shift in communication during a time when digital platforms have had a great impact on sharing information and global discussion. In this setting, the discipline of economics, which is intricately wrapped with social structures, has discovered a virtual forum to discuss its complexities, implications, and debates. At the crux of this phenomenon is Twitter, a microblogging platform that has developed into a vibrant forum for debates on a wide range of topics, including economics. The goal of this study is to navigate the complex landscape of economic discourse within the Twitter ecosystem, paying particular attention to the distinctive subnetwork shaped by the hashtag #Econtwitter.

Understanding the underlying aspects of this discourse is growing in significance as economic conversations move beyond academic settings and onto digital platforms. Economic ideas, policies, and discussions have gone beyond the boundaries of academic literature, influencing the digital sphere and shaping public perceptions, policy debates, and even fiscal choices. Echoing the sentiments articulated by David Colander, Richard P. F. Holt, and J. Barkley in their work "The Changing Face of Economics: Conversations with Cutting Edge Economists," the process of analyzing economic conversations assumes significance as economics stands at a crucial moment, transitioning from a steady adherence to the traditional triad of rationality, greed, and equilibrium toward a more multifaceted triad enclosing purposeful behavior, enlightened self-interest, and sustainability [1]. When we consider how economics affects policies, businesses, and individual financial decisions, the importance of this study becomes apparent. When economics meets the strong communication of social media, its impact becomes even stronger. This makes it possible for policymakers, economists, and everyone else to engage in discussions that could influence how economies develop. Understanding the structures behind these online economic discussions enables us to make wiser decisions, develop policies based on facts, and improve economic literacy.

The project is unique and original because of a number of important factors. While other studies have investigated into how social media plays a part in economic discussions, the present one stands out because it is focused specifically on the distinctive #Econtwitter dataset. This unique method involves a thorough analysis of this specialised subnetwork, enabling an extensive analysis catered to a particular community [2]. Notably, this project's distinctive hallmark is the integration of various methodologies, including network analysis, natural language processing, machine learning, and topic modelling. This multifaceted strategy reveals structural patterns, thematic content, and predictive insights, allowing for an in-depth understanding of the dynamics of the subnetwork. The inclusion of a classifier that predicts tweet clusters adds a predictive dimension to the research, which is a noteworthy feature. This predictive dimension demonstrates how machine learning can be used to understand the underlying patterns and content associations present in economic discussions. This strategy represents a cutting-edge methodology for extracting insights from social media platform discourse.

II. Background

In the era of digital communication, the core of social media and economics has become a fascinating field of study. Numerous studies [3] [4] [5] have been conducted on the sentiment analysis or categorization of tweets. Researchers are now delving into the complex dynamics of this changing landscape as a result of the discussions about economic phenomena, policies, and theories that have flourished in this digital space.

According to research published in [6], Twitter is having a greater impact on political discourse as politicians, journalists, and citizens actively use it. In a novel study, researchers from Austria examined Twitter interactions based on topics and professions, as well as connections to the media. Despite the possibility of outside involvement, their findings point to a dominant group of political experts in the Austrian political Twitter network. The study also demonstrates the emergence of specialised voices and sporadic discrepancies between Twitter and discussions in traditional media. Overall, the study clarifies how these dynamics and political participation are related.

The LDA method is successfully implemented and performs optimally in another paper [2] when processing tweet data. This includes topic extraction, topic modelling, generating index words that are in each topic cluster, and topic computer visualisation. With 1260 tweets and a 98% accuracy rate, LDA output demonstrates optimal performance in the process of word indexing in Sport topics.

[7] also examined a comparison of the Random Forest, k-Nearest Neighbour, and Support Vector Machine classifiers for classifying land cover. RFC and SVM's superior performance on both balanced and unbalanced data was discussed. The project's classifiers were motivated by the same thought.

The paper [8] discusses how retweets can reveal important details about the data and the Twitter network, which was also taken into account in this project. It also discusses community evolution in retweet networks.

Gephi offers a wide variety of layout options, making it challenging to select the best one. ForceAtlas2 was used by researchers in [9] to aid in their visualisation. In conclusion, ForceAtlas2 was created to operate as a continuous algorithm within the Gephi user interface. To give users a thorough understanding, the settings and restrictions of the algorithm were described.

There are numerous methods, but [10] discusses the benefit of utilising distance centralities for community detection. One of its main advantages is that no specific number of communities is required.

The literature essentially highlights the dynamic nature of digital economic discourse and the potential of social media sites like Twitter to shape public opinion, disseminate information, and shed light on societal perceptions. Our understanding of the complex interactions between economics, social media, and public discourse is improved by the research described in this section, which serves as a strong foundation for the current investigation into economic discussions within the #Econtwitter subnetwork.

III. Aims and Objectives

The main aim of this paper, as highlighted by the title, was to characterise and understand how Twitter is used to discuss economic topics. The aims and objectives for this paper are as follows:

- To understand how Twitter conversations reflect opinions on economy-related topics.
- To identify the key themes and topics surrounding economic topics that were tweeted about.
- To gauge the tone of Twitter discussions about the same, use sentiment analysis tools like VADER (Valence Aware Dictionary and Sentiment Reasoner).
- Use topic modelling with Latent Dirichlet Allocation (LDA) to classify and investigate Twitter conversations.
- To perform network analysis and community detection on the Twitter dataset
- To train a classifier model which can classify a tweet into a category or cluster

IV. Experiment Design & Methods

The project involves a comprehensive fusion of network analysis between user ids and their corresponding retweet ids [11], natural language processing, and machine learning techniques to explore user interactions and thematic patterns within economic discussions on Twitter. The main objective is to uncover hidden patterns, predict tweet categorization, and shed light on the dynamics of economic discourse. The project employs a multi-step process, including data extraction from Twitter, network analysis using tools like Gephi [12], community detection through K-means clustering [14], sentiment analysis using VADER [15], and thematic extraction using methods like Latent Dirichlet Allocation (LDA) and word clouds [16]. A word2vec model is trained to understand semantic relationships between words. This model's outcomes are then used in classification models like Random Forest and Support Vector Machine (SVM) for categorizing tweets accurately. The refined model is also utilized to predict the alignment of new tweets within existing conversational clusters. Sentiment analysis scores were included as an additional feature along with the tweet text in the model to assess their impact on accuracy. To achieve this, the following steps were performed (Fig. 1):

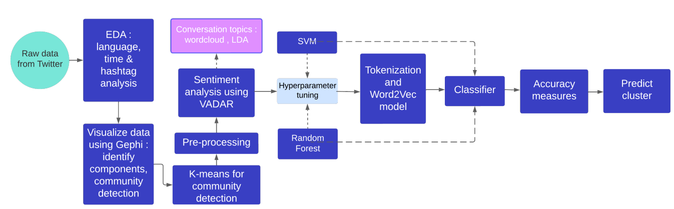

Fig. 1. Flowchart showing project outline

1) *Data collection*: The dataset was sourced using the Twitter API and spans a two-month period, encompassing February and March 2023, right before the

discontinuation of the Twitter API service. Approximately 34K tweets were collected under the hashtag #EconTwitter. These tweets contain various identifiers (Tweet ID, User ID, Retweet ID, Mention ID) with corresponding handles, timestamps, hashtags, and tweet content, constituting a rich array of communicative data.

2) *Exploratory Data Analysis*: The Twitter dataset originated from diverse geographic locations and language backgrounds. To facilitate the analytical scope of the project, exclusively those tweets composed in the English language were taken into account. Language distribution analysis was done to understand the primary language in which users post the tweets. It can help tailor the content, marketing messages, or campaigns to resonate with the target audience for localization strategies. Tweets were also analyzed on an hourly basis (Fig. 2) which helps identify the times when users are most active. This information can guide posting schedule to reach a larger audience and increase engagement. It also helps to provide insights into user behavior and habits. Hashtag analysis (Fig. 3) was done to find the most used hashtags during conversations. Hashtags behave as classifiers, grouping content according to specific topics, themes, or subjects. Taking a closer look at hashtags reveals information about topics that are popular and trending across the dataset. By helping to structure and annotate the content, this classification process makes it easier to spot patterns and new trends. It is possible to identify emerging trends, trending topics, or recurring themes that have attracted user resonance by closely examining the frequency and deployment of hashtags.

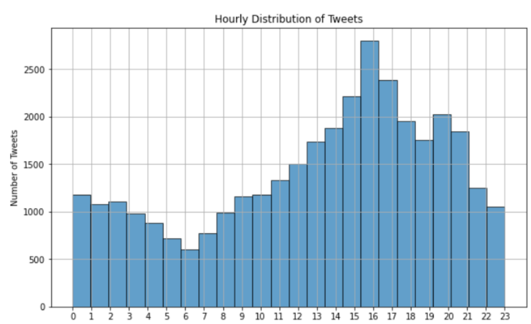

Fig. 2. The hourly distribution graph of tweets during EDA is plotted

3) *Pre-processing and Network Graph*: The Twitter dataset under study comes from a wide range of geographical origins and linguistic preferences. An exclusivity criterion was established, limiting consideration to tweets painstakingly written in the English language, to sharpen and channel the analytical trajectory. Following this curation, the dataset was effortlessly imported into the ”Gephi” software platform, creating an enhanced visualisation space and enabling thorough exploration. This dataset contained a plethora of strongly-connected

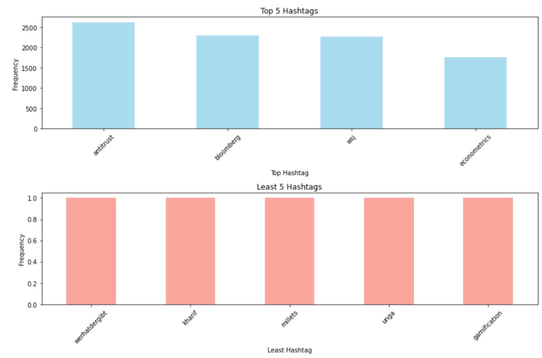

Fig. 3. Top and bottom 5 hashtags each extracted after EDA process with their respective frequencies

components, each of which embodied a distinct variety of interactions as shown in Fig. 4. However, one distinct component with clear significance emerged from this complex combination. Notably, compared to a total of 99 evident components, this obvious entity contributed a significant 93.97% of the entire dataset. Also, this strong influence manifested itself as a single giant component that embodied the dataset’s structural core. When this compelling feature was identified, a smooth transition was carried out to reintegrate the giant component into the immersive setting of the Jupyter platform. A number of essential tweet pre-processing procedures were carefully planned to flow through this transition. These procedural operations included a wide range of actions, such as the systematic removal of Twitter handles, special characters, symbols, URLs, and hyperlinks, as well as the subtle rectification of instances where words with repetitive characters were used (For example: ”looooong” to ”long”). Similar to this, word length optimisation was handled, leading to the recalibration of long words (for example: ”veryyyyyy” to ”very”). The purview included the thoughtful deletion of short words with length less than 3 that lacked deep analytical significance.

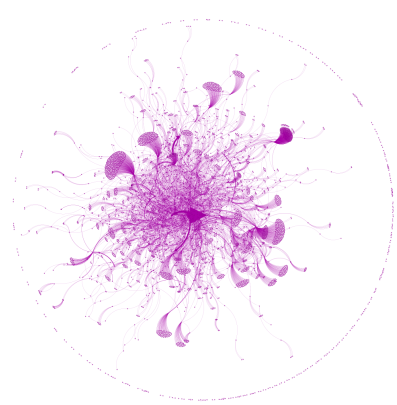

Fig. 4. Network Graph of entire dataset using Gephi software

4) *Community Detection*: When used with Twitter data, community detection is a valuable network analysis method that has the potential to reveal important insights and benefits. With the help of this methodology, it is possible to identify groups or clusters of users within a network who share certain traits, habits, or affiliations. By doing this, it reveals a deeper understanding of the underlying architecture governing user interactions and connections. The resulting insights can be used to modify content and engagement plans to better suit the preferences of each community. As a result of this personalised approach, which delivers content catered to the unique preferences of various user groups, user engagement is increased. We discovered subtopics or specialised discussions within the purview of each identified community that closely align with the project's focus. This discovery is important for identifying emerging patterns, discussions, or topics that demand attention and investigation. This accuracy and fineness can act as a compass for identifying new dynamics, conversations, or problems that demand careful analysis and strategic consideration.

   We used a variety of techniques in our pursuit of achieving this goal, with a focus on utilising the Gephi software's built-in community detection features. The modularity-based and statistical inference-based techniques for community detection were made possible by this software.

   In the modularity-based approach, a heuristic method was employed, which centers on modularity optimization. In terms of computational effectiveness, this method has proven to be superior to other widely used community detection strategies [17]. Additionally, as measured by the metric known as modularity, the quality of the detected communities demonstrates a high degree of effectiveness. The resolution parameter used in this instance was set to the default value, which is 1. As a result, 128 unique communities were identified as a result of using this method. However, a moderate level of community structure is indicated by the resulting modularity value, which is roughly 0.54.

   The statistical inference-based community detection strategy, in contrast, was founded on a non-parametric Bayesian framework that was based on the planted partition model. The advantage of using this alternative method is that it can locate statistically significant assortative modules within networks. This approach lessens the risk of systematic over-fitting, which is evident in both artificial and empirical situations, unlike techniques like modularity maximization [18]. It's interesting to note that during the community detection process, this statistical inference-based technique did not use any hyperparameters. As a result, this methodology allowed for the identification of 206 different communities.

   Given the suboptimal outcomes of the initially employed built-in methods, an alternative strategy was embraced. For community detection, we specifically used K-means, which was built from the ground up using Python code. The use of betweenness and closeness centralities as guiding parameters formed the basis of this strategy. In order to achieve the best results, a complex process of hyperparameter tuning was required. The result of this meticulous process was the identification of the ideal and manageable number of clusters, which was found to be 40 clusters.

   The frequency count of all the clusters helped us to get valuable insights into the distribution of users across different communities or groups within the network. Larger clusters may represent more active or engaged communities, while smaller clusters might indicate niche or specialized groups. For the scope of this project, top 3 influential and biggest communities were studied in details to analyse trends in communications.

5) *Sentiment Analysis*: The process of extracting the feelings, attitudes, and opinions contained within textual data is known as sentiment analysis, and it is a key component of natural language processing. Due to the enormous amount of user-generated content, sentiment analysis has a greater significance in the context of social media. A useful tool in this effort is VADER (Valence Aware Dictionary and sEntiment Reasoner). Notably, VADER shows a remarkable aptitude for managing the particular traits of Twitter data, which are distinguished by concision and informality. The tool's ability to navigate short texts with ease, a distinguishing quality that gels well with the tweet format, highlights its suitability for the project. The decision to use VADER was supported by a number of factors, including its domain specificity, awareness of valence dynamics (including sensitivity to negation, punctuation, and capitalization), prowess in deciphering slang and emoticons, and the notable absence of pre-processing requirements. Additionally, VADER excels at processing short texts, making it a wise choice for the condensed tweets that make up the majority of the Twitter landscape. The sentiment polarity scores were also used as a feature along with twitter text for our machine learning model.

6) *Topic Modelling and Visualization*: Due to their capacity to work with unstructured text data and extract significant patterns from it, LDA and word cloud analysis were particularly well suited for this project when applied to Twitter data. A collection of documents can be analysed using LDA to extract hidden thematic structures and identify the underlying themes that inform conversations. For determining trends, user preferences, and content categories, this was essential. The frequency of words in a dataset is represented graphically by a word cloud. It quickly draws attention to the most frequent terms while highlighting the words and ideas that users found to be meaningful. The unsupervised nature of LDA and word cloud analysis means that they both require little input, making them perfect for exploratory

analysis with no predetermined topics.

7) *Tokenization and Word2Vec*: It is essential to use these methodologies in order to extract semantic relationships between words and identify subtleties contained in user-generated content. Text is divided into discrete units, such as words or subwords, through the fundamental process of tokenization. This fragmentation makes it possible to carefully analyse and comprehend textual content at the micro level. This method captures the distinctive contributions that each element makes to the overall sentiment and context by treating them as individual tokens. This modality accepts hashtags and user mentions as separate tokens, facilitating the investigation of their impact on conversations, trends, and user interactions. Word2Vec is a crucial tool for capturing the semantic relationships between words at the same time. Words are converted into multidimensional vectors in order to achieve this. By creating a contextual understanding, this mechanism gives algorithms the ability to interpret word meanings based on their contextual surroundings. Word2Vec captures the complex contextual relationships between words, transcending simple word occurrence in contrast to traditional methodologies like bag-of-words. This feature makes it possible to measure word similarity and identify related terms, which is a critical skill for grouping topics and spotting trends. In contrast to the sparser representations typical of conventional techniques, the outputs it produces manifest as dense vector representations, thereby mitigating the dimensionality of data.
8) *Classification models*: SVMs, also known as support vector machines, are effective at identifying distinct decision boundaries between classes, which makes them useful for tasks like sentiment analysis and topic categorization. In contrast, Random Forest Classifiers (RFCs) make use of a collection of decision trees, giving them flexibility in handling complex and nonlinear patterns that are common in Twitter data. The inherent noise in Twitter data is a result of things like misspellings, slang, and abbreviations. In the face of such noisy data, SVMs and RFCs demonstrate resilience by focusing on broad trends and patterns as opposed to narrowly focusing on specific instances. Notably, these models perform better than others, especially when applied to unbalanced data sets. In terms of accuracy, RFCs and SVMs clearly outperform competing classification models, particularly in scenarios with skewed class distribution. The window size and vector size were two hyperparameters that were specifically adjusted as part of the word2vec model's refinement. Following the tokenization of tweets and training of the word2vec model, the next step involved the use of an SVM classifier, which was chosen for its demonstrated superior performance. The dataset was divided into training and testing halves at a ratio of 80:20. The meticulous calibration of hyperparameters, which resulted in an accuracy metric around 60%, was a prerequisite for training the ultimate classifier. Calculated measures like accuracy, precision, recall, and F1-score to assess the classifier's performance on the test set.

There was a significant problem while carrying out the aforementioned steps. Using the NetworkX library to visualise large network graphs has drawbacks, such as memory usage, sluggish rendering, and limited customizability. Gephi, a specialised graph visualisation tool, addresses these issues by providing improved layout algorithms, interactive exploration, and optimised rendering. Large network complexity can be handled more effectively, enabling quicker and more specialised visualisation. For instance, Gephi plotted an unfiltered graph much faster than NetworkX, taking only 5 minutes as opposed to NetworkX's 45 minutes.

V. RESULTS

A. EDA and pre-processing

- As shown in Fig. 2 and Fig. 3, the results of EDA showed that most users are active during the evening specially between 2pm and 5pm and least active during late nights and early morning. The popular hashtags used were antitrust, bloomberg, wsj and econometrics.
- About 92.2% of tweets were in English making it obvious the language distribution on the platform.
- The top 3 clusters in terms of frequency were 4, 10 and 18 with frequency ranging between 300-380. We also observed some clusters (at least 10) having frequencies below 25. (Fig. 7)

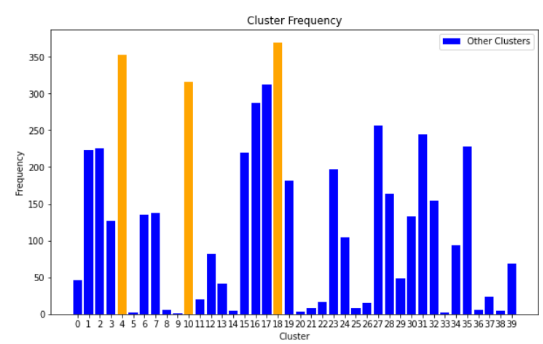

Fig. 5. Cluster frequency count after K-means is implemented with top 3 clusters highlighted in orange and others in blue.

- Preprocessing helps to have a cleaner data. For instance "RT @enkelejdahavari: (Y) The 2nd International Workshop on Migration and Family Economics organized by #IESEGiflame @Malynes2 will take p..." was changed to "2nd international workshop migration family organized iesegiflame take p..."

B. Network Analysis

- The network graph analysis was majorly done using Gephi software. For layout purposes, ForceAtlas2 was used as it was more visually appealing since it is based on the principles of force-directed layout, where nodes in the graph are treated as physical objects that exert attractive and repulsive forces on each other. There were 99 strong components and 1 weak component out of which 97% data was present in only one strong component. This strong component had only 1 giant component which was used for visual community detection.

- The more connected components are darker than the less connected ones as shown in Fig. 6. The most connected components are less as compared to others.

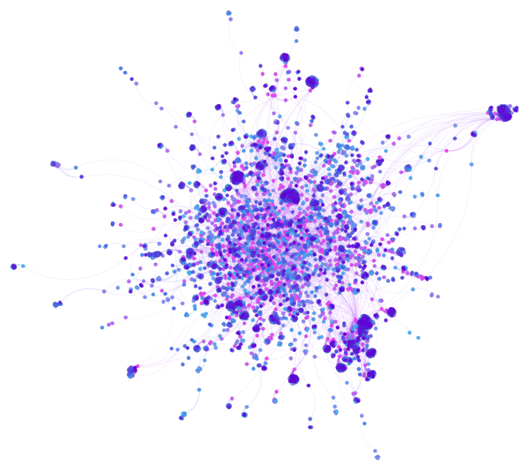

Fig. 6. Network graph showing colour-based connected components on the giant component

- 128 communities were detected in modularity based technique. The hyperparameter ie. resolution was 1 and gave a modularity of 0.58. (Fig. 7)

Fig. 7. Final graph from Gephi showing 128 unique communities detected in modularity based technique

- 206 communities were detected in statistical based technique where statistical inference was 51505.858. (Fig. 8)

Fig. 8. Final graph from Gephi showing 206 unique communities detected in statistical based technique

- The graph created using NetworkX and K-means shows 40 communities and is not as visually appealing as the ones from Gephi. We could clearly observe that nodes that are located nearer ended up in the same community. (Fig. 9)

Fig. 9. Graph visualised from python while using K-means to detect 40 communities

C. WordCloud & LDA

- The cloudword generated on the entire dataset shows that words like "Lina Khan", "Bloomberg", "ftc", "Antitrust" are very dominant. this can be due to the fact that during this time Lina Khan, chair of Federal Trade Commission (FTC) was in news a lot due to her new ideas and different approach. [19] (Fig.10)

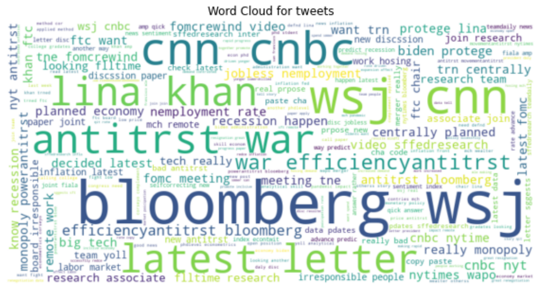

Fig. 10. Wordcloud created from the entire dataset with words as "Lina", "Khan", "bloomberg", "ftc" etc

- LDA was done on top 3 clusters since it was not reasonable to do on all 40 clusters. The results were as follows:-
  - Cluster 4.0 : The few topics included "paper", "answer", "discussion", "course" which might hint towards an academic community.
  - Cluster 10.0 : The few topics included "zerosm_gam", "absolte_advantage", "video" , "remote" which might hint towards a gaming community.
  - Cluster 18.0 : The few topics included "hosing_mogag", "fight", "labor" , "application" which might hint towards a housing or general public community assuming that due to word-limit restriction, hosing_mogag could mean housing mortgage.

  Interactive pvLDAvis notebook was also created to visualise this data.

D. Tokenization & Word2Vec

- After applying tokenization and word2vec, a TSNE visualization of the dataset helps to see how every word is associated with the other words in the data as shown in the Fig. 11 below:

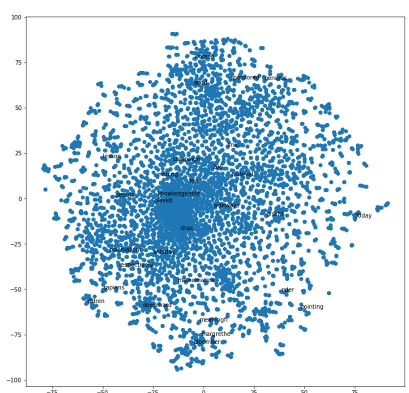

Fig. 11. TSNE visualization of the dataset with 30 random words from the dataset

E. Hyperparameter tuning

- *Tokenization and Word2Vec*: SVM Classifier and Random Forest was used to tune the hyperparameters of the Word2Vec model. SVM gave an accuracy score of 74% with window size=7 and vector size=300 whereas RFC gave 60% with window size=10 and vector size=100. RFC took less time for tuning as compared to SVM. (Fig. 12 and 13)

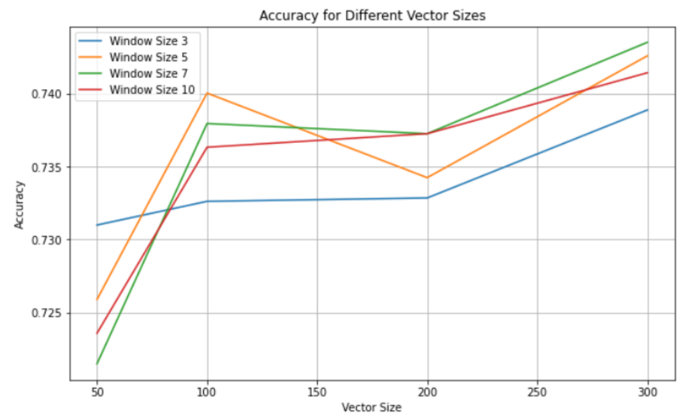

Fig. 12. Word2Vec plot for hypertuning vector size and window size using SVM model

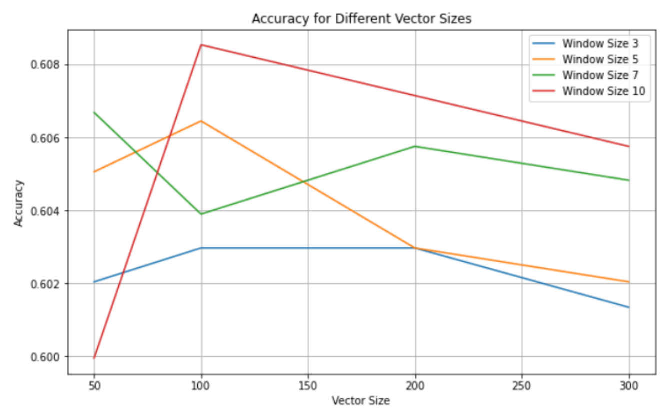

Fig. 13. Word2Vec plot for hypertuning vector size and window size using RFC model

- *k-means and SVM*: To define the best possible combination of number of clusters in k-means algorithm, silhouette scores were calculated. Best parameter values of clusters were also need to be figured out. For this silhouette score was calculated which gave total of 40 clusters with silhouette score of 0.61. For SVM, the best combination of Regularization parameter ie. C and kernel were 100 and 'rbf' by using GridSearchCV giving an accuracy of about 60% (Fig. 14 and 15). Apart from this, when the sentiment score was also added as a feature we observed a slight change in the accuracy from 60.66% to 59.99%. ROC curve for top 5 clusters have been plotted to validate the results. the same can be seen in Fig. 16. The values AUC ranges from 0.94 to 0.80.

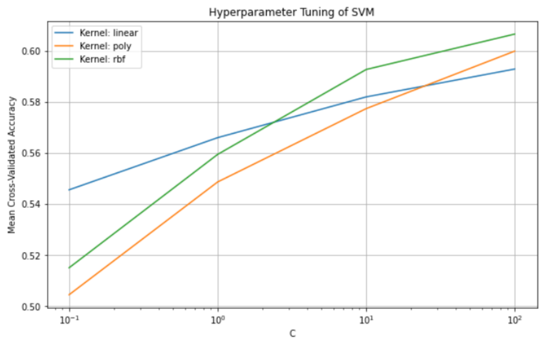

Fig. 14. SVM hyperparameter tuning plot to find the best possible values of vector size and window size

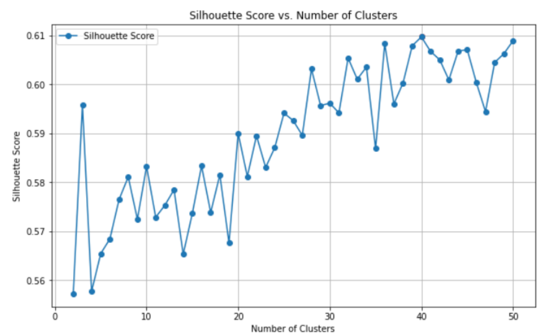

Fig. 15. K-means hyperparameter tuning plot to find the best possible values of vector size and window size

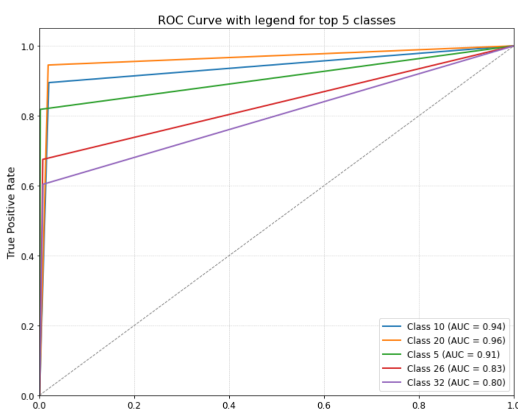

Fig. 16. ROC graph for top 5 classes

VI. Discussion

Using network analysis, natural language processing, and predictive modelling, the study has revealed the complex dynamics of economic discussions within the #Econtwitter subnetwork. User interactions were visualised through network analysis, which identified a sizeable "giant component" that represents the inter-connectedness of economic discourse. K-means clustering successfully categorised users and highlighted various conversation clusters. Word clouds and LDA-powered thematic exploration revealed the most prevalent themes in tweet content. Semantic embeddings from Word2vec deepened our comprehension. Sentiment analysis added to the discourse's richness by exposing its emotional underpinnings. SVM's predictive modelling demonstrated its capacity to classify and forecast tweet alignments within pre-existing clusters with an accuracy of 60%. The study's scope is expanded from observation to anticipation by this predictive component. Comparing with the results from [7], the accuracy for this research has not been as good which can depend on the dataset used as well. There are a few limitations to this project. First one is that the machine learning models are all unsupervised which may lead to lack in ground truth. Second, there is no commonly used acronym or term for "VADAR" in the fields of technology, computer science, or data analysis. Third, in sentiment analysis, we did not take into account any emojis which may have had a good impact and improved the model's performance.

VII. Conclusion

In order to characterise and comprehend the economic discussions occurring on the Twitter platform for the #Econtwitter subnetwork, the project employed a variety of methodologies. Techniques used for EDA and pre-processing helps to clean and understand the data better. The better way to analyse a huge network is by using the Gephi software rather than NetworkX. Visualization techniques like wordcloud and LDA are easy to understand and perform giving niche insights into the data. It was evident that the SVM classifier gave an accuracy of about 60% and that adding the sentimental analysis part (senti scores) did not affect the accuracy much. These techniques aided in the network analysis of the relationships between Twitter users and those who retweeted them. This study demonstrates how well data-driven methodologies work in a world where the exponential growth of data keeps up with the rapid pace of digital transformation to understand the complexities of human discourse. It acts as a lighthouse, offering nuanced viewpoints on how economic debates, sparked by websites like Twitter, shape the dynamic landscape of ideas, viewpoints, and insights as the internet continues to influence public perceptions and societal narratives.

VIII. Declaration

Declaration of originality
I am aware of and understand the University of Exeter's policy on plagiarism and I certify that this assignment is my own work, except where indicated by referencing, and that I have followed the good academic practices.
Declaration of Ethical Concerns
This work does not raise any ethical issues. No human or animal subjects are involved neither has personal data of human subjects been processed. Also, no security or safety critical activities have been carried out.

IX. Refrences

[1] J. Barkley Rosser, R. P. F. Holt, and D. C. Colander, The changing face of economics: Conversations with cutting edge economists. Ann Arbor, MI: University of Michigan Press, 2004.

[2] E. S. Negara, D. Triadi, and R. Andryani, “Topic modelling twitter data with latent Dirichlet allocation method,” in 2019 International Conference on Electrical Engineering and Computer Science (ICECOS), 2019, pp. 386–390.

[3] D. Zimbra, A. Abbasi, D. Zeng, and H. Chen, “The state-of-the-art in Twitter sentiment analysis: A review and benchmark evaluation,” ACM Trans. Manag. Inf. Syst., vol. 9, no. 2, pp. 1–29, 2018.

[4] E. Kouloumpis, T. Wilson, and J. Moore, “Twitter sentiment analysis: The good the bad and the OMG!,” Proceedings of the International AAAI Conference on Web and Social Media, vol. 5, no. 1, pp. 538–541, 2021.

[5] V. R. K. Garimella and I. Weber, “A long-term analysis of polarization on Twitter,” Proceedings of the International AAAI Conference on Web and Social Media, vol. 11, no. 1, pp. 528–531, 2017.

[6] M. Haman, “The use of Twitter by state leaders and its impact on the public during the COVID-19 pandemic,” Heliyon, vol. 6, no. 11, p. e05540, 2020.

[7] P. Thanh Noi and M. Kappas, “Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using sentinel-2 imagery,” Sensors (Basel), vol. 18, no. 2, p. 18, 2017.

[8] B. Evkoski, I. Mozetič, N. Ljubešić, and P. Kralj Novak, “Community evolution in retweet networks,” PLoS One, vol. 16, no. 9, p. e0256175, 2021.

[9] M. Jacomy, T. Venturini, S. Heymann, and M. Bastian, “ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software,” PLoS One, vol. 9, no. 6, p. e98679, 2014.

[10] L. Wu, T. Bai, Z. Wang, L. Wang, Y. Hu, and J. Ji, “A new Community Detection algorithm based on Distance Centrality,” in 2013 10th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), 2013, pp. 898–902.

[11] B. Evkoski, I. Mozetič, N. Ljubešić, and P. Kralj Novak, “Community evolution in retweet networks,” PLoS One, vol. 16, no. 9, p. e0256175, 2021.

[12] “The open graph viz platform,” Gephi.org. [Online]. Available: https://gephi.org/. [Accessed: 16-Aug-2023].

[14] Cai, Zeng, Wang, Li, and Hu, “Community detection method based on node density, degree centrality, and K-means clustering in complex network,” Entropy (Basel), vol. 21, no. 12, p. 1145, 2019.

[15] S. Zahoor and R. Rohilla, “Twitter sentiment analysis using lexical or rule based approach: A case study,” in 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 2020, pp. 537–542.

[16] R. Kusumaningrum, S. Adhy, and S. Suryono, “WCLOUDVIZ: Word cloud visualization of Indonesian news articles classification based on latent Dirichlet allocation,” TELKOMNIKA, vol. 16, no. 4, p. 1752, 2018.

[17] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, “Fast unfolding of communities in large networks,” J. Stat. Mech., vol. 2008, no. 10, p. P10008, 2008.

[18] L. Zhang and T. P. Peixoto, “Statistical inference of assortative community structures,” Phys. Rev. Res., vol. 2, no. 4, 2020.

[19] D. Kerr, “Lina Khan is taking swings at Big Tech as FTC chair, and changing how it does business,” NPR, NPR, 07-Mar-2023.