Esports and gaming community on Twitter: A social network and content analysis approach

By

WONG YIK WAI Eagle

Ip Wang Chun Tiga

This work is submitted to

Faculty of Education of The University of Hong Kong

In partial fulfillment of the requirements for

BASc(SDS)

Date : 2022-05

Supervisor: Dr. Feng shi hui

Abstract

With the growth of technology, games evolved and were no longer confined to home consoles and offline environments. Gamers are linked together through the world wide web, and competing in video games has become a sports category. Esports has proliferated in recent years and is becoming a professional sport, with an enormous amount of competitions organized worldwide every year. Although there have been widely adopted approaches for network analysis and sentiment analysis targeting online communities, the Esports community has not yet been analyzed by researchers. This research aims to understand the attitudes of Twitter users towards Esports, identify the key influencers, and analyze the network structures. Content sentiment analysis and social network analysis are done using the Twitter data with relevant hashtags and keywords. In this research, it is found that the attitude of Twitter users tend to be neutral and positive under the influences of esports players, the crypto industry and the gaming industry.

Acknowledgment

We would like to take this opportunity and thank our supervisor, Dr Feng shi hui, who guided us throughout the completion of this final year project.

Content Page

Abstract	**2**
Acknowledgment	**2**
Content Page	**3**
Introduction	**5**
1.1 Research Objectives	**6**
1.2 Potential contributions	**7**
1.3 Literature review	**7**
1.3.1 Community attitudes	**7**
1.3.2 Network analysis	**8**
1.3.3 Theoretical basis of analysis	**9**
2. Data preparation	**9**
2.1 Data collection	**9**
2.2 Tweets Raw Data for sentiment analysis	**10**
2.2.1 Preprocessing for sentiment analysis.	**11**
2.2.2 Brief analysis of data after preprocessing and cleaning	**13**
2.3 Data for network analysis	**14**
2.3.1 Preprocessing for network analysis	**14**
2.2 Analysis methods	**15**
2.2.1 Content Sentiment analysis (RQ1)	**15**
2.2.2 Social network analysis (RQ2,3)	**15**
2.3 Ethical consideration	**17**
3. Results	**17**
3.1 Sentiment Analysis	**17**
3.1.1 Textblob	**17**
3.2 Sentiment analysis result	**19**
3.4 Network Analysis	**20**

3.4.1 There exists an interconnected network around Esports 25

3.4.2 References to the crypto industry were prevalent 25

3.4.3 There are a number of influencers from different backgrounds who tend towards in-group interaction 25

3.4.4 Distinct communities and categories of users exist in the network 26

4. Discussion and Interpretation 26

4.1 Sentiment Analysis 26

4.2 Social Network Analysis 29

4.2.1 Why are they the key influencers? 29

4.2.1.1 Esports players (streamers) 29

4.2.1.2 The crypto industry 30

4.2.1.3 The gaming industry including managers, online shop platforms and streaming platforms 30

4.2.2 How do the subcommunities form? 31

4.2.2.1 Esports and gaming relevant users 31

4.2.2.2 The crypto industry 32

4.3 Interpretation 32

5. Limitations 33

5.1 Bot detection 33

5.2 Languages 33

5.3 Hardware 33

5.4 Data collection period 34

6. Conclusion & Future research direction 34

Reference 35

Appendix A : Lexicon-based and deep learning approach to sentiment analysis 40

1. Introduction

Since online games became popular in the late 1990s, video games are no longer limited to home consoles, and gaming partners are not limited to those who know each other; internet users who have never met each other will discuss their strategies online, and the online world is like a reciprocal community. The traditional view of electronic games is that they are a form of childlike playfulness, but Esports has overgrown in recent years and is becoming a professional sport, even becoming a competition event at the Asian Games in Hangzhou in 2022. In 2013, it was estimated that 70 million people watched Esports online and there was more online viewership of the League of Legends championship than that of the FA Cup Finals.

Esports generally refers to an organized and competitive approach to playing video games (Witkowski, 2012). It is a type of sports activity where people develop and train mental or physical abilities using information and communication technologies (McCutcheon, Hitchens & Drachen, 2018). Esports has been incorporated into sports and has a complete industry chain as well as changing the media industry for streaming. With the professionalization of Esports, institutions in Asian countries are offering courses in competitive training, game data analysis, et cetera. Also, the Esports players share the same experience with athletes of having their own fans and the top players become superstars such as Faker from T1lol who have 482917 followers on Twitter (Heaven, 2014). However, there is one difference between Esports players and traditional athletes is that they usually stream their gameplays and interact with their fans. This helps to build a positive relationship between players and fans and hence a positive attitude to Esports by streaming Esports games (Sjöblom, et, al., 2017; Xu, Kim & Billings, 2022). Through this study, the attitude of Twitter users towards Esports, influential users on Twitter about the discussion on Esports, and analysis of the Twitter e-sport community can be done.

Asia's sports and Esports industry is snowballing. In 2020, Asia’s Esports generated US $543.8 million in revenue, showing an increase of 4.9% over 2019, and it was expected to keep growing in 2021 and reach US $600 million in revenue (Niko, 2021). The Esports viewership in Asia boomed in 2020 with a 21% increase, growing to 618.4 million Esports spectators in 2020. The Asia market accounts for more than 54% of the global market. It was expected that the Esports market size in China would rise to 215.7 billion yuan in 2020 (Statista,2021). As China is the world's largest Esports consumer market, and its popularity among young people, the Esports industry is likely to provide sustained employment opportunities during its development in the Asian region. It is anticipated that the Esports industry will continue to provide employment opportunities during its development. To this end, further research is essential to enable Asia's Esports industry to maximize its revenue in the future.

Attracted by the rapid growth of the Esports industry, the Crypto industry starts capitalizing on it. Their markets are trying to fuse by each entrepreneur sharing similar characteristics of high competitive and rapid changes (Dom, 2021).The fusion attracts both Esports fans and cryptocurrency users to contact the industry of the other side so that both industries acquire more followers and become larger (Chris, 2019).

1.1 Research Objectives

In this study, the attitude of Twitter users towards Esports, influential users on Twitter about the discussion on Esports, and analysis on the Twitter Esports community are investigated. The objectives of this study include:

1) Analyzing the attitude of Twitter users towards Esports.;
2) Identifying the influential users in the online discussion network of Esports on Twitter;
3) Analyzing the sub-communities in the online discussion network of Esports on Twitter.

In light of the existing research and the research gaps being identified, three research questions are established. The research questions that guide this study include :

RQ1) What are the attitudes of Twitter users towards Esports?;

RQ2) Who are the top influencers supporting Esports and promoting Esports towards the public on Twitter?;

RQ3) What are the structural characteristics of the subcommunities in the online discussion network of Esports on Twitter?

With the above research questions, this research aims at exploring the attitudes toward Esports from the Twitter user's perspective and identifying the key influencers. By achieving the objectives, the significance of this research can provide views and references for understanding the Esports network communities on Twitter.

1.2 Potential contributions

By investigating the above objectives, the Esports industry can find significant ways to grow further. Knowing the attitude of Twitter users as well as their comments allows the industry to understand what they are demanding of the industry, especially for negative comments. Relieving their concerns or arguments can increase the opportunities for the Esports industry. By identifying the key influential users and the network structures, The industry can set clearer promotion strategies for raising images and attracting more people.

1.3 Literature review

1.3.1 Community attitudes

One of the analyses that can be performed on a Twitter community is examining their attitudes. Community attitudes refer to how a community perceives certain subjects; they can be directly tied to that community or outside the community. In the case of Esport, the community analysis will be performed based on the whole Twitter community, and Esport is an inclusion of the Twitter community. Similar research on community attitude analysis has shed light on the Australian community's attitudes towards people with disabilities; results indicate that the negative attitude experienced by specific disabilities groups hinders their ability to engage in ordinary activities (Thompson et al., 2012). This and other research conducted on community attitudes indicates that the implication of these researches are significant as they can display more insight about specific social issues, business projects, or community groups for different purposes in real-life situations(Subramaniam et al., 2020; Thompson et al., 2012; Doogan, Buntine, Linger, & Brunt, 2020).In recent years, more and more research on community attitudes has approached this topic from a data-driven approach, with the implementation of computational analytics tools, even the attitudes or values upheld by a certain community can be analyzed or predicted. Justin & Henry (2015) analyzed attitudes of urban residents with methods like content analysis and multivariate statistics and also utilized the AFINN package for sentiment analysis. Mahmud, Fei, Xu, Pal, and Zhou (2016) utilized data collected from Twitter API and performed classification through Sequential minimal optimization to extract Attitudes from tweets. Kharde and Sonawane (2016) summarized the common approach used in sentiment analysis, which includes Naive Bayes, Max Entropy, and Support Vector Machine, two of these approaches were used in Ardianto, Rivanie, Alkhalifi, Nugraha, and Gata's (2020) research on people's sentiment about Esports education and results indicates that the Naive Bayes algorithm can better predict sentiment towards e-sport education. Nonetheless, it is evident that a lack of research on the attitudes of Twitter users towards Esports, while Twitter is the most popular channel for Esport organizations to update about their status or promotion, that the community is well-established.

1.3.2 Network analysis

A social network is formed with connections of nodes and edges. Social media platforms like Twitter allow users to connect with each other by sharing information such as retweets, comments and like a message, video or image (Himelboim, et al., 2017). In the directed social network, the relationship among users can be found through the direction of linkage. It is also found that there were sub communities in the network formed by more interactive users because of the freedom in the network. Social network analysis allows investigators to find out the community structure and the key influential connectors in the network by calculation of different degrees.

Network analysis aims to uncover the nature of a social network with systematic approaches. A pivotal publication on network analysis by Scott (1988) argues that the social network acts as a compelling model for us to understand social structures. With the help of different graph theories and other models, it enabled us to be able to study certain relationships among members of society (Scott,1988). Serval studies implemented the network analysis approach to understand Twitter communities (Pew Research Centre,2014; Zunera Malik, Sham Haidar,2020; Gruzd, Wellman, & Takhteyev, 2011; Fornacciari, Mordonini & Tomaiuolo, 2015). Especially in Pew Research Centre's study (2014), six types of topic networks were identified and analyzed. One of the examples is the Polarized Crowds, where political conversations take place on Twitter, users among the community were usually two polarized groups that often did not communicate with each other, and there were obviously opposing liberal and conservative factions that discuss the same issue but each spoke about it separately, ignoring opinion from the opposing party. Grandjean (2016) and Gruzd, Wellman & Takhteyev (2011) took a similar approach in that they analyzed the Twitter community from a structural approach. Another approach taken by scholars was to locate the influencers among a social network. Lahuerta-Otero and Cordero-Gutiérrez (2016) for example, conducted research on influencers among the Japanese automobile network with the help of a data-mining research tool invented by the University of Salamanca called PIAR. However, with well-established network analysis methods on Twitter's community, there is a lack of network analysis on the Esports community while it is becoming more and more popular among the world.

1.3.3 Theoretical basis of analysis

Social capital theory provides a theoretical basis for the coming network analysis of our project. According to Bourdieu, one of the primary dimensions captured by the social capital theory is how a social agent can utilize its social network to facilitate its own actions (Bourdieu, 1986). As stated in research question 2 , 3, this paper aims to locate the top influencers and the structural characteristics of subcommunities in Esports. These two aims are derived on the basis of social capital theory as directions to uncover properties in the network. Top influencers in the network refer to those with an abundant amount of social networks, meaning that they are able to achieve certain objectives through invoking their network. In the field of Esports, some characters like this include the Esports commentators and streamers; those people might be able to reach out to their network to promote Esports events or to advertise. The structural analysis of the network is a reference to the other dimensions emphasized by the social capital theory, including the reciprocity, solidarity, and trust among social agents (Bourdieu,1986). With this theoretical basis, relationships among the individuals in subcommunities can be rationalized and unveiled through social network analysis to aid us in understanding the network.

2. Data preparation

2.1 Data collection

Data collection process was initiated by focusing on the common hashtags and wordings, including: **#Leagueoflegends, #csgomajor, #ValorantEsport, #hltv, #data2esport, #pubgesport, #esports, #gaming , #MSI2022 , #esportsarena , #algs , #CSGOEsport**. These tags were used to search tweets for a month. They were selected according to the popular hashtags for Esports found by RiteTag and the most-watched games on Twitch in October 2021, which also have official competitions organized. RiteTag is a platform introducing Twitter’s hashtag grading according to their usage and identifying relevant hashtags that are always used together (RiteTag, 2021). Twitch is an interactive online streaming platform for gaming, entertainment, music, et cetera (Twitch, 2021). Using Twitter's application programming interface (API), tweets collected include at least one of those hashtags. Twitter streaming API was also applied to collect real-time data for a month. The raw data included information such as the author of a tweet, the tweet, the time stamp for each tweet, mentions, and replies. On the basis of this data, users data and available profile information within the overall dataset were also identified and harvested for analysis.

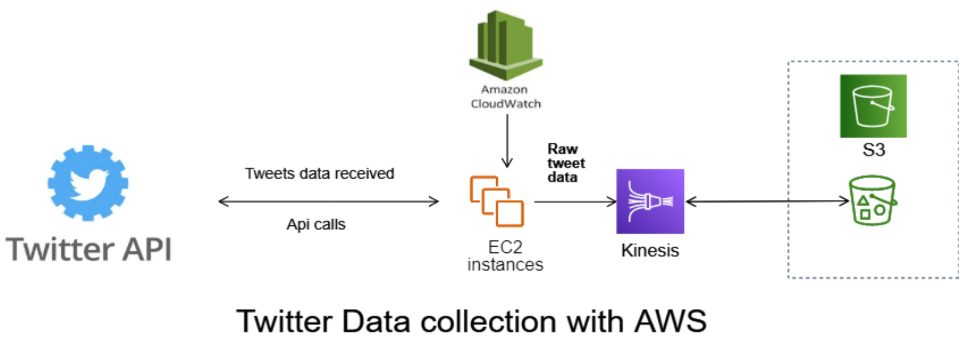


Figure 2.1. The architecture of Twitter data collection with AWS

Figure 1 provides the complete architecture of the data collection process used in the current project. In order to connect to the Twitter API, the python package Tweepy is used and the code is placed inside the Amazon EC2 instance for cloud computing. Upon the data extraction from Twitter API, the raw data is passed through the Amazon kinesis data firehouse which delivers the data to Amazon S3 for data storage. Since Twitter API has a rate limit for data extracted in a certain period, Amazon cloudwatch is used to trigger the data collection program periodically to get new Twitter data from time to time during the data collection period.

2.2 Tweets Raw Data for sentiment analysis

Features	**Description**	**Types**	**Sample**
username	User name of the	Nominal	klever_io

	Twitter account		
description	Description of the account	Nominal	Twitch Affiliate, Lord of Socks
location	Location of the account	Nominal	California, USA
following	Number of following users of the tweet owner	Discrete	5111
followers	Number of followers of the tweet owner	Discrete	276021
totaltweets	Total number of tweets of the tweet owner	Discrete	11219
retweetcount	Retweet count of the tweet	Discrete	1065
text	Content of the tweet	Nominal	You won't want to miss this @IG_Galaxy vs. @kl …
hashtags	Hashtags used for the tweet	Nominal	'Klever', 'CODM', 'ClashRoyale', 'esports'

Table 2.1. Raw Data for sentiment analysis

A total of 59039 of tweets are collected, along with 9 other related features as shown above. For sentiment analysis, all tweets collected (n=59036) are used but only the text column of the data is selected and included in the analysis.

2.2.1 Preprocessing for sentiment analysis.

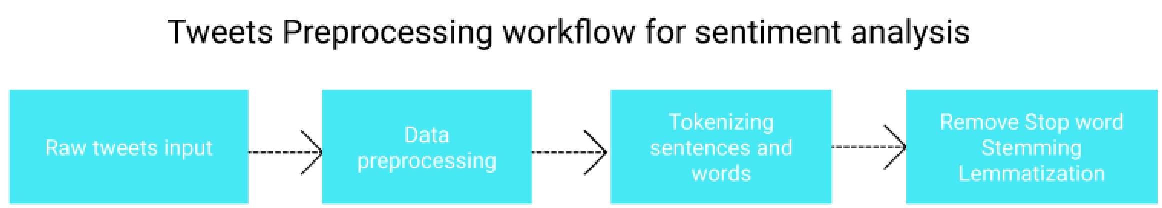


Figure 2.2

Data cleaning is performed following data collection to remove any extraneous information. HTML entities such as < > &amp may be present in data acquired from Twitter. Regular expressions can be used to remove them directly. These entities are converted to regular HTML tags using Python's HTML parser package, such as turning < to “<” and converting & to “&”. As a result, links and special HTML characters are eliminated. During the decoding data transformation process, complex symbols are turned into simple and easier to comprehend characters. Other unrelated features of tweets, such as user mention that begins with a @ sign and emojis, are masked with specific words instead of also using regular expressions. One thing that has to be mentioned is that only English-language tweets are filtered for the sake of uniformity and to reduce the time needed for translation.

Proceeding to the next data preparation step, Tokenization is performed on words and sentences inside the tweets. Tokenization is a common practice in NLP as it helps the machine to understand text sequentially and contextually. Through tokenization, text that is originally a lengthy string is turned into an array with each word as a string.

Tweets before tokenization	Tweets after tokenization
When was I last outside? I am stuck at home for 2 weeks.	['When', 'was', 'I', 'last', 'outside', '?', 'I', 'am', 'stuck', 'at', 'home', 'for', '2', 'weeks', '.']

Table 2.2 tweets before and after tokenization

This step is essential for later as deep learning models are performed on the dataset for sentiment analysis purposes. The array of text obtained through tokenization is transformed into integer indexes to feed them into the neural networks.

At the final steps of the data preparation stage, three standard steps from the NLP pipeline are used, including removing stop words, stemming and lemmatization. All three steps are also achieved through the NLTK library in python, with stop word removal performed first. During this step, the stop words in the array are removed, such as “now”, “at”, “then”, etc. Stemming, the following step, extracts the stem or root form of a word, where the stem does not necessarily express the full semantic meaning. Lemmatization, the last step of data processing, reduces any linguistic word to its general form that preserves its semantic meaning. These two methods of word form normalization are linked and unique, and both can fulfill the goal of effective morphological consolidation.

2.2.2 Brief analysis of data after preprocessing and cleaning

Figure 2.3. Wordcloud of tweets

Here is a brief overview of the data after pre-processing and cleaning. Figure x.x is a frequency plot above that displays the most common words among all the tweets, and the subsequent table is the frequency table. It is observable that half of the lemmatized words are related to the research topic (Esports), while others are trivial wording that is used in tweets for promotional purposes, such as retweets.

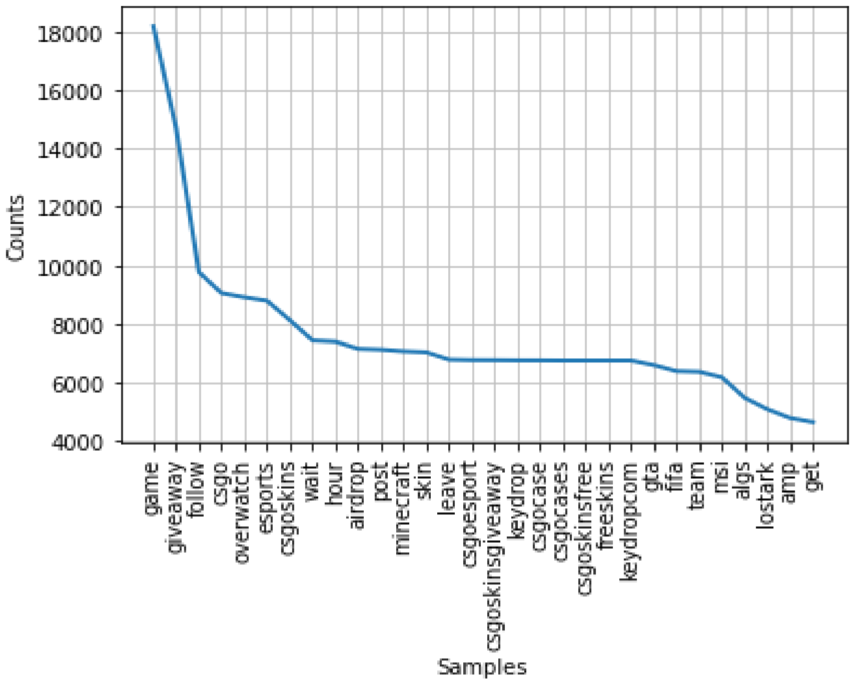


Figure 2.4 Sorted frequency plot for Lemmatized words after removing stopwords

Word	index	token	stem	pos	counts
game	10	games	game	NN	18176
giveaway	9578	giveaways	giveaway	NN	14683

follow	147	follow	follow	VB	9760
csgo	1453	csgo	csgo	NN	9042
overwatch	323	overwatch	overwatch	NN	8901
esports	18	esports	esport	NN	8781
csgoskins	145872	csgoskins	csgoskin	NN	8119
wait	10386	wait	wait	NN	7428
hour	18985	hours	hour	NN	7388
airdrop	136441	airdrop	airdrop	NN	7147

Table 2.3 Most common lemmatized words

2.3 Data for network analysis

For the Twitter streaming 4 datasets, they are the streaming data using the keywords selected. They were collected through Twitter Streaming Importer in Gephi, and the Twitter User's networks were built at the same time. The streaming of data ran once a week for a month, and each set contained about 3000 nodes.

Features	**Description**
ID	User name of the Twitter account
friends_count	Number of following users of the tweet owner
followers_count	Number of followers of the tweet owner
Degree	degree centrality of the tweet owner
Modularity	modularity class of the tweet owner
Eigenvector Centrality	eigenvector centrality of the tweet owner
Closeness Centrality	closeness centrality of the tweet owner
Between Centrality	between centrality of the tweet owner

Table 2.4. Raw data for network analysis

2.3.1 Preprocessing for network analysis

No additional preprocessing is needed for the data in network analysis.

2.2 Analysis methods

Different datasets are applied for different analysis because the data for sentiment analysis do not show the connections between the Twitter users. The data only retrieved the content of Tweets and the brief detail of the Twitter accounts such as name, no. of followers, etc. Therefore, other datasets showing the retweet, comments, and like connections are required to be collected for social network analysis.

2.2.1 Content Sentiment analysis (RQ1)

Sentiment analysis is used for addressing our RQ1 to find out the users' attitudes towards Esports on Twitter. It is a text analysis method used to understand the opinions, emotions, and subjectivity of text. Using Twitter API to collect the comments from the users with common hashtags and divide them into 'positive', 'negative,' and 'neutral' with python. Sentiment analysis can measure the attitude of a group towards esports or assess the 'positiveness' as a personality trait of a single user, especially the key influencers. TextBlob in the python library is implemented to measure the average polarity of the tweets and hence divide the data into the mentioned three groups. For polarity equals above 0, it is positive. For polarity equals 0, it is neutral. For polarity equals below 0, it is negative. Also, word clouds are generated to visualize the comments of Twitter users according to their attitudes. Therefore, what people like or dislike about Esports are identified.

2.2.2 Social network analysis (RQ2,3)

Network analysis is used to find out the top influencers of the Esports network on Twitter as well as the structure of the networks. From the data collected through Twitter API, it is hoped that the PageRank algorithm can be applied in order to locate the significant nodes in the network that we construct on the Esports community. The PageRank algorithm suggested that if there are important links to a page, the links to other pages will also become essential (Xing & Ghorbani, 2004). The network analysis is done using Gephi. Gephi can create and visualize the network by inputting the dataset of nodes and edges as well as calculating the statistics used in analyzing the structure of the network. Also, it can collect Twitter streaming data with the plugin 'Twitter Streaming Importer" and visualize the directed network immediately by entering the selected keywords and choosing the type network (e.g. hashtag network, users network) to be made. Using Gephi, the Twitter influencers and the communities are detected.

For research question 2, the centrality and degrees of the dataset are calculated to find out the key nodes, which mean the influential users. In-degree (incoming tweets) and out-degree (outgoing tweets) are used to identify two-way users with the values ≥1 and one-sided users with the values = 0 in the directed graph. While it was found that the number of followers is not related to the number of retweets and mentions that a user receives (Cha et al., 2010). Centrality provides insights in a network about the roles of different nodes and hence their importance. Degree centrality is the number of direct connections an individual has in the network. Closeness centrality is about the shortest distance between the nodes. Betweenness centrality quantifies the importance of nodes in connecting other nodes. Eigenvector centrality assigns relative scores to the nodes and measures the importance of a node. Given the social structural position of these individuals, they are regarded as having disproportionate influence over content, information and ideas that flow in the Esports network in Twitter. For the influential nodes, they are identified with the eigenvector centrality higher than 0.7 and /or the degree centrality ranked in the top 3 in the dataset.

For research question 3, modularity and the number of subcommunities are first calculated. Then, the structural characteristics are examined by analyzing the degree distribution and several network-level measures. Directional networks are also built and visualized as forward and backward edges respectively by extracting retweets and mentions from each tweet's text. Also, the communities are colored according to their modularity classes. Modularity is a community detection algorithm that allows fast unfolding of communities in the networks (Thu & Wai, n.d.). Only communities including more than 3 % of total users are seen as significant.

The network is constructed using Gephi with the real-time streaming data in Twitter using the selected keywords. In the networks, the nodes are the Twitter users and they are connected with the retweet, comments and like as edges. After entering the keywords, the nodes pop up and connection is then built eventually to form the network.

2.3 Ethical consideration

Ethical approval is acquired through the Faculty of Education, and Faculty Research Ethics Committee as online Twitter data is collected for the data analysis.

3. Results

3.1 Sentiment Analysis

3.1.1 Textblob

As mentioned previously that this dataset is unlabelled, and sentiment categorization can only be achieved through unsupervised learning. To this purpose, various sentiment analysis tools have been reported, while the most extensively utilized among them are TextBlob and Vader [ ]. For this research, TextBlob is used to calculate the polarity and subjectivity of the tweets. Polarity refers to feelings and sentiments, whereas subjectivity refers to ideas, thoughts, and facts. An application of the TextBlob python package is required for this.Both subjectivity and polarity functions from TextBlob returns a score within the float range of [0.0-1.0] and [-1.0-1.0].That the higher the subjectivity score, the more subjective the text is. For polarity, the score will be classified into three sentiments (negative, neutral, and positive) based on the polarity score. When the score is above 0, it is classified as positive. When it is equal to 0, it is neutral; else negative.

The methodology of how TextBlob obtains both subjectivity and polarity scores is based on a lexicon approach. Each word is labeled with a polarity score and subjectivity score inside the lexicon based on their meaning. Take the word “Amazing” (Loria,2014) as an example; it has a polarity and subjectivity score of 0.8, 1.0 or a polarity and subjectivity score of 0.4, 0.8 based on the word sense. Further computation might also be performed by Textblob on the polarity score for the instance of negation or appearance of modifier before the word.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{1}$$

TextBlob obtains polarity and subjectivity score from the n words and phrases in the sentence, and the final polarity and subjectivity score is calculated with an averaging technique that divides the summation of score $\mathscr{X}i$ for i the number of words and phrases by the n words and phrases.

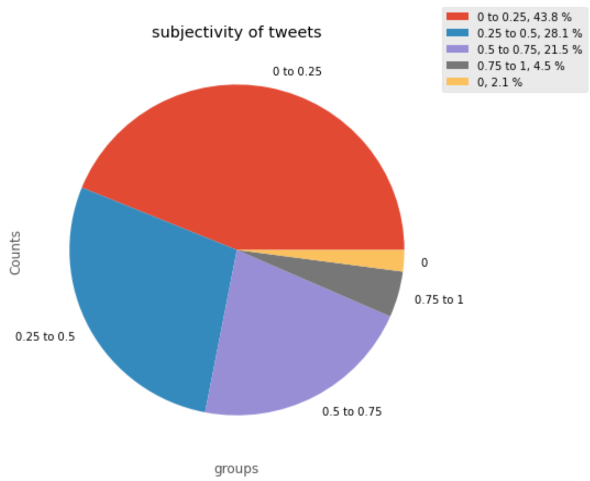


Fig. 3.1. Grouped frequency of subjectivity score

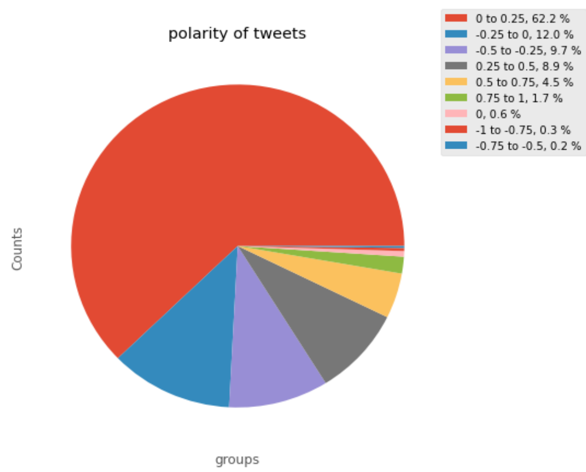


Fig. 3.2. Grouped Frequency of polarity score

Fig 3.1 and Fig 3.2 above demonstrate the grouped frequency of both subjectivity and polarity score of the datasets. It is visible that over 60% of the tweets have a neutral or semi-positive sentiment, and almost all of the tweets contain some level of subjectivity.

TextBlob is also used as the baseline model for the following steps where another lexicon sentiment analysis approach VADER, as well as two deep learning sentiment approaches, the Bidirectional Encoder Representations from Transformers (BERT) and Convolutional Neural Networks (CNNs), will be used to perform sentiment analysis on the same dataset. Machine learning metrics will be applied to compare the performance of each model later(See Appendix A). The reason why TextBlob is used as the baseline is that the manpower requirement for running the analysis is very low. Thus, sentiments of tweets are located easily with a reliable lexicon-based approach.

3.2 Sentiment analysis result

Upon the analysis of the tweets in the dataset *(n = 59039)*, three types of sentiment were labeled through Textblob. As shown in Figure 3.3 and 3.4, about 40% (~25000 tweets) of the tweets is labeled as neutral sentiment, followed by about 35% (~21000 tweets) being labeled as positive sentiment. The remaining 25% (~13000 tweets) of the tweets are labeled as negative sentiment.

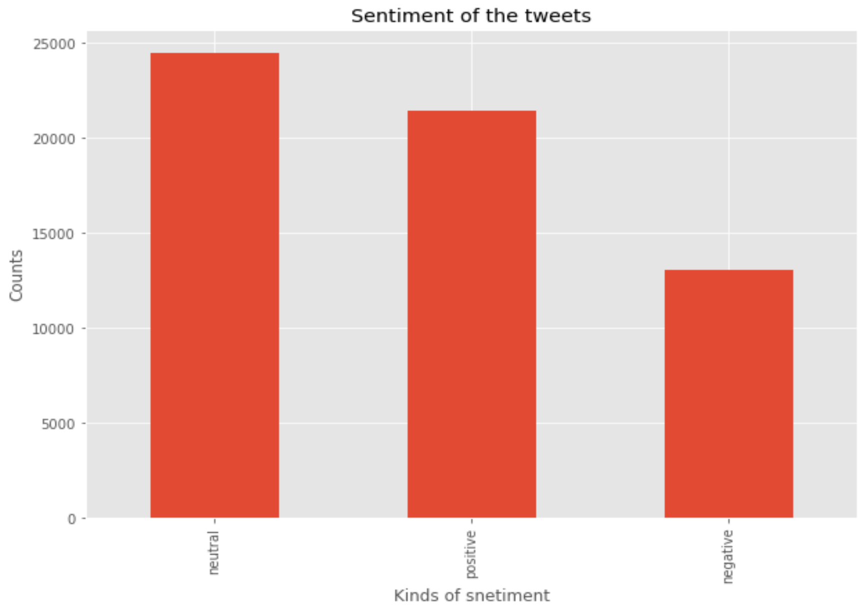


Figure 3.3 Counts of different sentiments among the tweets

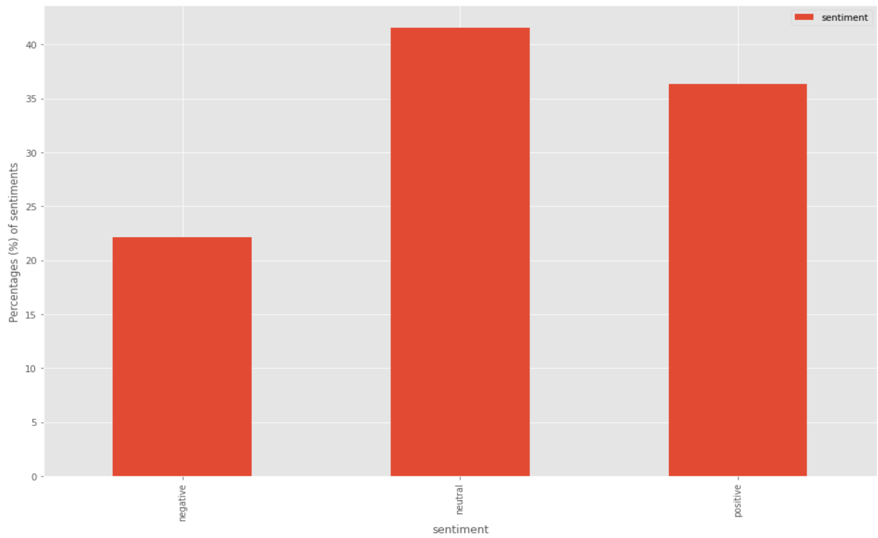


Fig 3.4 Percentage of different sentiments among the tweets

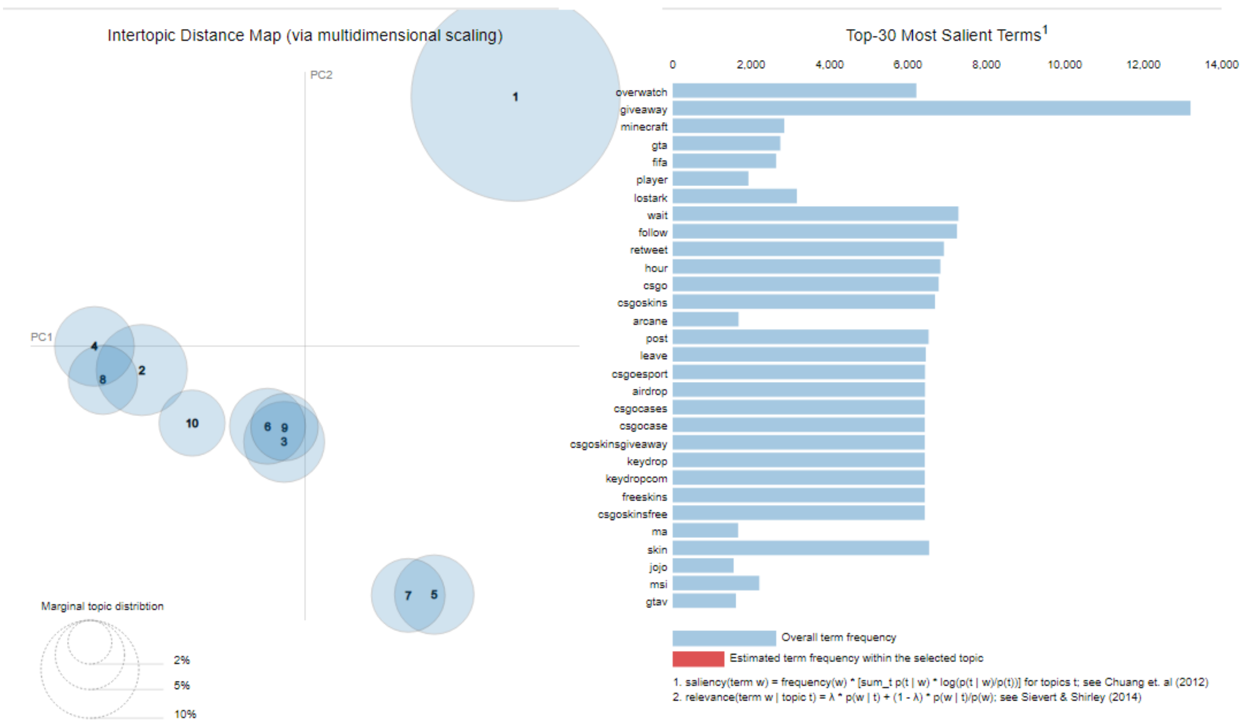
3.2.1 Further analysis for each sentiment

To take a step further after obtaining the results of

Positive tweets

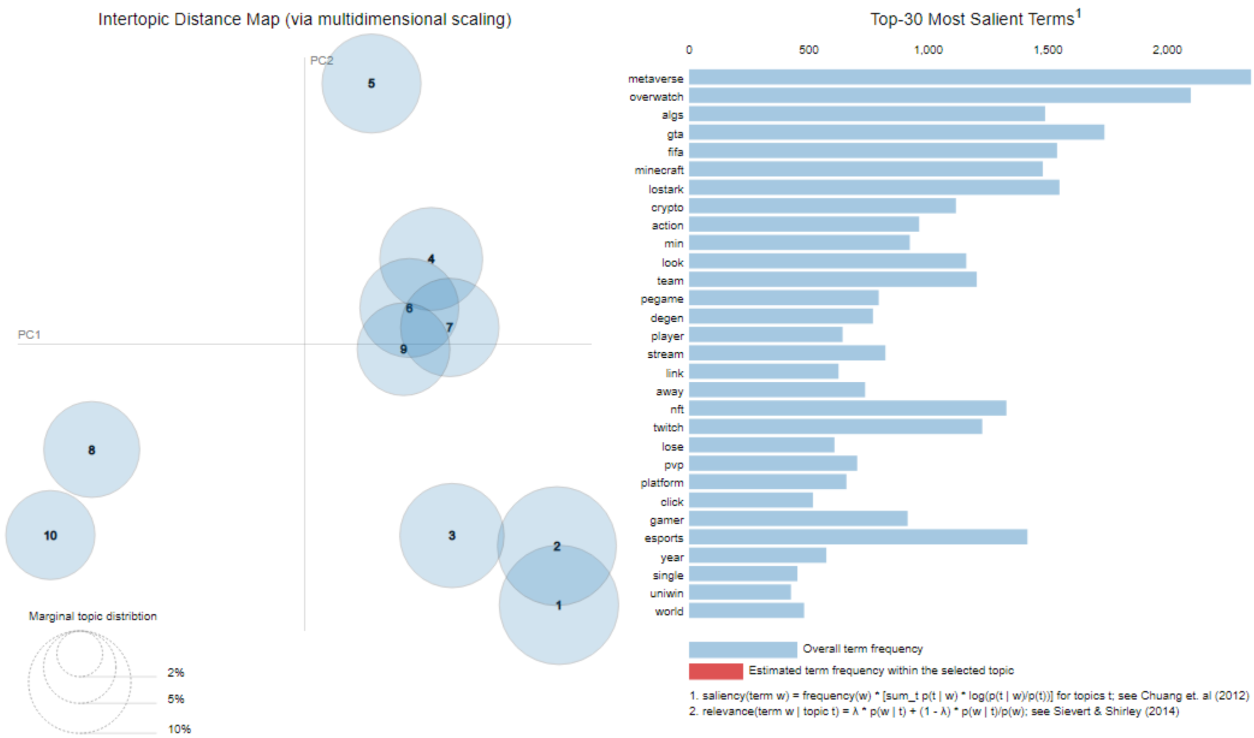
Neutral tweets

```
LDA Topics
Topic 0:
giveaway wait follow
Topic 1:
fifa algs fut
Topic 2:
player ma amp
Topic 3:
lostark stream playlostark
Topic 4:
arcane jojo style
Topic 5:
staydope gangstance rockstargames
Topic 6:
overwatch review msi
Topic 7:
minecraft leagueoflegends lose
Topic 8:
lostark msi ask
Topic 9:
gta beta gtaonline
```



Negative tweets

	LDA Topics
Topic 0 - 9	game link gta fifa overwatch stream year game world game lostark gamedev algs action min game legend episode game amazon msi metaverse game crypto game team nft



3.4 Network Analysis

4 streaming datasets are chosen instead of 1 streaming dataset for 1 month to do the network analysis. Choosing 4 streaming datasets to do the social network analysis and comparison, it can find whether the formation and structure of networks are similar to ensure the results are more reliable and prevent there being too much irrelevant data such as tweets from Bots.

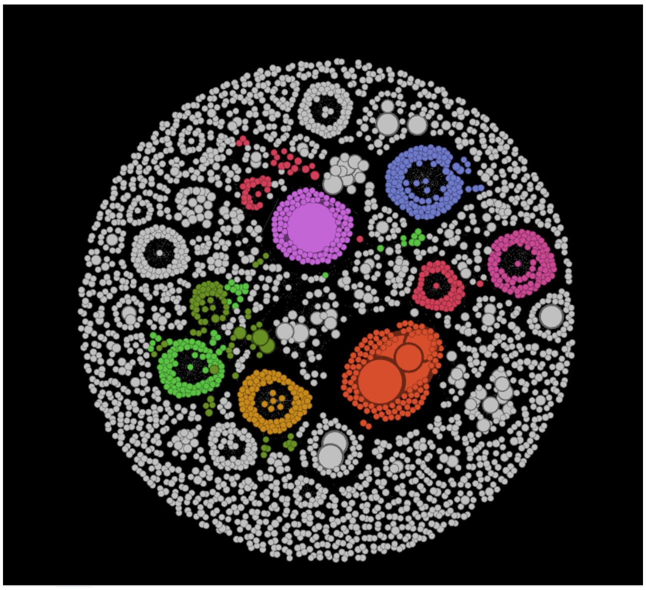


Figure 2. Streaming network week 1

In this directed network, there are 3003 nodes and 4438 edges. The average degree is 1.478 and the average betweenness centrality is 1.278. The eigenvector centrality is 0.116 with 100 iterations. There are 400 weakly connected components and 2982 strongly connected components with average modularity of 0.968 and 411 communities in total. However, only 8 communities are significant including more than 3 % users which are modularity class 211(4.43%), 214(3.76%), 41(3.7%), 73(3.33%), 25(3.3%), 26(3.23%), 251(3.2%).

For the influential nodes, they are identified with the eigenvector centrality higher than 0.7. In the network, 3 key influencers are found who are @poodletoken, @hoshi_u3 and @bsc_daily and 2 of them belong to class 41 except @hoshi_u3. @poodletoken is an account related to the crypto industry having 138459 followers with 132 degrees, 1 eigenvector centrality, 1 closeness centrality and 192 betweenness centrality in this network. For @bsc_daily, it is a news account related to the crypto industry with 381212 followers, 165 degrees, 0.772 eigenvector centralities, 0.6 closeness centralities and 6 betweenness centralities. For @hoshi_u3, it is an artist account belonging to class 26 with 0.898 eigenvector centralities and 192 degrees. They are also the top 3 in degree centrality.

For the sub-communities, 8 large communities are identified. For class 211 in blue, the most important node is @skyversenft with eigenvector centralities of 0.032 which is about Non-Fungible Toke. For class 214 in green, the key node is @defina_offcial which is a game introduced by blockchain. For class 41 in orange, it is about the crypto industry as mentioned. For class 73 in red and 25 in brown, the key influencers are Twitter simple users concerned about crypto. For class 26 in pink, it is about art. For class 251 in light red, the key influencer is @elrondnetwork which is a blockchain platform for distributed apps. In this network, they are not closely linked with others.

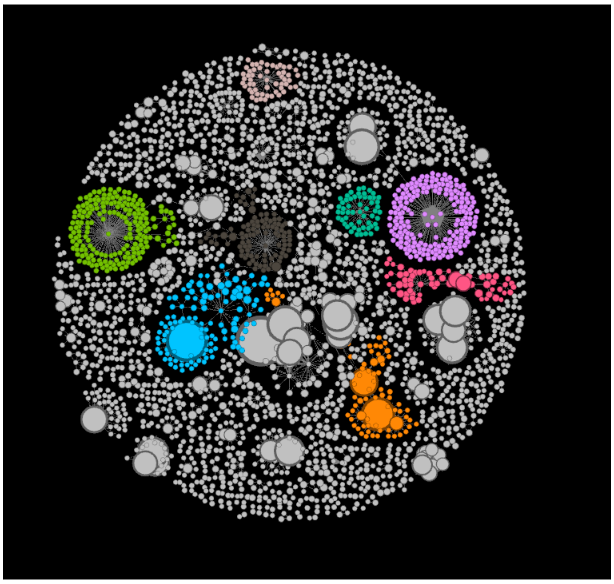


Figure 3. Streaming network week 2

In this directed network, there are 2946 nodes and 4450 edges. The average degree is 1.511 and the average betweenness centrality is 1.178. The eigenvector centrality is 0.295 with 100 iterations. There are 465 weakly connected components and 2908 strongly connected components with average modularity of 0.968 and 476 communities in total. However, only 4 communities are significant, which are modularity class 302 (6.89%),85 (6.62%),65 (4.04%),and 129 (3.29%).

In the network, 4 key influencers are found who are @samaras_mima, @steamgodfather, @aten-mima, and @samuelrunnacles in terms of eigenvector centrality. Besides, @steamgodfather, the other 3 accounts are relevant to the cryptocurrency and NFT with eigenvector centralities of 1 ,0.719 and 0.703 and degree centralities of 32, 29 and 17 respectively. @steamgodfather is an account promoting Steam, a platform for purchasing games having an eigenvector centrality of 0.815 and 118 degrees which rank 3rd in degree centrality. The top 2 in degree centrality are @amanda64392853 and @loudshake with 203 and 183 degrees who are related to the crypto industry.

For the sub-communities, 4 large communities are identified. For class 302 in pink and 129 in black, the most important nodes are @eroekpc and @cz_binance with eigenvector centralities of 0.007 and 0.117 who are concerned about Non-Fungible Token and bitcoin respectively. For class 85 in green, the key node is @pegaxyoffcial which is a horse-racing game in the Metaverse. For class 65 in light blue, its key node is @steamgodfather which shows that it is composed of gamers. In this network, these communities are connected with some intermediate nodes.85 is connected to 129 via a group of users then 129 is connected to 65 by the communities of @ samaras_mima (the central big node in gray color) with two-sided direction. Finally, 65 has directed arrows to 302. It can be seen that 129,65 and 302 are closely related.

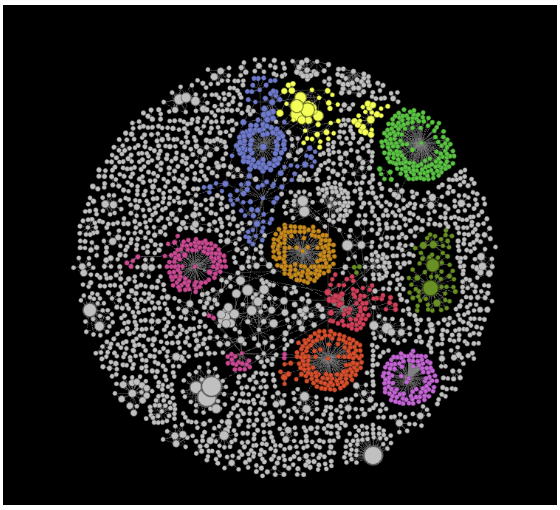


Figure 4. Streaming network week 3

In this directed network, there are 2826 nodes and 4067 edges. The average degree is 1.439 and the average betweenness centrality is 1.618. The eigenvector centrality is 0.421 with 100 iterations. There are 425 weakly connected components and 2791 strongly connected components with average modularity of 0.953 and 443 communities in total. 6 communities are significant which are modularity class 200 (5.45%),140 (5.38%), 233 (4.49%), 158 (4.07%), 310 (4.03%), 3 (3.26%).

In the network, 5 key influencers are found who are @biosstorm (eigenvector=1, degree =24), @shoutgamers (eigenvector=0.939, degree=9), @twitchtvonline (eigenvector=0.939, degree=9), @cero_engine (eigenvector=0.754, degree=27) and @blazedrts (eigenvector=0.723, degree=14) in terms of eigenvector centrality. Besides @cero_engine, other accounts are in the same modularity class of 228 while @blazedrts is related to the crypto industry and the remaining accounts relate to the gaming esports industry. Creo Engine is a gaming platform dedicated to game developers worldwide to establish their games and hence also has a connection to the esports industry. In terms of degree centrality, @patterson_io (degree=127), @heritag37682270 (degree=91), and @morningstargg (degree=81) are the top 3 and the first 2 are crypto-related while the 3rd is a streamer.

For the sub-communities, 6 large communities are identified. For class 200 in dark blue, @morningstargg is the key node which is a twitch streamer for gaming. For class 140 in green, class 158 in pink, class 3 in purple, and class 310 in brown, they are related to NFTs. For class 233 in red, @solcraftroyale is a new multiplayer Play2Earn project. The structure is so closely connected for 200,233, 158, and 3 that they stay in the middle with expanded linkage to 140 and 310.

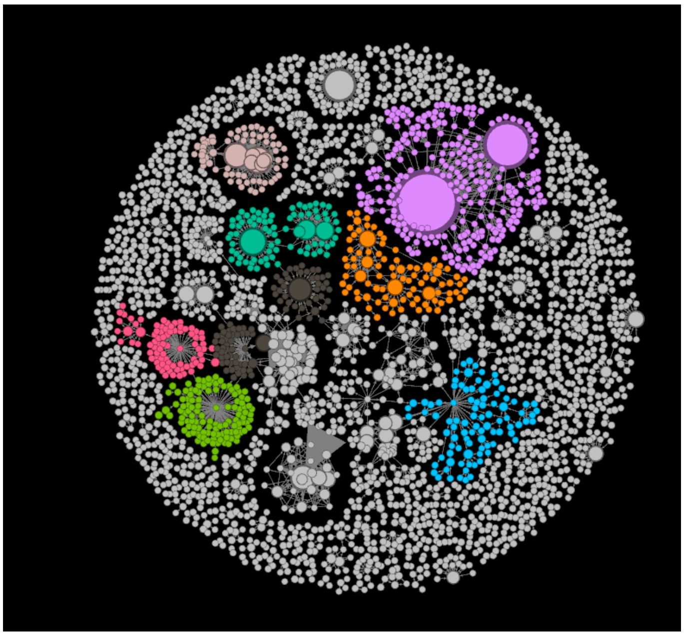


Figure 5. Streaming network week 4

In this directed network, there are 3023 nodes and 5050 edges. The average degree is 1.671 and the average betweenness centrality is 1.289. The eigenvector centrality is 0.111 with 100 iterations. There are 469 weakly connected components and 2992 strongly connected components with average modularity of 0.955 and 487 communities in total. However, only 4 communities are significant, which are modularity class 4 (8.67%), 3(3.84%), and 218 (3.04%).

In the network, only 1 key influencer is found who is @d6ixtoken (eigenvector =1, degree =255) in terms of eigenvector centrality above 0.7 and the 1st in degree centrality. It is an account of the first fully functional Esports Cryptocurrency. The 2nd and 3rd in degree centrality are @div6ix (degree =178) and @lilieshome_ (degree=116). For div6ix, it is not only the founder of @d6ixtoken but also a UK Esports Organisation. For lilieshome_, it is a fanpage of idols.

For the sub-communities, 3 large communities are identified. For class 4 in pink, class 3 in green and class 218 in blue, the key nodes are d6ixtoken, @amirsamurainob1 and @AfterlandGames, which are all related to play to earn games and cryptocurrency. The middle structure of this network is divided into two parts: class 4 is less connected to other subgroups, while classes 3 and 218 are connected by other subgroups.

After comparing the data over 4 weeks, the key influencers are consistent that the crypto-related accounts are identified for all weeks while the gaming and esports accounts appeared as key influencers after week 2. The artist and idol account only popped up in 1 week so they might not be the major key influencers for the Esports Twitter networks.

3.4.1 There exists an interconnected network around Esports

Our analysis resulted in overall social networks around Esports totaling about more than 10000 nodes and edges whose overall activity are portrayed in the total complete networks map. The vast majority of users were part of an interconnected Esports social network. As evidenced by the wide ranges of in-degree and out-degree in the ego network and streaming networks, there was considerable variance in users' contributions to the hashtag discussions around Esports.

	Overall Degree Centrality	Ranges of Degree Centrality
Streaming network 1	1.478	0- 192
Streaming network 2	1.511	0-203
Streaming network 3	1.439	0-147
Streaming network 4	1.671	0-255

Table 5. Degree centrality of the streaming networks

3.4.2 References to the crypto industry was prevalent

There is at least one large sub-community in the network predominantly shared, mentioned or replied to others around the crypto industry. Although our study focus was on Esports promotion within the context of development, Esports is becoming a prospective niche for the crypto industry that the crypto industry treats Esports space as its market and capitalizes on it. This contributes to the growth of the Esports industry so the data is kept for further analysis.

3.4.3 There are a number of influencers from different backgrounds who tend towards in-group interaction

The above network analysis revealed different key individuals from different backgrounds with disproportionate influence on the content that flows within the network. Compared with the five networks, the key influencers in the Twitter networks about Esports are 1. Esports players, 2. the crypto industry and 3. gaming industry including managers, online shop platforms and streaming platforms with high eigenvector centralities and hence the importance. In the analysis, the cryptocurrency-related nodes are found to appear in all networks which indicates their importance to the recent Esports industry and have significantly more connections and communications on Twitter.

3.4.4 Distinct communities and categories of users exist in the network

The communities found in the network were not identified a priori, but rather based on user behaviour on Twitter, that users who interact more frequently among each other compared with all other users to form one community. This suggests that although there is a much larger interconnected system around the topic of Esports, there are also unique groups of users that have patterns of interactions and form a community. Esports and gaming relevant (organizations, players) and crypto industry related (NFTs, Play to Earn and cryptocurrency) Twitter users are identified to form the majority of subgroups in the results.This distribution remained consistent among the communities, indicating that interactions do not occur within an industry alone, but rather extend to connect individuals and companies as well. Hence, the subcommunities are connected to form the networks.

4. Discussion and Interpretation

4.1 Sentiment Analysis

After analyzing the result of sentiment analysis from a quantitative approach, qualitative analysis is performed in the following sections. Among the tweets that have a positive sentiment, it is observed that some of the words with a high polarity score. For instance,the word “great” that appears in the positive word cloud below has a polarity score of 1. The word “welcome” also carries 0.8 polarity scores, meaning that both these words carry an extreme positive sentiment. One of the reasons a more significant portion of these tweets has a positive sentiment is due to the operational characteristic of the Esport community. Under the pandemic situation, different social aspects were affected thus people are inevitably likely to use Twitter as a resort for their negative emotions.It is however that the video game and Esport industries are rather affected, benefiting from the pandemic situation as it acts as a way of entertainment. Bryl et al (2021) reveal that the Esports industry gained from the pandemic as video games act as sustenance for many people. Besides, it is found that Esport events have a positive temporal effect on people’s sentiment about Esport as they might be anticipating for their favorite Esports teams to participate in the event or cheer for them. Thus tweets that carry a positive sentiment were published on Twitter.

Fig x.x Word cloud of positive sentiment

Result indicates that more than one-third of the tweets have a neutral sentiment. For tweets that have a neutral sentiment, it is discovered that most of the tweets are promotion campaigns or advertisements, which are primarily objective. These messages often contain words that are nouns or verbs that have a neutral sentiment. These words often do not appear in the lexicon of Textblob and were not labeled with a polarity score as they cannot be classified into a sentiment category when they are assessed individually. Mattila & Salman (2018) performed research on the sentiment of advertisement tweets and result indicates that most of them carry a neutral or positive sentiment. Although, another point raised by Mattila & Salman (2018) suggests that social media managers of these business accounts might post tweets that carry a positive sentiment as a positive correlation is found between the number of likes, and comments with the sentiment of tweets.

Fig x.xWord cloud of neutral sentiment

Negative takes up the least portion among the analyzed tweets. These tweets often carry words that have a negative connotation. For instance, words such as “hate” carry a polarity score of -0.8, and “pathetic” with a polarity score of -1.0 appear in the tweets with negative sentiment, as shown in the word cloud below. One of the reasons associated with the appearance of negative tweets is the same as positive tweets, which is the coming of online or offline Esports events. Thelwall et al’s (2011) research discovered that popular events are correlated with the increase of negative tweets. One of the reasons for the negative tweets about the event to appear is suspected to be complaints made about the event or the performance of certain Esports teams. Although, tweets might be wrongfully classified as negative tweets due to the lexicon-based approach used does not fully cover the contextual meaning of certain words. For instance, the word “Evil” appears in the word cloud below but it is actually an Esports Organization.

Fig x.xWord cloud of negative sentiment

4.2 Social Network Analysis

4.2.1 Why are they the key influencers?

This research was done to understand Esports' development through communication and knowledge sharing on Twitter. The goal was to identify opportunities for more strategic, targeted and effective Esports promotion in the future. In this research, networks of individuals that hold the potential to promote, support and enhance knowledge exchange around Esports on Twitter are created.

In this study, we identified 3 types of key influencers by the eigenvector centrality. They are esports players, the crypto industry and the gaming industry. As these influencers pose the largest influence on the Esports network, identifying these users may provide the insight for targeting communities and generating promotion strategies. Traditionally, influencers on social media are identified based on the number of followers. However, previous works had proven that the relationship was low for followers and the importance of users and centrality became the key factors for identifying key nodes (Cha et al., 2010). If the number of followers was the consideration of importance, many of the individuals identified in the networks would have been skipped. Overlooking these socially influential users may have resulted in missing out on strategic opportunities with high latent potential for impact (Antonakaki, Fragopoulou, & Ioannidis. 2021).

4.2.1.1 Esports players (streamers)

For professional Esports players and streamers, the top players are superstars with legions of fans around the world. Esports leagues and competitions have emerged all over the world. Similar to other sports competitions, Esports also has its own groups of audiences and so do the players. According to McCutcheon, Hitchens, and Drachen (2018), There was about 8.5 million online viewership for the League of Legends championship which was more than the number of viewers of the ITV broadcast of the FA Cup Final in 2013. Similarly, Esports players also have their fans compared to football players and they are less dedicated than professional athletes. The major difference is that professional Esports players streaming video of playing the games are essential while athletes are not necessarily practicing in front of their fans (Heaven, 2014). In the view of fans, the interaction with their idols brings strong attraction to them and it is unique in Esports. Streamers build a more engaging, intimate and trustworthy connection with their followers (Xu, Kim & Billings, 2022). The high exposure of professional Esports players increases their influence online and attracts more users to contact the Esports industry. Therefore, they become the strong influencers on the social network of Twitter about Esports. In return, according to the Social Capital theory, the Esports players/streamers can gain more money and more fans by being the influencers in the network.

4.2.1.2 The crypto industry

For the crypto industry, it is capitalizing on Esports and the two industries are moving closer to each other. According to Dom (2021), the blockchain and Esports enterprises are converging their markets that there are more and more gaming directly relevant to NFTs and Play To Earn are developed. This changes the Esports industry and increases the influence of the crypto industry in the Esports social network since their relationships are difficult to separate. Added by Dom (2021), Esports and cryptocurrency have a special connection because of their high level of competitiveness and rapid changes. For competitiveness, both of them involve harsh competition where Esports players compete for champion and investors compete for money. For rapid adoption curves, both industries change at a rapid speed and people have to adjust immediately. In short, it is influential in the Esports networks because they are mixing and forming a new type of gaming. Hence, the crypto accounts can attract more people to invest in cryptocurrency and expand the industry widely.

4.2.1.3 The gaming industry (managers, online shop platforms and streaming platforms)

For the gaming industry, it developed and grew rapidly in the past decade and became more mature in producing more game genres and hence increasing the attraction of the streaming platforms to the users. With the advance of technology in these years, the gaming industry grows and creates its own markets economically. To match the requirements of different consumers, more game types were developed and it was found that the Esports genres showed a positive relationship with affective motivation (Sjöblom, et, al., 2017). In their research about the association of games, gratifications and streaming, it was suggested that competitive game streaming especially Esports games provides a more intense watching experience for the audience similar to sports games and hence a feeling of stress and escapism which meant that the audience was able to escape into the competitive gaming streams and watch their favorite teams and players, removing themselves from the worries of life. This increases the audience's concentration on the streaming of Esports games.

Besides, the growth of technologies enhances the streaming function to attract and affects more users worldwide. Broadcasters are operators with platforms to distribute audiovisual content in live, on-demand and online modalities such as Twitch. According to McCutcheon, Hitchens, and Drachen (2018), Esports streaming has contributed to the major change in the broadcast media environment and better streaming ensures deeper engagement with Esports. These factors enable the gaming industry to pose large influences on online users and promote Esports to them. As reciprocity, the Esports industry can be more recognized by the public as a promising field with prospect development instead of just an entertainment.

4.2.2 How do the subcommunities form?

The research presents the identification of a dynamic online social system in the context of Esports on Twitter. In a short 1-month time period, more than 10000 unique users focused on Esports are identified. These Esports-focused users formed distinct identifiable communities and transferred knowledge effectively and efficiently within the groups and even expanded to other users.The potential of social influence in the social media space around Esports is reflected in this research.

As the reasons for being influential has been explained above, this part will focus on how the Esports and gaming industries affect the users.

4.2.2.1 Esports and gaming relevant users

The aims of streams are to obtain positive association and provide satisfactions by transferring cognitive information successfully (Sjöblom et al., 2017). The appearance of video games streaming is significant in representing the growth of user-generated content and the democratization of the media. In the research done by Xu, Kim and Billings (2022), they concluded that the parasocial relationships and loyalty from viewers to streamers explained the positive effects of streamers' credibility on the brand attitude as well as the Esports industry. The credibility builds close relationships between streamers and viewers especially for professional gamers that better skills and strategies can attract more users. Closer relationships usually bring trust and loyalty that viewers can attach to the streamers directly and hence influences their attitude. Therefore, the subcommunities are formed by positive and loyal Twitter users to the Esports industry.

4.2.2.2 The crypto industry

For the crypto industry, it is influential because the application of cryptocurrency on Esports gaming attracts both Esports fans and crypto enthusiasts (Chris, 2019). For Esports fans, they focus on the big pay from professional Esports such as the prize pools from Esports games. For crypto enthusiasts, they concentrate on the rise of cryptocurrency. Both of them are distracted by money. This expands the original networks of Esports from people who enjoy gaming and watching competitions only to capturing users who prefer to earn money. People who are originally concerned with only one of the industries only can acquire reciprocity from the other. Esports fans can have opportunities for earning money and crypto enthusiasts can have entertainment while playing games.With an increase in the number of games mixing with cryptocurrency and a broader audience, the number of subgroups of the industry rises and the connection is stronger starting with the crypto industry.

4.3 Interpretation

In the above research, it is found that the major attitude of Twitter users towards Esports is neutral because most of them are advertisements. Also, negative attitudes related to dissatisfaction are identified. These findings can be interpreted to find methods to attract them and those neutral users affected by those promotions to be positive. To promote Esports positively, positive wordings such as ‘exciting’ and ‘relaxing’ can be added to the advertisements in order to give a good image to the public at first sight. Besides, benefits gained from Esports can also be mentioned to remove the negative beliefs about Esports, for example, leading to game addiction. By promoting the positive aspect of Esports, the attitudes of Twitter users can be affected as they might allot more time to have more engagement in the Esport industry.

Upon understanding the influential users and the sub-communities in the Esports social network, clearer promotion strategies can be set for improving brand images and attracting more people. Three-sided corporations can be adopted among the Esports industry, the crypto industry and the streaming platforms. They are not only the key influencers and the major communities in the networks, but also the Esports and crypto industry are already using streamers for promotion respectively. They can think about the relation of their contents to personal integrative motivations. For example, both industries can spectate the habits and motivations of their targets and identify the same characteristics. Then, asking the streaming platforms to seek the most famous and suitable streamers to promote them. As a result, all of them can gain from the increase in the number of followers by attracting each targeted audience to touch three of them.

5. Limitations

5.1 Bot detection

During data collection, tweets from bots are not identified and excluded. In Twitter, there are Bots which are not human accounts and share the assigned information automatically. In the Twitter data collection, it is unavoidable to collect Tweet data from Bots if they used the selected keywords. In this research, although data is preprocessed for sentiment analysis, for network analysis, it is difficult to preprocess to identify the Bots and remove them. Hence, the network formation might be slightly affected but not the big picture about the key influencers and the structure of sub-communities.

5.2 Languages

Only English is focused in this Twitter research. To collect data, English is chosen as the language criteria that the data only includes tweets in English. This creates a limitation that the attitude or influencers might be different from the results in those countries using other languages such as Asian countries.

5.3 Hardware

Thirdly, our devices are not professional enough to handle large amounts of data and hence unable to run larger networks for analysis. In data analysis, more data is preferable, but it significantly depends on the functionality of the hardware of the computers to analyze large amounts of data. This limits the data analysis in a certain amount, and hence the solution is to divide the large datasets into 4 small datasets for analysis. This also helps to increase the reliability of the research if they get similar results.

5.4 Data collection period

As Only 1 month is used for data collection, the effects of key influencers might not be in the long term. Another challenging factor lies in the number of current key influencers and in their mode of interaction. In this study, the Twitter data during a 1-month period only is collected, and identified only a few types of key influencers as well as the users' attitudes. It may be possible that a certain influence and attitude only exist for a short period. A longer period (e.g., 6 months) can find additional influencers and prevent only a few users from dominating the discussions.

6. Conclusion & Future research direction

Reference

Analyzing Social Network using Gephi (Thu & Wai, n.d.)

Antonakaki, D., Fragopoulou, P., & Ioannidis, S. (2021). A survey of Twitter research: Data model, graph structure, sentiment analysis and attacks. Expert Systems with Applications, 164, 114006.

Ardianto, R., Rivanie, T., Alkhalifi, Y., Nugraha, F. S., & Gata, W. (2020). Sentiment Analysis on E-Sports For Education Curriculum Using Naive Bayes and Support Vector Machine. *Jurnal Ilmu Komputer dan Informasi*, *13*(2), 109-122.

Arnaboldi, V., Conti, M., La Gala, M., Passarella, A., & Pezzoni, F. (2016). Ego network structure in online social networks and its impact on information diffusion. Computer Communications, 76, 26-41.

Bakshy, E., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011, February). Identifying influencers on twitter. In *Fourth ACM International Conference on Web Seach and Data Mining (WSDM)* (Vol. 2).

Blei, D. M., Ng, A. Y., & Jordan, M.I. (2003). Latent Dirichlet Allocation. Journal of Machine Learning Research 3 (2003) 993-1022.

Bourdieu, P. (1986). 'The forms of capital' in J.G. Richardson. Westport (ed.) Handbook of Theory and Research for the Sociology of Education, CT: Greenwood Press, 241-258.

Bryl, Ł., Majewska, J., & Truskolaski, S. (2021). Effects of the Covid-19 pandemic on sport, video game, and tourism industry: Sentiment analysis of press, internet articles, and Twitter data. *Towards the "New Normal" after COVID-19—A Post-Transition Economy Perspective*, 137-151.

Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, K. (2010, May). Measuring user influence in twitter: The million follower fallacy. In Proceedings of the international AAAI conference on web and social media (Vol. 4, No. 1, pp. 10-17).

Chen, Y. (2015). *Convolutional neural network for sentence classification* (Master's thesis, University of Waterloo).

Chris, M. (2019). What Crytpocurrency Could Mean To eSports (And Vice Versa). TWICE, 34(1), 22.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Doogan, C., Buntine, W., Linger, H., & Brunt, S. (2020). Public perceptions and attitudes toward COVID-19 nonpharmaceutical interventions across six countries: a topic modeling analysis of twitter data. *Journal of medical Internet research*, *22*(9), e21419.

Fornacciari, P., Mordonini, M., & Tomaiuolo, M. (2015, September). Social network and sentiment analysis on Twitter: Towards a combined approach. In *KDWeb* (pp. 53-64).

Grandjean, M. (2016). A social network analysis of Twitter: Mapping the digital humanities community. *Cogent Arts & Humanities*, *3*(1), 1171458.

Gruzd, A., Wellman, B., & Takhteyev, Y. (2011). Imagining Twitter as an Imagined Community. American Behavioral Scientist, 55(10), 1294–1318. https://doi.org/10.1177/0002764211409378

Heaven, D.(2014). Esports: Pro video gaming explodes with big prize pots. Retrieved from https://www-newscientist-com.eproxy.lib.hku.hk/article/mg22329823.900-esports-pro-video-gaming-explodes-with-big-prize-pots/#.VR9OzeEyRxI

Himelboim, I., Smith, M., Rainie, L., Shneiderman, B., & Espina, C. (2017). Classifying Twitter Topic-Networks Using Social Network Analysis. Social Media Society, 3(1), 205630511769154.

Hollander, J. B., & Renski, H. (2015). *Measuring urban attitudes using Twitter: An exploratory study*. Lincoln Institute of Land Policy.

Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media* (Vol. 8, No. 1, pp. 216-225).

Kharde, V., & Sonawane, P. (2016). Sentiment analysis of twitter data: a survey of techniques. *arXiv preprint arXiv:1601.06971*.

Kim, Y.(2014). Convolutional Neural Networks for Sentence Classification.

Lahuerta-Otero, E., & Cordero-Gutiérrez, R. (2016). Looking for the perfect tweet. The use of data mining techniques to find influencers on twitter. *Computers in Human Behavior*, *64*, 575-583.

Loria, S. (2014). en-sentiment.xml(Version 1.0)[Source code].https://github.com/sloria/TextBlob/blob/eb08c120d364e908646731d60b4e4c6c1712ff63/textblob/en/en-sentiment.xml

Mahmud, J., Fei, G., Xu, A., Pal, A., & Zhou, M. (2016, March). Predicting attitude and actions of twitter users. In *Proceedings of the 21st International Conference on Intelligent User Interfaces* (pp. 2-6).

Mattila, M., & Salman, H. (2018). Analysing Social Media Marketing on Twitter using Sentiment Analysis.

McCutcheon, C., Hitchens, M., & Drachen, A. (2018). ESport vs irlSport. In Advances in Computer Entertainment Technology (Lecture Notes in Computer Science, pp. 531-542). Cham: Springer International Publishing.

Niko.(2021). Esports in Asia. Retrieved from https://nikopartners.com/esports-in-asia/

Pew Research Centre (2014). Mapping Twitter Topic Networks: From Polarized Crowds to Community Clusters. From http://www.pewinternet.org/2014/02/20/mapping-twitter-topic-networks-from-polarized-crowds-to-community-clusters.

Powers, D. M. (2020). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *arXiv preprint arXiv:2010.16061*.

RiteTag. (2021). RiteTag. Retrieved from https://ritetag.com.

RiteTag.(2021). Popular hashtags for esports on Twitter and Instagram.Retrieved from https://ritetag.com/best-hashtags-for/esports.

Scott, J. (1988). Social network analysis. *Sociology*, *22*(1), 109-127.

Statista Research Department. (2021). Market size of eSports market in China from 2017 to 2020 with estimates until 2022 (in billion yuan). Statista. Retrieved from https://www.statista.com/statistics/1018659/china-esports-game-market-value/

Sjöblom, M., Törhönen, M., Hamari, J., & Macey, J. (2017). Content structure is king: An empirical study on gratifications, game genres and content type on Twitch. Computers in Human Behavior, 73, 161-171.

Subramaniam, M., Shahwan, S., Abdin, E., Goh, C. M. J., Ong, W. J., Tan, G. T. H., ... & Chong, S. A. (2020). Advancing research to eliminate mental illness stigma: the design and evaluation of a single-arm intervention among University Students in Singapore. *Frontiers in Psychology*, *11*, 1151.

Taha, A. A., & Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC medical imaging*, *15*(1), 1-28.

Thompson, D., Fisher, K., Purcal, C., Deeming, C., Sawrikar, P. (2012). Community Attitudes to People with Disability: Scoping Project. SSRN Electronic Journal. 10.2139/ssrn.2014423.

Thelwall, M., Buckley, K., & Paltoglou, G. (2011). Sentiment in Twitter events. Journal of the American Society for Information Science and Technology, 62(2), 406–418.

Twitch.(2021). Twitch. Retrieved from https://www.twitch.tv.

TwitchTracker. (2021). TWITCH GAMES STATISTICS. Retrieved from https://twitchtracker.com/statistics/games

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Witkowski, E. (2012). On the Digital Playing Field. Games and Culture, 7(5), 349-374.

Xing, W., & Ghorbani, A. (2004). Weighted PageRank algorithm. *Proceedings. Second Annual Conference on Communication Networks and Services Research, 2004,* 305-314.

Xu, Q., Kim, H., & Billings, A. (2022). Let's Watch Live Streaming: How Streamer Credibility Influences Brand Attitude in Esports Streamer Marketing. Communication and Sport, 10(2), 271-290.

Appendix A : Lexicon-based and deep learning approach to sentiment analysis

Sentiment analysis with other approaches

For RQ1: *What are the attitudes of Twitter users towards Esports?*; It will be completed with the Textblob approach as explained above. The sub research question of question 1, RQ1.1*: Comparison of lexicon approach and machine learning approach on locating sentiments among tweets*, aims at providing a more holistic view of the other approach for locating sentiments of tweets.

Model introduction:

- **VADER**

  Vader is another lexicon sentiment classifier with additional rule-based enhancements for sentiment analysis (Hutto & Gilbert, 2014).Vader's lexicon has more than 7000 words, and now it also supports emojis .Generally speaking, data for sentiment analysis has to be manually for subsequent feature engineering and model training, resulting in a cost in time and manpower. It is, however, that for the lexicon and rule-based classifier, VADER, like Textblob, there is no need for the above steps. And research has indicated that the results of SVM and other classifiers on multiple datasets were outperformed by VADER s(Hutto & Gilbert, 2014). Vader returns four labels as its output, including ***neg,*** which indicates negative emotion, ***neu*** which indicates neutral emotion, ***pos*** which indicates positive emotions, and ***compound,*** which returns values from -1 to +1, with -1 being the most negative and +1 being the most positive. Each pos, neu, and neg score reflects the portion of the text in the sentence containing each sentiment. While the compound score is calculated similarly like Textblob where the summation of scores will be made and normalized within the output scale -1 to 1. For sentiment analysis in this research, the compound score will be used as the classifier for sentiments.

- **BERT**

BERT refers to Bidirectional Encoder Representations from Transformers, and it is originally developed by Google AI developers (Devlin et al., 2018). Currently, there are many version of BERT that is developed on the basis of this model for NLP. One of the primary concepts that BERT is built on is transfer learning, as transfer learning is an efficient approach in NLP that pre-trained models will be used for training the current model. The word bidirectional in the model refers to how the model will take the surrounding words into consideration when understanding a word, this processing aids us to discover the contextual meaning of that word. Transformer refers to the transformer model that will be used in BERT, and the transformer model is proposed by Vaswani et al (2017) that utilizes the attention mechanism. With this model, the masked tokens from the original text are scanned at once instead of beginning from one direction. This mechanism further allows the understanding of the contextual meaning between words.

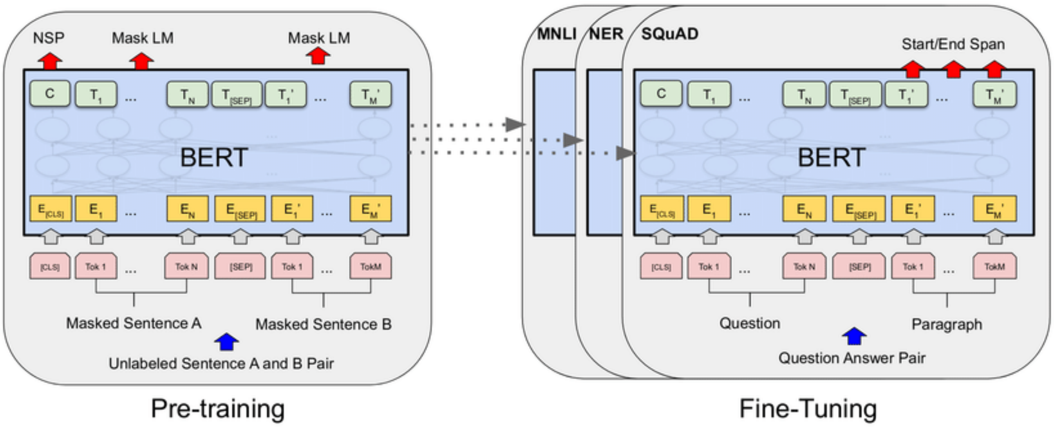


Fig x.x Pre-training and Fine-tuning steps of BERT (Devlin et al., 2018)

The pretraining step of BERT includes Masked Language Modeling (Masked LM), in which a few words in a sentence are randomly masked, and the BERT model is trained to guess what the masked words are. The other is Next Sentence Prediction (NSP), in which the second sentence of two consecutive sentences is replaced by another sentence at a fixed rate, and the model is trained to guess whether this sentence is the next sentence of the first sentence (Devlin et al., 2018).BERT can later be fine-tuned based on specific tasks like natural language inference. All parameters in the pre-trained model will be used in this step.

- **CNN**

Convolutional Neural Network (CNN) is a model that is powerful in image recognition but it can also be used for sentiment analysis.CNN's structure will not be mentioned here, instead, some of the different between model input of sentence classcification and image classcification will be inspected.As NLP input is a text or document, unlike inputs for compurter vision which is picture pixels. The sentences or documents are embedded with the embedding layer at the input and represented as a vector matrix with each row representing a word, the total number of rows representing the sentence length, and the total number of columns representing the dimensionality (Chen, 2015).For this project, the dimensionlity is set to be 300. So when a 300 dimensional embedding on a sentence with ten words, we get a matrix with a 10x300 input.

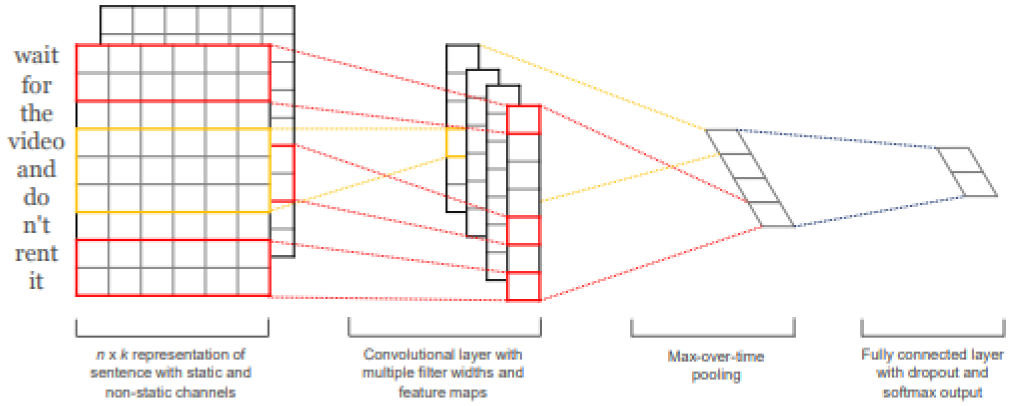


Figure x.x Architecture of CNN for sentiment analysis (Kim,2014)

In image recongnition tasks, filters are applied as a patch to the full picture no matter the size of the image, but in NLP, filters are applied to all dimensions,the shape of the filter equals to the size of the filter times the size of the embeddings. For instance , if the input matrix is 10x300 like above, the filter heights can be 2 to n base on how the sequence of words are grouped, and the width is 300 like the input matrix. The output of each filter will be pooled to produce feature vector for determing the sentiment.

- **Comparison metric**

In order to answer the sub-research question RQ1.1*: Comparison of lexicon approach and machine learning approach on locating sentiments among tweets*. Evaluation metric including pricision,recall, F1 score and accuracy.The value of these metrics were computed base on these value : True positive (TP), True negative (TN), False positive (FP), False negative (FN). True positive (TP) refers to the number of positive tweets that are identified accurately as positive sentiment.True negative (TN) refers to the number of negative tweets that are appropriately identified as such (Powers , 2020).The same goes for FP and FN.With these values, the values of different evaluation metrics can be obtained.

- **Accuracy**

Accuracy refers to the proportion of the correct prediction made by each model.It can be calculated with the following formula

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

- **Precision**
  Precision refers to the portion of tweets that is identified as positive is fatually positive.The calculation of it is by deviding the Number of True Positives (TP) by the Total Number of True Positives (TP) and False Positives (FP).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

- **Recall**

Recall aims to address the portion of actual positive tweets are identified.It is observable that the accuracy and recall formulas are very similar. The sole variation is in the denominator's second component, which is False Positive for precision but False Negative for recall.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

- **F1 Score**

In order to evaluate model performance comprehensively, it is necessary to look into both precision and recall. The F1 score is a useful statistic that takes into account both of them.F1 score is defined as to provide a more balanced description of model performance using the harmonic mean of precision and recall (Taha & Hanbury,2015).It is calculated by the following equation.

$$\text{F1 Score} = \frac{2 \times (\text{ Precision } \times \text{ Recall })}{\text{Precision } + \text{ Recall}}$$

- **Measurement of each model**

Table x.x CLASSIFICATION RESULT OF APPROACHES

Approach	Accuracy(%)	Percision(%)	Recall(%)	F1-score(%)
VADER	49%	44%	44%	44%
BERT	98%	98%	98%	98%
CNN	74%	72 %	75%	72%

BERT performs the best

- **In-depth review of the BERT MODEL**

  Almost all prediction are correct, the BERT model perfroms exceptionally on classifying sentiment of text in this dataset.

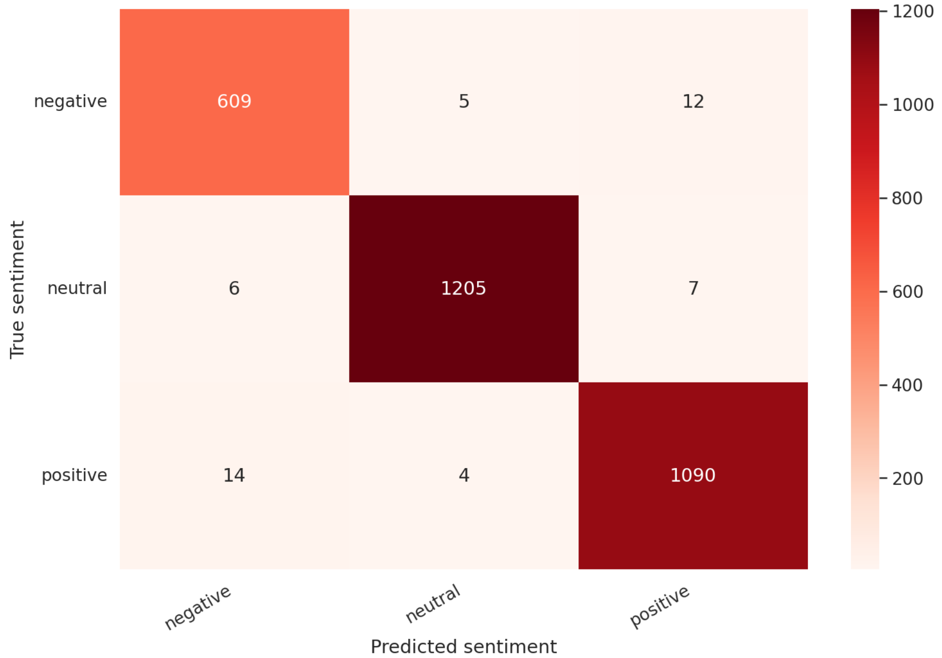


Fig x.x Confusion matrix of bert model predictions

Train and validation accuracy increase when both validation and train loss decreases throughout the 10 epochs.

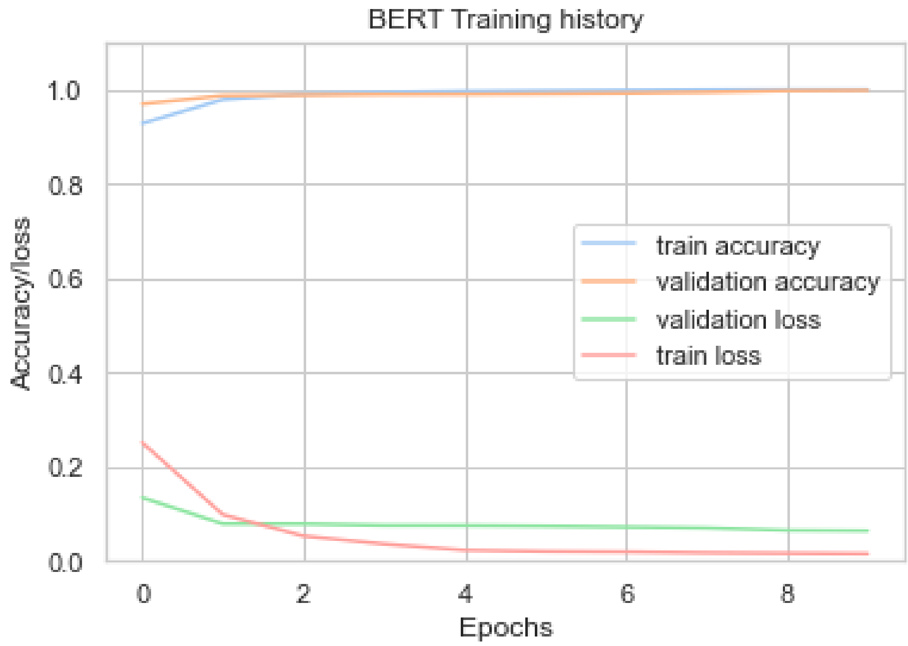


Fig x.x Training process of BERT