Department of Computer Science

Submitted in part fulfilment for the degree of MEng

Analysing Language Patterns in Variety Streams: What Makes Audiences Come Back?

Myles Ragay

19 May 2021

Supervisor: James Walker

ACKNOWLEDGEMENTS

I want to thank my mentors Dr James Walker, Dr Anders Drachen, and Charlie Ringer for their constant guidance through the course of this project. Their advice over the course of the year has shaped my growth and has polished skills that I will continue to use for the rest of my life. I would also like to note their dedication as they have continued to communicate with me even as the deadline is right around the corner.

I want to thank my housemates for continuous support and reassurance whenever I was in doubt of my skills. Especially to Luke Davies who I bounced ideas with at the start of this project, having someone to talk ideas with me really helped me visualise the project.

Finally, I want to thank my parents. The way they have raised me helped complete this project, they have conditioned me with perseverance to work through the pressure and stress of a deadline to complete this project to the best of my ability.

STATEMENT OF ETHICS

This paper uses a dataset that is openly available to the public. Despite effort to anonymise usernames, the files contain information that will help identify the streamer, but that is the extent of the personal information you can obtain.

There is no motivation to identify the streamers in the project. However, through inference, the identification of the streamer whose chats I analysed may have been achieved as consequence of my project.

All my results have been provided with the relevant data and a Jupyter Notebook to replicate my experiment. All tools, resources and the like have been properly referenced using the industry standard IEEE referencing procedures.

TABLE OF CONTENTS

TABLE OF TABLES

Executive summary

This paper explores the problem of quantifying the traits needed to gather an audience on Twitch. Currently, there are far too many people out there trying to make it big on Twitch with no concrete guide. Many of the sources existing today are purely based on qualitative advice, none of which has any supporting evidence based on numerical values.

In addition to this, the paper explores the lack in academic research in the field of livestream chat data by proposing a method to identify key traits that will help streamers gather a larger audience by using quantitative data. This method will be applied to variety streamers for two reasons: 1) will serve as a basis for future research of variety streamers as it will be the first entry into academia, 2) variety streamers are a good control measure to create a fair experiment.

The aim of this project is to find which traits, whether it be personality traits or traits that describe the environment of the chat, gather an audience. In order to do this, one must create a machine learning model that identifies the most prevalent topics in a chat. This model is the LDA model and has been justified and optimised to best of my ability against criteria set in order to collect the results needed.

The model is then applied to streams I chose against a criteria I personally set, which have also been justified. The results are then displayed intuitively in human-readable format so convenient inference of data. Any considerations and potential improvements have also been discussed, to give those doing further work on the field of variety streamers a strong foundation to start with.

After applying the model to three different chat logs, and analysis of nine topics created by the model (three per chat log), there are four identified traits that will help streamers gather more audience members. This paper then discusses the possible implications of the method and results.

1 Introduction

The concept of watching entertainers play video games has been around since the earliest days of YouTube [1] and livestreaming just enhances that experience. Livestreaming, which is most referred to as streaming, is the act of recording yourself and broadcasting that recording simultaneously. There are many streaming platforms but one of the most popular is Twitch.tv [2], most known as Twitch.

Twitch primarily focuses on live streams of games, however there are categories for art, music, and even debates. Twitch also has a lot of features to enhance the viewer's experience such as donations. This is where the user can reward the streamer with a currency called bits [3]. Another feature is the appearance of a chat window.

The chat window (or 'chat') is a scrolling window where audience members can send their messages, to the streamer or amongst each other, and view previously written messages. The window is only a certain size, so once enough messages fill the window previous messages will scroll out of view. These messages sent are unique to the livestream community as most use 'leet speak' [4], which is essentially gamer terminology. This 'leet speak' would be commonly used alongside Twitch emotes [5].

Some streamers tend to stick to one 'main' game, for example Grandmaster Hikaru Nakamura streams chess [6] almost always, however there are streamers that play a variety of games, these streamers are appropriately named 'variety streamers'. These variety streamers play many different games on stream, and successful variety streamers have a consistent fanbase that attends every stream due to some type of quality they have or the type of environment they have created, rather than for practicality (such as watching a professional in order to get tips on gameplay).

Being able to discover these attributes may also allow us to quantify them, for example Twitch may create an extension to sort streams in terms of these qualities; an example of this is if a viewer wants to watch something funny, they may be able to sort streams in terms of 'comedy'. Not only this but if these statistics were available to public view, it could teach streamers which qualities they are lacking in and how to improve to gather a larger audience.

1.1 Hypothesis

I believe that by analysing the messages being sent in chat, we can identify certain patterns in the language used that may demonstrate why audiences keep coming back. The method in which I will carry out this investigation is by using natural language processing (NLP), which will be explained in Section 3 about related works. By using NLP, it is possible to identify the most prevalent topic of words which may allow us to identify the reasons why these audience members keep coming back.

Variety streamers attract audiences across many different games. Therefore, we can assume that the audiences do not tune into the stream for practical reasons like learning how to get better as there are many dedicated professional streamers that exist. This means that variety streamers attract their audiences due to their character or the environment created on stream. The aim of this project is to find out what these attributes are.

Therefore, in this project, the variety streamer being focussed on is a control variable so the human aspect (their influence on their audience) stays constant during the experiment, the independent variable (what changes) will be the game being played and the dependent variable (what is being observed) will be the topics produced. Having the streamer stay constant throughout the experiment should make the observations fairer.

This thesis is reasonable as firstly, as of January 2021, there is nothing about variety streamers in academia. Since there is no existing works on variety streamers in academia, my primary driving motivation is to be the first one to able to add knowledge of variety streamers in academia. Secondly, the experiment has been made as fair as possible and is purely investigative rather than opposing existing theories. This will hopefully provide a good introduction to the theme of variety streamers in academia.

2 Twitch Chat and The Dataset

This section briefly describes the main features of Twitch chat and how it has been translated into data that can be manipulated by discussing the dataset.

2.1 Twitch Chat

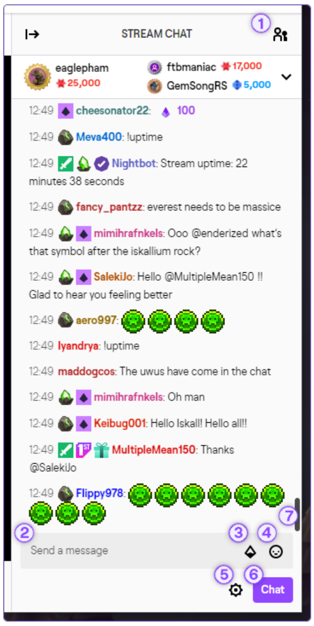


Figure 1: *A screenshot of Twitch chat*

Figure 1 [26] displays the features of Twitch chat. The main features are the names of viewers, the messages, the time stamp and the scrolling window. One can see the names of users highlighted in colour, this includes audience members but also any staff such as the streamer, moderators and admin. The messages follow the colon; these are what the viewers send to interact with the streamer, fellow viewers and even with the stream itself (via commands i.e. '!uptime', which allows them to see how long the stream has been running for). The time stamp precedes the name of the users, this allows other users to keep track of the messages in a temporal manner. Finally, there is the scrolling window which contains all the mentioned features within itself.

2.2 The Dataset

The dataset is comprised of two parts: a folder that contains a .csv file of chat logs from 1,951 streams, and another folder that contains the metadata (data that describes other data) for these streams. The files containing chat logs are saved as {streamer}_{number}.csv; {streamer} is an anonymised streamer name and {number} is the number of the stream, where ‘0’ would be the streamers first stream, ‘1’ would be their second, etc.

Time	User	Message
0	accaadf29ca2c7d45a3a4082afba0f25bb483507	mcriversidehey
11	accaadf29ca2c7d45a3a4082afba0f25bb483507	i got all skins kappa
67	accaadf29ca2c7d45a3a4082afba0f25bb483507	i think its a fake here
79	accaadf29ca2c7d45a3a4082afba0f25bb483507	i thik its a fake
217	13ddffd54e1852da63e76fc7a9ec72d2c97d78bb	is this just a video
342	13ddffd54e1852da63e76fc7a9ec72d2c97d78bb	gg your twitch account
357	13ddffd54e1852da63e76fc7a9ec72d2c97d78bb	obvious scams are obvioua

Figure 2: *Screenshot of a spreadsheet containing a portion of chat – the raw data*

Figure 2 shows the first 8 rows (including the header row) of a file that contains a stream’s chat log, and each file has data gathered from a single stream session. The ‘Time’ column shows the time, in seconds, between the start of recording the chat and when the message was sent. ‘User’ contains an anonymised username of the user who sent the message. Finally, the ‘Message’ column presents the message that was sent. Notice the similarities between Figure 1 and Figure 2.

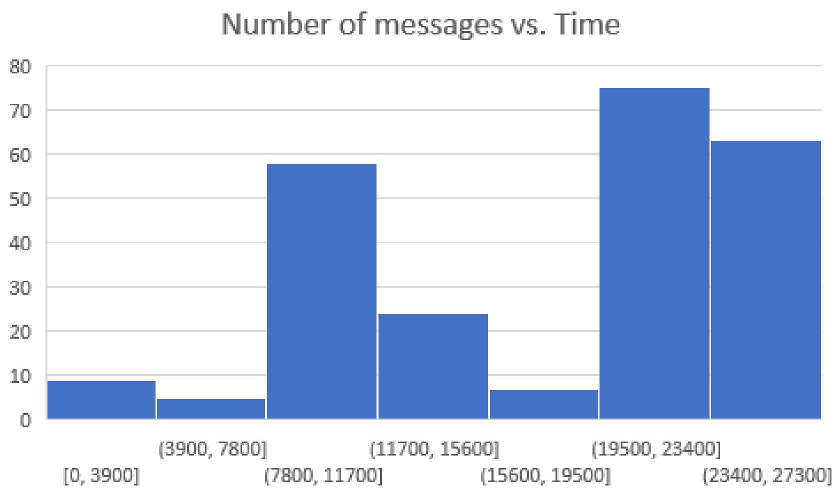


Figure 3: *Table that showcases 'Number of messages vs. Time'*

The data can be expressed in various high-level formats, one such way is measuring the volume of messages within a certain time by graphing the number of messages sent against time. For example, in Figure 3 you can see there is roughly 75 messages sent in between 19,500th and 23,400th second of the chat being record. This may be useful to identify important events within a stream as the volume of messages are unusually high compared to the other time frames, but this is not within the scope of my project so it will be left at this speculation.

The metadata, which is explained in Table 1, is a document that contains information about the stream, which was collected at the start of the process to gather and document the chat logs.

Metadata type	Explanation
user_view_count	The number of views the stream has accumulated throughout their career
user_broadcaster_type	Details whether the stream is a ‘partner’, ‘affiliate’ or otherwise – relating to one of Twitch’s programs that let streamers monetise their content [25]
stream_game_id	Each game has a unique ID number that allows the Twitch API to recognise the game being played on stream
stream_type	Details whether the stream is being streamed ‘live’, or if it is a ‘replay’ of a past stream, or otherwise
stream_viewer_count	The number of viewers watching the stream at the time of starting the data collecting process
stream_start_date	Details the date and time (in GMT) that the data collecting process started
stream_language	The language of the stream i.e., ‘en’ for English, ‘sp’ for Spanish, etc.

Table 1: *Metadata*

3 Related Work

I split my research into two categories: research findings on Twitch Chat, and the methods used in order to explore the hypotheses stated in those papers. This would give me a strong foundation of existing works and a basis of knowledge on the Machine Learning techniques used and some of their applications.

The academic research I have conducted is primarily on general findings of Twitch chat, with one paper into the usage of grammar in sentences. There is, however, many materials out there that seem useful that are not academic.

These sources are from specialists, such as the streamers themselves. There are also publishing platforms such as The Verge, who cover stories about, but not strictly, technology. These stories are sources from a stakeholder's point of view and can give more insight into the subject matter.

3.1 Existing Research into Twitch Chat

To preface, the investigation of livestream chat is a very fresh field of research; this shows when trying to search the terms 'variety' or 'stream' or 'streamer' in conjunction with each other – resulting in no results. However, much research already exists exploring other areas of chat, most employing various Machine Learning techniques to aid them.

Ringer et al. in [7] has compiled a large-scale dataset of chat text. This includes the metadata of the stream; this includes information such as the game being played on stream, duration of stream and how many viewers are present. A data cleaning process has been employed so the text is 'presentable', and they have overcome the problem of different emotes.

The bigger the audience is, the more messages will be sent in chat and therefore the quicker a message will scroll out of view. Therefore, communities on Twitch have created 'crowdspeak' [8], which may seem incoherent at first however have allowed large audiences to communicate with each other. This crowdspeak may include a large spam of emotes, or copypastas (which are large blocks of text that are spammed using copy and paste).

However, Seering et al. in [9] has, due to the banning of some users that have used spam messages, labelled all spam as undesirable behaviour. This leads to the discussion that all spam is not crowdspeak, and that a suitable model must be created in order to determine whether one is disruptive or adds to the coherence of the stream. Clearly, this conclusion is clashing with the one found in Ford et al. [8], which goes to show that due to this area of academia being so new there is the ability to argue against findings as nothing is yet set in stone.

Bulygin et al. in [10] has stated that large audiences are event driven. Meaning that rather than communicating amongst themselves, they use crowdspeak to react to on-stream events. This only applies to large audiences, as the smaller a crowd is the more space there is to chat amongst themselves. Large livestream audiences are also likened to the crowd in a sports arena, as any time a major event occurs the smaller voices fade away due to the loudness of the crowd.

While a lot of the papers mentioned above study exclusively large streams, Poyane in [11] studies a range of streams including smaller ones. Additionally, Poyane added a layer of personality to the experiment by adding information such as: gender, race and webcam. What was discovered was that larger streams have more negative communication, creating evidence for the conclusion found by Seering et al., garnering more harassment and toxicity.

While gender and webcam influenced the negativity and there is some importance to this conclusion, the dataset does not include this information. Therefore, I would need another way to measure how ‘personal’ a stream is strictly by using speech. This led me to delve into English grammar, specifically about personal pronouns [13]; this states that there are some words (‘you’, ‘I’, ‘*name of person*’) that are used when referring to a person.

In an early paper about livestreaming, Hamilton et al. in [12] has said that viewers participate for one of two reasons: being drawn to the unique content of a particular stream, and the gratification of interacting with the stream. They found in their analysis of the streams that the atmosphere created is reflective of the personality of the streamer, this means that there is some way to objectify a streamer’s qualities. This is a good foundation to start looking at which quality of a streamer attracts their audience.

3.2 Different Topic Modelling Techniques

Murakami et al. in [15] defines topic modelling as “a [unsupervised] machine learning technique that identifies ‘topics’ in a given corpus”. Topic modelling does this by detecting word and phrase patterns in the given text and clustering them under a topic.

This works very well for the project as Twitch chat has many words that are specific, either to the game or to a streamer. This means that training a supervised model (one that would need predefined lists or tags) would take too long as there are so many words to add.

Following the decision to use topic modelling, I must now decide on which method to use to analyse my chosen texts. One of the first methods used in NLP is latent semantic analysis (LSA) [16]. This method analyses the patterns of tokens in the text and assumed that words that are close in meaning will occur in similar contexts, therefore clustering said words into the same topic.

LSA overcomes the problem of synonyms; one of the earliest problems in NLP, it is when two or more words would mean the same thing – the model would weight or index these words differently when in theory they should be treated the same. However, it does not overcome the problem of polysemous words; this is when one word would have different meanings depending on the context it is used in.

This led me to search for a better alternative, and in the process, I found probabilistic latent semantic analysis (PLSA), based on LSA. Hofmann in [17] concluded that PLSA was better than LSA in most cases due to the stronger statistical foundation it has. PLSA also deals with the polysemy problem as it directly minimises word complexity. Theoretically, PLSA is a much better alternative to LSA; the data collected also show that is has much better precision when it clusters words into topics.

Another method researched is latent Dirichlet allocation (LDA) [18]. This method is better suited to my project due to its process. It is a 3-level topic model where the topic node is sampled repeatedly within the corpus. This means it goes through the corpus repeatedly, each time getting more accurate with indexing tokens into a topic. This is better for my project that LSA and PLSA because Twitch chats, while quite large, are often made of very short and quick messages; so, the 3-layer process can mitigate any inaccuracies caused due to short messages. Blei et al. in [18] also mentions how it is "highly illustrative" and how it provides "useful inferential information"; this means it will be useful for when I want to analyse my own results.

Searching for another model that performs better on shorter texts I found biterm topic model (BTM) [19]. Generally, topic models reveal topics within a text by spotting patterns in a large text, this means short texts suffer. The number of tokens in a short text play a less discriminative role, meaning the topics they are allocated in are less accurate. Limited contexts also make it harder to identify the way in which some words are used, again making it harder to cluster words. This is where BTM comes into play, a topic model made for short texts. Cheng et al. in [19] concludes it outperforms LDA on shorter text, having a higher accuracy level, while also being simpler and easier to implement.

3.3 Literature synthesis

Analysis of existing works (as of January 2021) within academia and other alternative sources of information yield very little information about the audiences in chat, the language used in chat and the streamers. This can be attributed to the fact streaming is a very new concept, with the earliest academic source being Hamilton et al.'s work [12] being published in 2014.

However, since then there has been work of interest produced. Most explore the language used in chat with the context of the volume of the audience. Works such as [8][9][11] all perform analysis of chat with the size of the audience in mind. One paper [10] tried to break from that limited scope by introducing other factors such as gender, race and whether they used a webcam – however the conclusion still weighed heavily on the volume of audience.

One academic paper found [12] explored the streamer, and discovered the atmosphere created on stream was reflective of the personality of the streamer.

In all these papers, they used NLP (topic modelling to be specific) to analyse and explore their results. There is so much work the methods used [16][17][18][19] it was not hard to find a general method to build my project around. These methods are quantitative, each topic weighs each word a value in terms of how important it is.

Using [12] and the methods founds in [16][17][18][19] I started to form my hypothesis. By looking at the atmosphere created in chat, I can use topic modelling to analyse the language used and pick out any traits. Since the atmosphere is reflective of the streamer, the streamer would have these traits. If we analyse the chat of a streamer who consistently pulls in viewers, then we can quantify the traits one would need to attract viewers. I immediately thought of variety streamers, because whatever the game or activity they do on stream, they attract audiences.

There is a need for the research, both in academia and in general interest. By going on Twitch and sorting channels by views you will see countless channels with very low audience members, this would be the general interest. The academic interest is that all guides on how to be a 'better streamer' are all qualitative with no quantitative evidence.

4 Methodology

In order to purely focus on and explore variety streamers I had to create an informal definition of what a variety streamer is as there is none in academia. This is because I decided to cull the non-variety streamer related data from the dataset, this makes it easier to see which chat logs were are form variety streamers.

Variety streamer – A streamer who streams more than one game.

Using this definition, I narrowed down the files I wanted to use in my experiments using the following constraints: a streamer whose stream has been recorded more than once, and across these streams the streamer has played two or more different games.

4.1 Choosing the ideal data

Now that it is possible to separate the data, from the data desired to the undesired data using the definition stated, it is time to discuss the ideal type of data.

The ideal streamer is one that streams lots of different games, officially stating here that it is preferred to explore streamers that stream three or more games rather than just two. This is because this gives me the chance to identify certain characteristics in the audience across many more games rather than limiting the results by only exploring two games, just to satisfy the definition. By exploring many different games, it is possible to see which topics exist across these genres – this may mean that the topic is independent from the genre and is rather influenced by the audience, from this I can infer that it is one of the attributes I am looking for.

I would also like if these games were of different genres to each other. This is because if games are in the same genre there will clearly be a constant use of genre specific language observed, this is not wanted because we are trying to identify attributes of personality and environment and not the genre of the game being played, which is all the genre specific language would convey. However, if language from different game genres is explored then the genre specific language used will not be common across the results, therefore putting less importance on them.

4.2 Choosing LDA

I have chosen LDA as my model. This is because as Blei et al. [18] mentioned, it produces results just as good as PLSA and LSA, but due to its 3-layer process it is more flexible. Twitch chat varies in length, this means the corpus that the model will use varies – focussing on the importance of flexibility. Furthermore, there are a plethora of tutorials online that support LDA that I can use as resources to better my project.

I have considered using PLSA as according to Blei et al. [18], it is easier to implement. I have also considered using BTM but there are not many tutorials online to support it.

4.3 Choosing Python and GenSim

I have chosen Python [20] as the language to write my code in. This is simply because it has a plethora of libraries, a collection of functions that eliminates the need to code from scratch, that support machine learning. It also has libraries that make visualisation of code and results produced extremely readable, making it easy for me to infer the data. This is good for my project as it allows me to easily infer my results and come to conclusions.

The two most prominent libraries for LDA are GenSim [21] and scikit-learn [22]. You cannot compare the two, as both models achieve the same thing for the scope of my project – so one is not objectively better than the other in that regard. However, I have chosen GenSim as it seems more specialised for my task as it is purely used for topic modelling and NLP, whereas scikit-learn feature various data analysis tools.

4.4 Applying LDA

This section goes through the process of how LDA is being applied to a chat of my selection.

First, the data must be pre-processed. This is done in two steps: removing unnecessary columns, and removing stop words (in NLP, refers to common words that do not add any meaning to a sentence [27]).

I removed the time column as my project only focuses on the language used and not the frequency of words and therefore there is no need for the time column. The removal of stop words allows the model to focus on special words used, it also mitigates inaccuracies as the model would not take these stop words into account.

```
["i", "me", "my", "myself", "we", "our", "ours", "ourselves", "you", "your", "yours", "yourself", "yourselves", "he", "him", "his",
"himself", "she", "her", "hers", "herself", "it", "its", "itself", "they", "them", "their", "theirs", "themselves", "what", "which",
"who", "whom", "this", "that", "these", "those", "am", "is", "are", "was", "were", "be", "been", "being", "have", "has", "had",
"having", "do", "does", "did", "doing", "a", "an", "the", "and", "but", "if", "or", "because", "as", "until", "while", "of", "at", "by",
"for", "with", "about", "against", "between", "into", "through", "during", "before", "after", "above", "below", "to", "from", "up",
"down", "in", "out", "on", "off", "over", "under", "again", "further", "then", "once", "here", "there", "when", "where", "why", "how",
"all", "any", "both", "each", "few", "more", "most", "other", "some", "such", "no", "nor", "not", "only", "own", "same", "so", "than",
"too", "very", "s", "t", "can", "will", "just", "don", "should", "now"]
```

Figure 4: *The list of stop words used in my LDA model*

The next step is to code the model, which will be included in the submission, run the program and let it produce the set of results. Figure 5 shows the results produced after applying the LDA model to our data, using pyLDAvis [24].

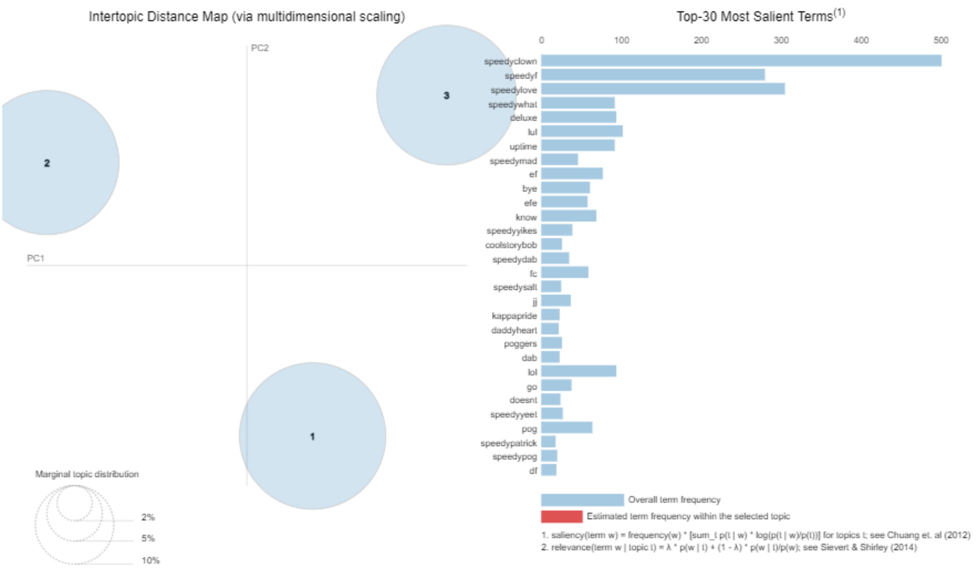


Figure 5: *Default result produced using pyLDAvis*

Using a library called pyLDAvis one can produce results into a convenient, human-readable format that can be intuitively understood. Figure 5 is what the results would look like without interacting with it, a ‘default’ result. From this output you can identify the ‘Top-30 Most Salient (a fancy word for important) Terms’, which can provide a brief insight into what the chat as whole contains without having to analyse the topics.

In addition to this, there is an ‘Intertopic Distance Map’ that visualises how different each topic is to each other. The further away a topic is, the more different they are, so if these topics overlap then that means they are similar. The size of each area represents how prevalent that topic is in the chat.

4.5 Analysis of topics

In order to explore the topics produced one must interact with the default result. By simply clicking on the blue circle labelled ‘1’ the panel on the right side of the result will display the ‘Top-30 Most Relevant Terms for Topic 1’, as Figure 6 shows, and since I have specified in my model to produce only three topics, I will do this three times for every chat I apply my model to.

The main features of the results shown in Figure 6 are: the red bar, the blue bar and the percentage number of tokens [24], which is shown in brackets at the end of the title of the panel. The blue bar represents overall term frequency, this is how often this term appears in chat. The red bar represents estimated term frequency within the topic, this is an estimation of the number of times this term appears in the topic; the higher the frequency is, the higher the importance put on that term within the topic. Finally, the percentage number of tokens displays the percentage of tokens that this topic covers. In Figure 6, the LDA model has grouped 34.7% of all tokens in chat into this topic. Linking this to my hypothesis, if this topic were inferred to describe a ‘comedic’ trait then I would say that the chat was 34.7% comedy.

In my experiment I will be focussing on the terms that appear in to be most prevalent in each topic; each topic orders terms by order of frequency, so the most prevalent terms will appear higher up the list. I will also explore the percentage of tokens the topic covers.

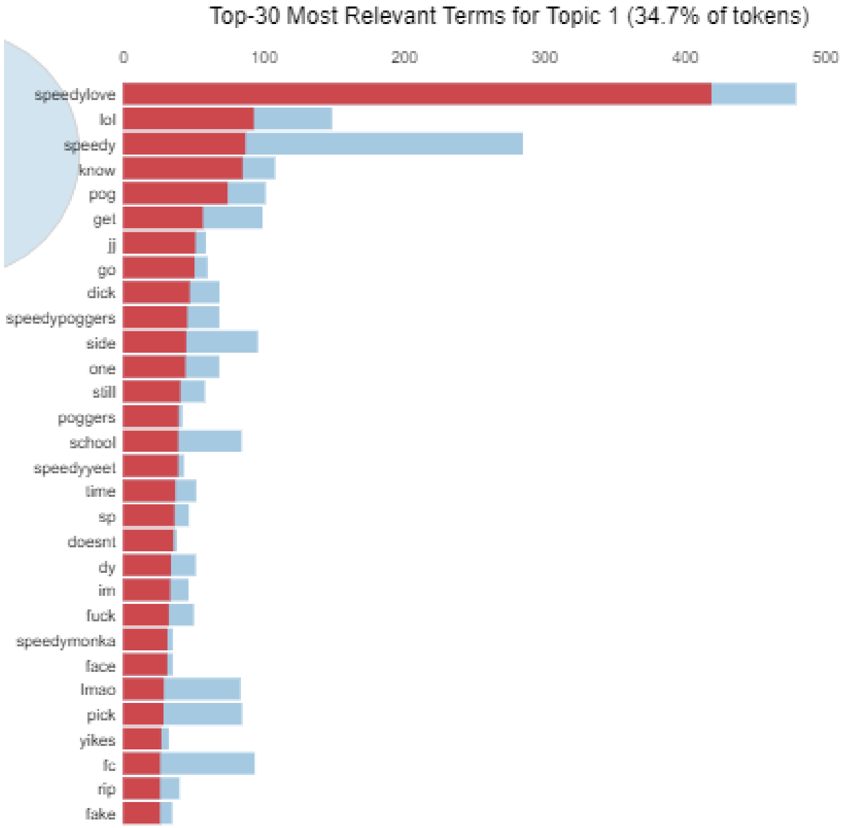


Figure 6: *Example of final results to analyse*

4.6 Evaluation

This section discusses the problems faced and decisions made, ensuring that all decisions have been justified appropriately. In addition to this, any considerations made to improve the work shall be discussed.

The main problem was found when exploring 4 or more models. As you can see in Figure 7, the intertopic distance map shows topics 2 and 4 overlapping. These two topics overlapping mean that the topics are similar, therefore the topics are not unique as they share some terms. This is unfavourable as it may mean that these topics may represent the same trait. Hence the decision to cap the number of topics to just three.

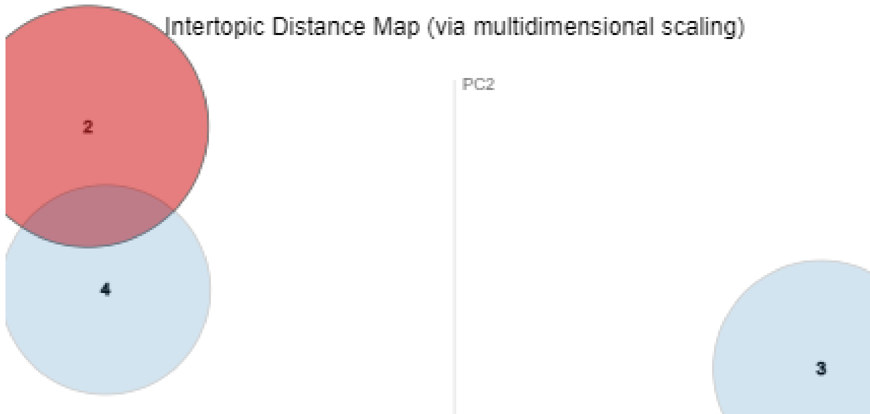


Figure 7: *Image showing overlap when using 4 topics*

Furthermore, the LDA model is only being applied to one chat at a time. Chats are relatively small, therefore only containing a small number of relevant terms. By applying the model specified to create many topics on a small chat, many unimportant topics were created that are filled with useless terms. As shown in Figure 8, where Topic 3 is filled with nonsensical terms such as ‘ef’, ‘jj’, ‘cfa’, etc. Adding more reason to cap the number of topics created to a small amount.

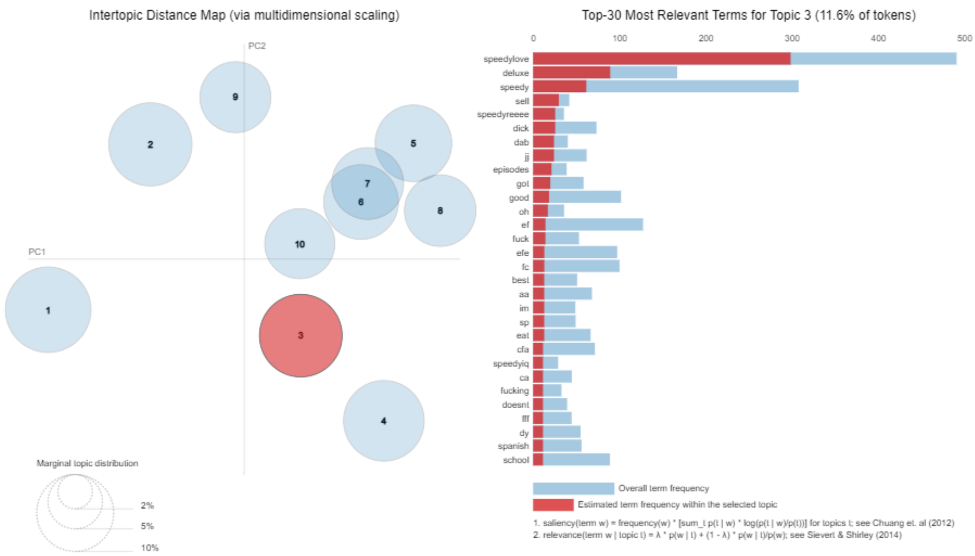


Figure 8: *Image showing Topic 3 filled with many useless terms*

One solution considered was to remove all terms with three or less characters. This solution however was quickly disregarded as potentially important terms such as ‘ty’ and ‘np’ could potentially be found in a topic. These terms are abbreviations; ‘ty’ translates to ‘thank you’ and ‘np’ translates to ‘no problem’. These are important terms as they facilitate the trait of friendliness, which is an important trait to discover as it can describe a streamer’s chat as friendly and welcoming.

There have also been considerations to implement lemmatisation and stemming as it has been done in many similar works [7][10][11]. Lemmatisation [28] is the process of grouping inflected forms of a word into a single item. Table 2 shows the terms ‘lool’, ‘loooool’ and ‘lul’ will be grouped into the single item ‘lol’. However, this has been disregarded due to the way the results are represented. Since one must estimate the value of the red and blue bars (as no value is given), it would be more sensible measure the number of inflected forms over the top-30 terms. For example, if the inflected forms in table 2 were found in the top-30 terms, then we can calculate 10% (3/30 as a percentage) of the topic would be “lol”. This statistic just seems more tangible than estimating the value by sight.

lol	lool
	loooool
	lul

Table 2: *Example of lemmatisation on the term "lol"*

Stemming [28] is the process of reducing inflected words to its root word. However, this been disregarded as it will conflict with emotes. Table 3 shows that streamer specific emotes will be reduced to just the streamer name by using the streamer xQc and his emotes as an example. Before stemming, each emote would be categorised into different topics as they are used in different context, but if they were to be stemmed into ‘xQc’, the model would display that as a single token. This will interfere with the quality of topics being created.

xQc	xqcRage
	xqcSad
	xqcPog

Table 3: *Example of stemming on xQc's emotes*

Finally, there is a slight professional issue. The dataset anonymises the streamer, however by analysing the chat one can infer their identity. By close inspection of the language used, I explored emotes used. These emotes include streamer specific emotes. By analysing which streamer specific emotes are used, and the volume at which they are used, one can infer the streamer.

This is not an ethical issue as the streamer would be operating as normal, knowing that everything is being recorded either by Twitch logs, VODs, clips etc. It does cause a slight professional issue however, as it makes the anonymity process just slightly redundant.

5 Results

In this section I will be showcasing the results produced from applying the LDA model to my chosen Twitch chats. To satisfy the criteria set in Section 4.1, I have chosen one variety streamer that has streamed three games with each game coming from its own genre. This was possible due to the metadata; one can find out which games a streamer plays via the game ID and thus easily finding out which genre the game belongs to.

Call of Duty: Black Ops 4 (BO4) is a first-person shooter (FPS), in this genre the player sees the game through the eyes of their player character with a weapon (normally a gun) at the centre of the screen. World of Warcraft (WoW) is an example of a massively multiplayer online role-playing game (MMORPG), in this genre the player plays the role of a character, generally, in a fantasy or science-fiction world where they can interact with other players. Dead by Daylight (DbD) is an example of a survival horror game, where the player must ensure their character's survival while the game attempts to scare the player.

To preface, all contexts given in terms of how an emote is all personal, my experience as an avid Twitch user has given me confidence to impose my own perspective on the results. Additionally, images of any emotes mentioned can be found in Appendix C.

5.1 Results from Call of Duty: Black Ops 4 stream

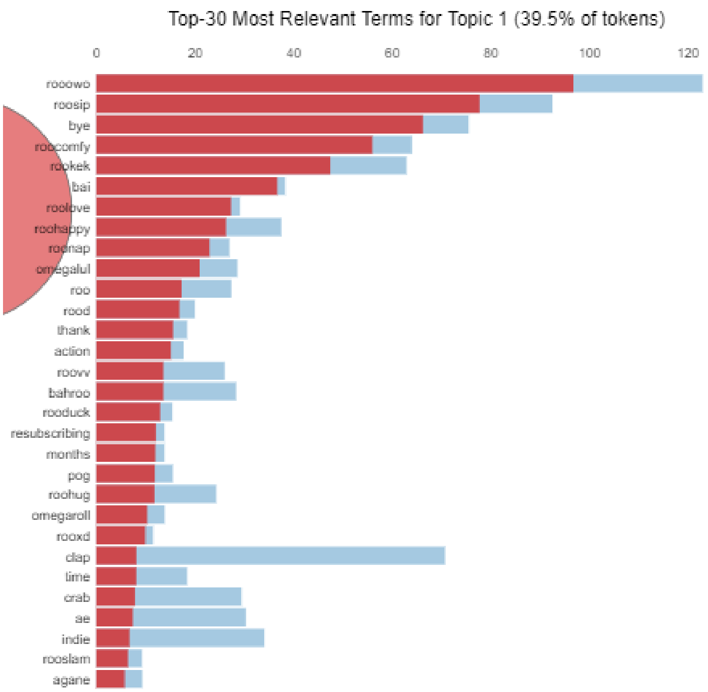


Figure 9: *BO4's Top-30 Most Relevant Terms for Topic 1*

Figure 9 shows the most frequent terms in Topic 1, which covers 39.5% of all tokens, are all streamer specific emotes from the streamer AdmiralBahroo [30]. Every term containing 'roo' is some form of the streamer's name, this accounts for 16 of the 30 terms (53%) showcased in the results. In order to use streamer specific emotes, one would have to spend buy access by subscribing, which in turn helps the streamer as it is one of their forms of income. Therefore, we can label Topic 1 as 'supportive' as the audience is providing the streamer financial support.

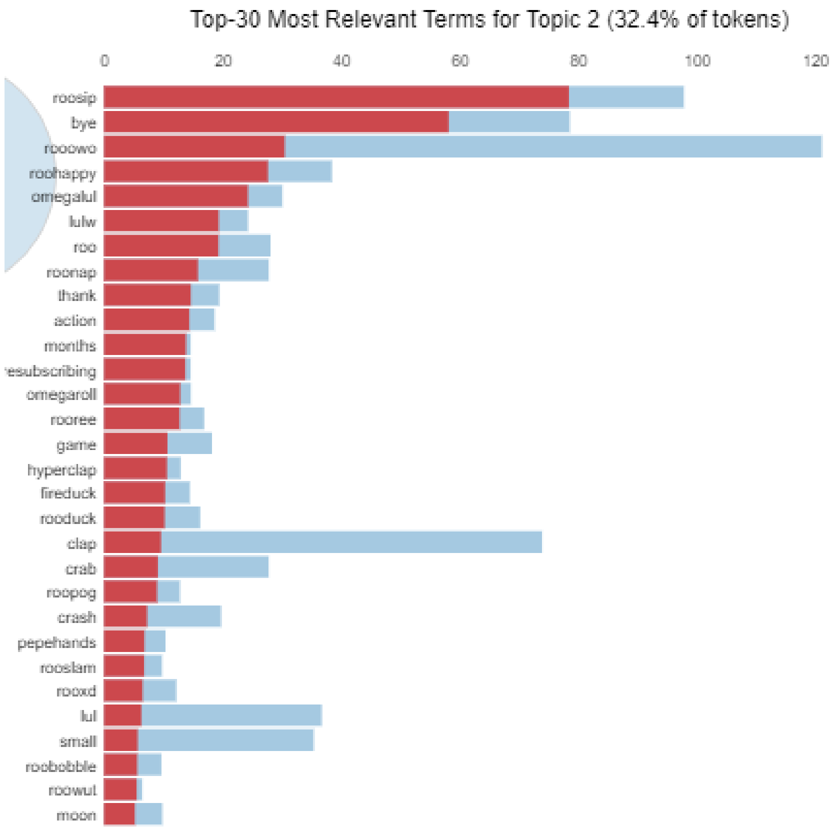


Figure 10: *BO4's Top-30 Most Relevant Terms for Topic 2*

In Topic 2, displayed using Figure 10, the most prevalent group of terms are: 'thank', 'months', 'resubscribing', 'pog', 'clap', 'roohappy', 'action'. These are all terms used by a bot in chat to celebrate when a viewer resubscribes. The term 'resubscribing' signifies a repeated action, and as mentioned before subscribing is a supportive action. This continued support denotes a sense of loyalty, hence why Topic 2 will be labelled as 'loyalty'.

Topic 3, shown using Figure 11, is labelled as 'conversation'. This is because I focussed on the terms 'small', 'indie', 'company' which are very high on the list of relevant terms, meaning that these terms are important. Since these terms were also grouped closely together, we can assume that they were used in conjunction with each other. Clearly there was a discussion in chat about small indie companies, since BO4 is made by Treyarch, whose father company is Activision and from this it is possible to infer this discussion is not game related.

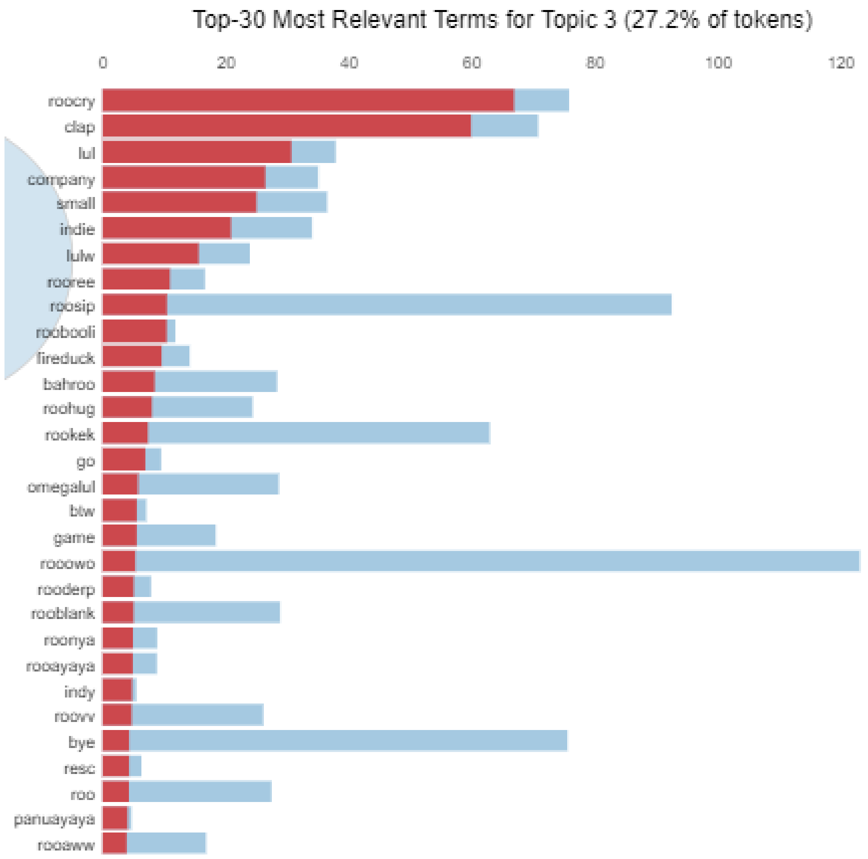


Figure 11: *BO4's Top-30 Most Relevant Terms for Topic 3*

5.2 Results from Dead by Daylight stream

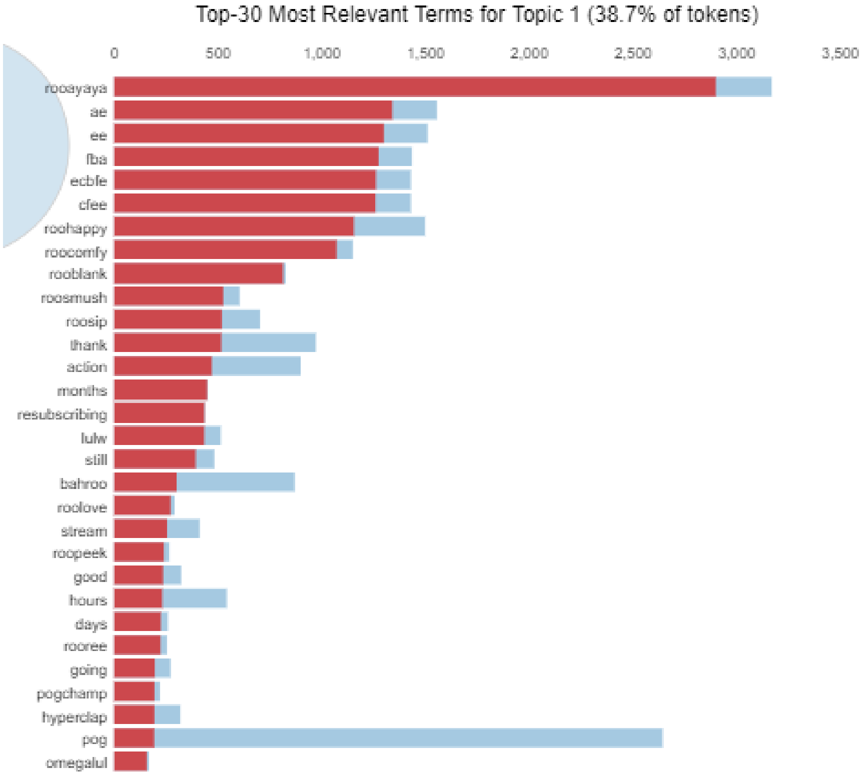


Figure 12: *DbD's Top-30 Most Relevant Terms for Topic 1*

Due to the results displayed in Figure 12, Topic 1 of the DbD stream is labelled as 'supportive'. This is due to the same reasons found by analysing Figure 9. In short, it is due to the high frequency of 27% (8/30 in percentage form) and relevancy of the streamer specific emotes.

Results shown in Figure 13 show a lot of the 'loyalty' terms having high relevancy. Therefore, my reasons for labelling Topic 2 of the DbD stream are similar to my reasons explained during my analysis of Figure 10.

Figure 14 introduces a new trait. Due to the high frequency of the term 'pog', and relating terms such as 'monkas', 'roopog'. These terms all have the connotations of 'excitement'. From Figure 14, you can notice that the red and blue bars are well above the rest, seeming to be almost double the next most relevant term.

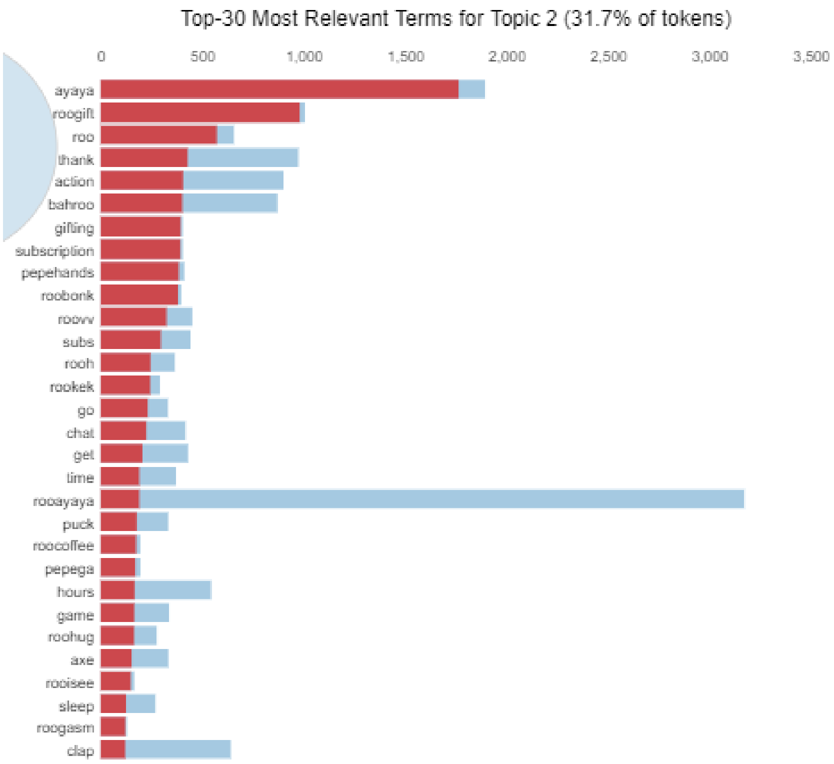


Figure 13: *DbD's Top-30 Most Relevant Terms for Topic 2*

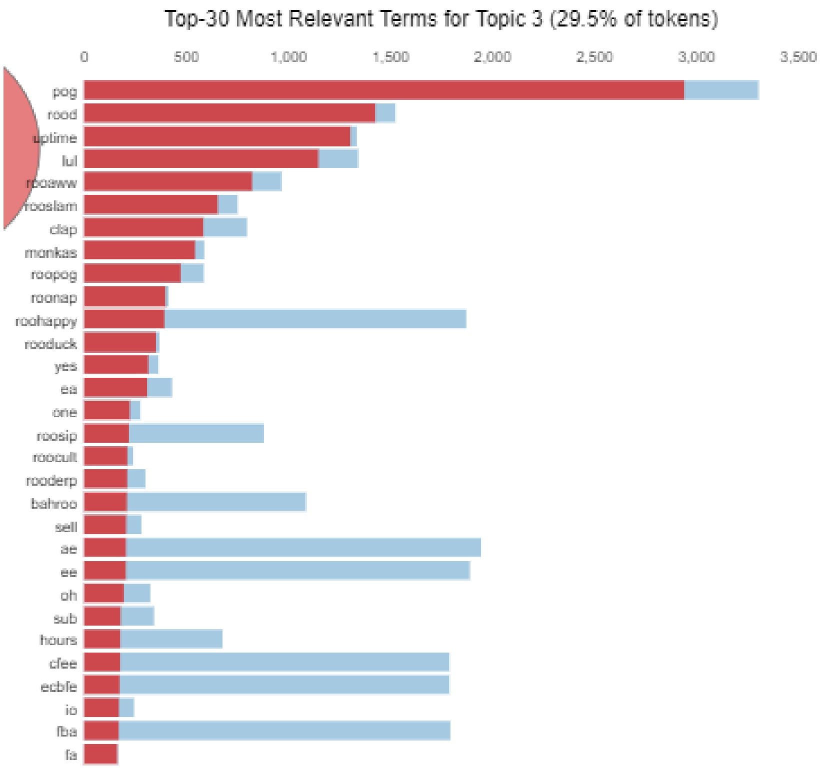


Figure 14: *DbD's Top-30 Most Relevant Terms for Topic 3*

5.3 Results from World of Warcraft stream

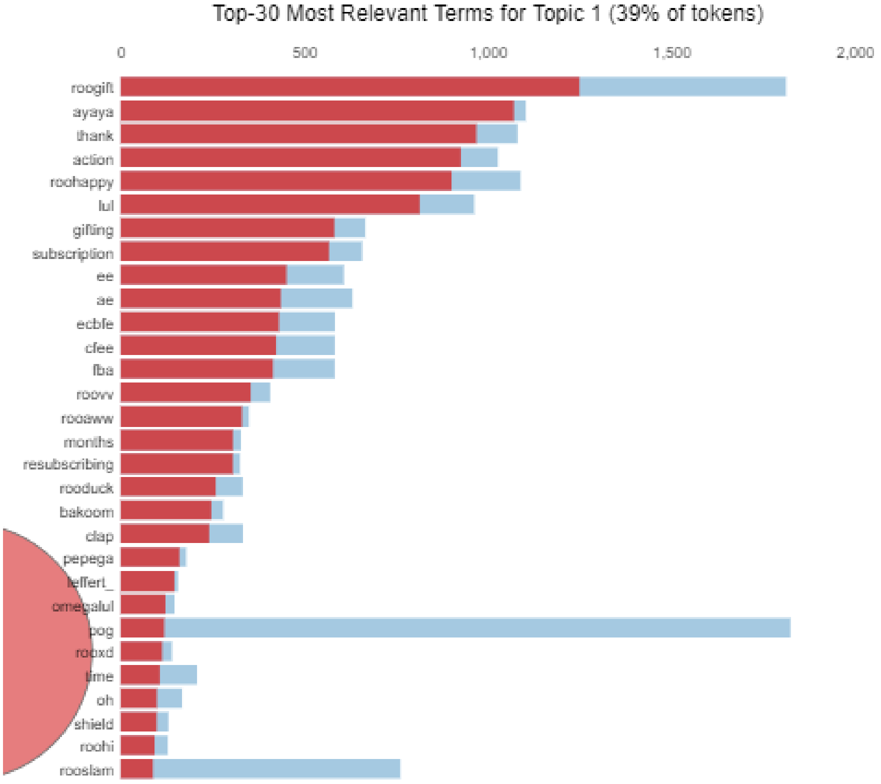


Figure 15: *WoW's Top-30 Most Relevant Terms for Topic 1*

Figure 15 shows a break in trend. The most significant group of terms are: ‘thank’, ‘action’, ‘gifting’, ‘subscription’, etc. This is the group that describes the topic of ‘loyalty’. In the case of the previous two streams analysed, we have noticed that the results for Topic 1 described the ‘supportive’ trait. However, encompassing 39% of all tokens in chat, the WoW stream’s Topic 1 is labelled as ‘loyalty’.

The WoW stream’s Topic 2 has ‘pog’ as its most significant term, as shown by Figure 16. In addition to this, the term ‘pog’ is accompanied by ‘pogchamp’, which is a variant of ‘pog’. This adds evidence that Topic 2 should be labelled as ‘exciting’. However, this topic also contains the terms ‘pepehands’, ‘jebaited’, and ‘roocry’ which suggests something unfortunate has happened on stream. Therefore, you could argue this topic should be labelled ‘sad’. But due to the high frequency of ‘pog’ and its variations completely overshadowing the other terms, I shall label this topic as ‘exciting’.

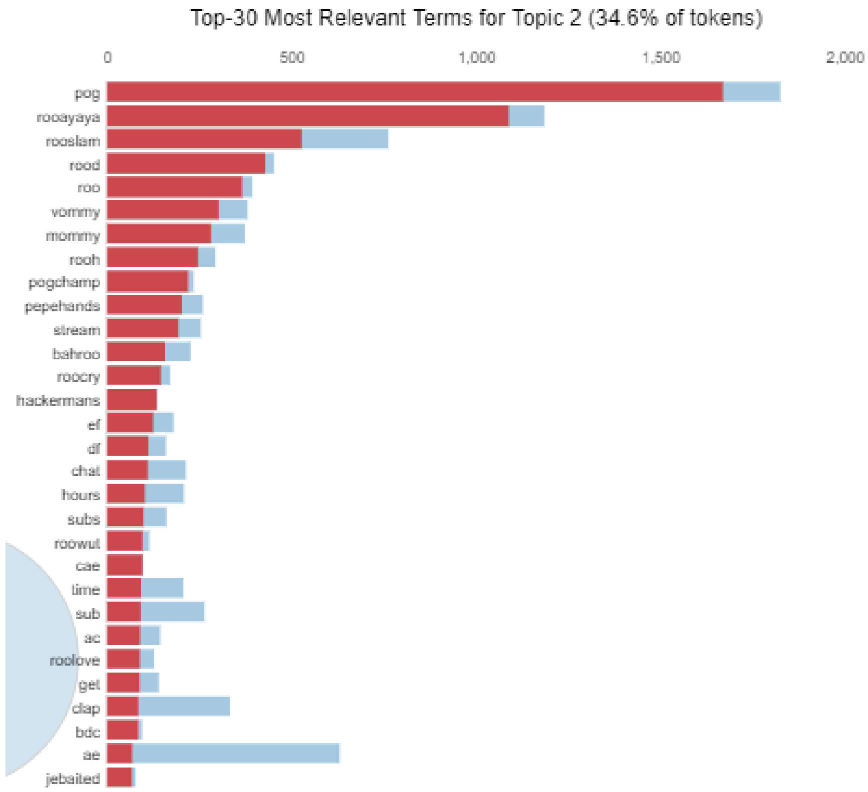


Figure 16: *WoW's Top-30 Most Relevant Terms for Topic 2*

For the final topic explored, shown in Figure 17, there is a high frequency of streamer specific emotes. Therefore, just like when analysing BO4's Topic 1 and DbD's Topic 1, this topic shall be labelled 'supportive'. What is interesting to note is that while the 'supportive' topic encompassed 39.5% and 38.7% in the BO4 and DbD streams respectively, it comes last at topic 3 at a low percentage of 26.4% - the lowest percentage number of tokens recorded.

One reason could be due to that unfortunate event mentioned earlier, denoted by the group of tokens that would have labelled Topic 2 as 'sad'. It would make sense that an unfortunate event would overshadow anything exciting.

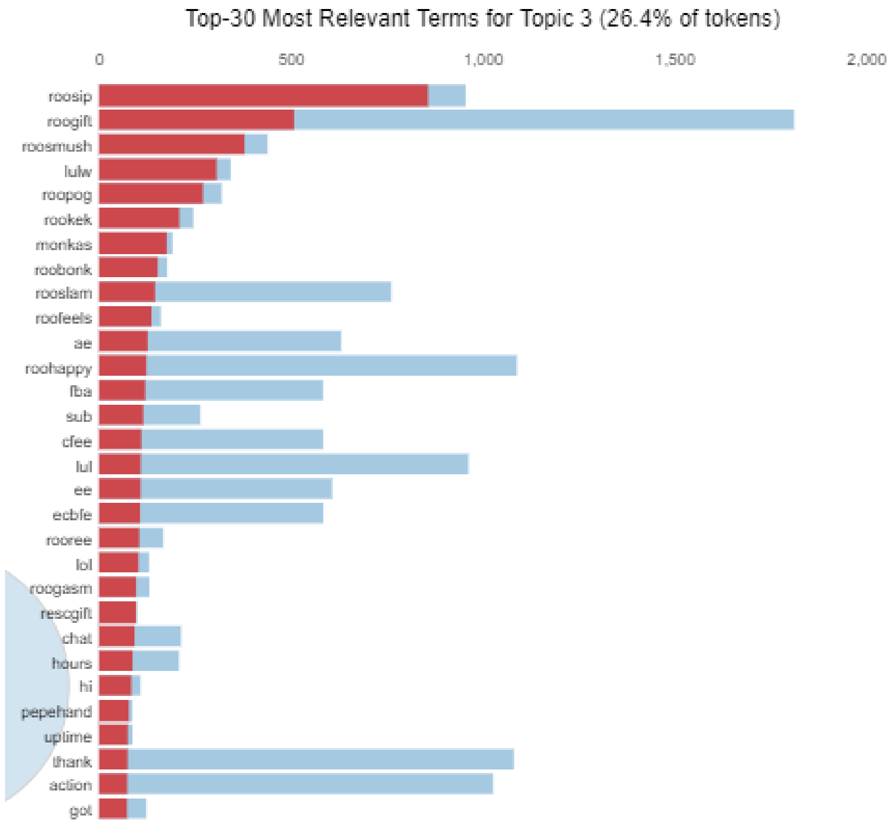


Figure 17: *WoW's Top-30 Most Relevant Terms for Topic 3*

5.4 Discussion

		Games		
		BO4	DbD	WoW
Topics/ Traits	Supportive	39.5%	38.7%	26.4%
	Loyalty	32.4%	31.7%	39%
	Conversation	27.2%		
	Exciting		29.5%	34.6%

Table 4: *Table showing percentage number of token statistics of all topics analysed*

Table 4 displays the results explored in Sections 5.1-5.3. By analysing which terms make up a topic, it was possible to identify the most prevalent traits within said topic, and thus label it. I observed that there were four significant traits which I labelled: 'supportive', 'loyalty', 'conversation', 'exciting'. Each trait is made up tokens, these tokens appear in the chat a certain number of times, and this number has been transformed into a percentage that represents the percentage number of tokens. For example, the 'supportive' trait makes up 39.5% of the BO4 stream chat.

I will use the following equation to calculate an average:

$$n\% = \frac{Sum\ of\ trait's\ percentage\ number\ of\ tokens}{Number\ of\ streams\ analysed}$$

Where 'n' represents the percentage that trait will appear in the chat.

		Average % appearance
Topics/ Traits	Supportive	34.9%
	Loyalty	34.4%
	Conversation	9.1%
	Exciting	21.4%

Table 5: *Table showing average percentage appearance of terms in a topic*

Table 5 clearly represents and synthesises my results. It contains the traits I labelled the topics with, and a respective percentage that represents how often a token of that trait will appear in the chat.

It is discovered that the most prevalent trait across all chats is 'supportive' at 34.9% average. Linking back to Hamilton et al. [12], if the chat is discovered to be supportive, then so is the streamer. If both audience and streamer support one another, it creates a very friendly and welcoming atmosphere that invites people to join. 'Loyalty' is the next most prevalent trait at 34.4%. If such an inviting atmosphere is created, loyalty is a consequence of that because it keeps inviting people back to interact with the stream. The next trait at 21.4% is 'exciting'. Of course, if a stream is exciting, it is entertaining and therefore why it is one of the traits that causes audiences to return. Finally, at 9.1% is the 'conversation' trait. Due to the fact it only appeared once across all the streams, one may argue that it is just an anomaly. According to the metadata, the streamer's audience is relatively large therefore there should be no room for coherent conversation to occur. However, it was prevalent enough to be chosen and it occurs at 9.1%. While it may not happen often, there is room for it to occur and with how friendly and inviting the chat environment is due to how supportive the other users are, you cannot completely disregard it.

6 Conclusion and Future Work

In conclusion, the utilisation of modern NLP has allowed for me to achieve my hypothesis. It has also helped me identify four traits that can cause audiences to keep coming back to a certain stream.

My results are broad reaching as it covers a wide scope of subjects because from the results discussed in sections 5.1-5.3, we have managed to label each topic a unique trait. In addition to this, we also managed to quantify these traits by analysing how many times a trait would appear in the chat. This covers quantitative and qualitative analysis, two very different topics.

We covered quantitative analysis when explaining the reasons why these traits are important and how one can argue it causes streamers to return to a stream. And secondly, we covered qualitative analysis as we managed to identify the average appearance of a trait in Table 5. In addition to this, the red and blue bars in sections 5.1-5.3 were measured by frequency – it gives a value.

My results are also multi-faceted. There are several ways you can perceive the qualitative data of the traits. Explained in section 5.4 for example, the trait ‘conversation’ – one could argue it is an anomaly that should be ignored, but another could argue it is still valid data as it is consequence to the other traits found, namely ‘supportive’.

By successfully quantifying the traits found, the possibility of sorting Twitch chat by one of these traits becomes possible. If a viewer wanted to join a new community, they could sort each streamer by ‘supportive’ and this task will be easier. As we discussed, a ‘supportive’ chat is one that is friendly and welcoming – this is exactly what that viewer would want.

Furthermore, by identifying these four traits we can start to build a quantitative case study of how to gather viewers to your stream. This will teach new streamers what traits are important. Additionally, if Twitch allows existing streamers to see their own statistics, these existing streamers would be able to see what they are lacking in and therefore how to improve.

In terms of future work, this is the first paper that focusses on variety streamers in academia. Hopefully, it will be possible that others use this paper as a good starting point on that topic. Using the methods stated in this paper, one could attempt to find different traits, possibly one that drives away their audience. In any case, further works should consider what was discussed in section 4.6 to improve my method and possibly produce better results.

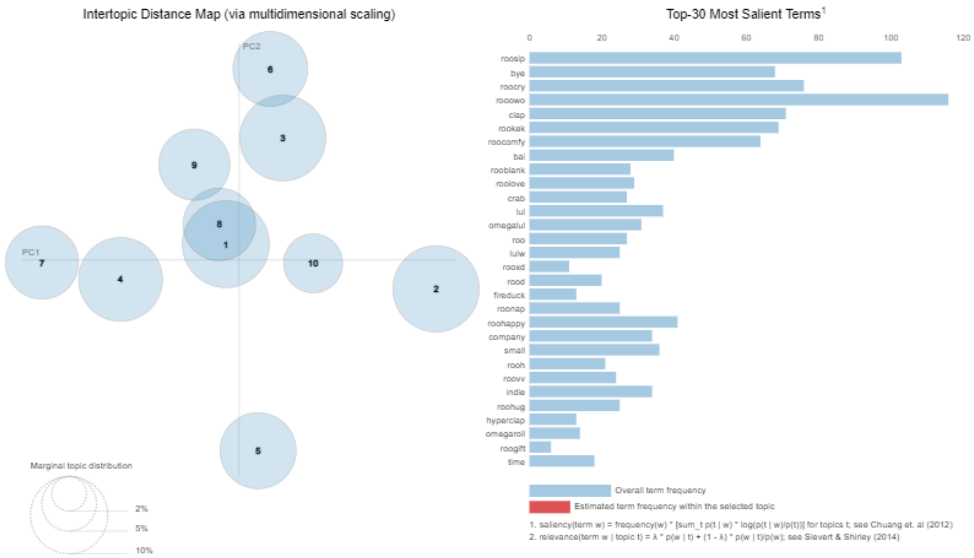
Overall, this project provides the first insight into variety streamers, and to become a fun and interesting entry point into this topic in the industry I will certainly be able to fill a gap in knowledge, which stated in section 1.1 was my driving motivation. As they say on Twitch…

First!

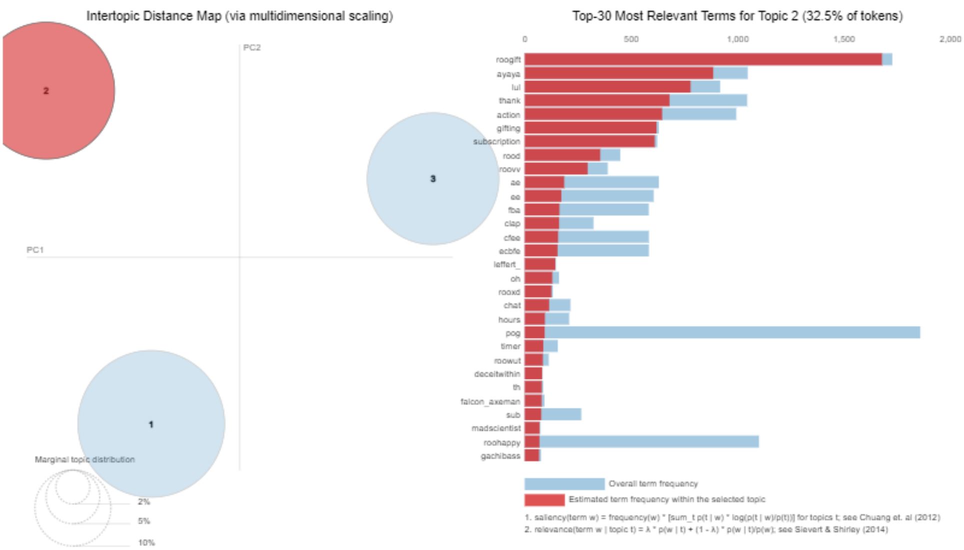
Appendix A – Default LDA results

The purpose of this appendix is to show the default results. These results were not used in the experiment but may provide you with more insight to the data.

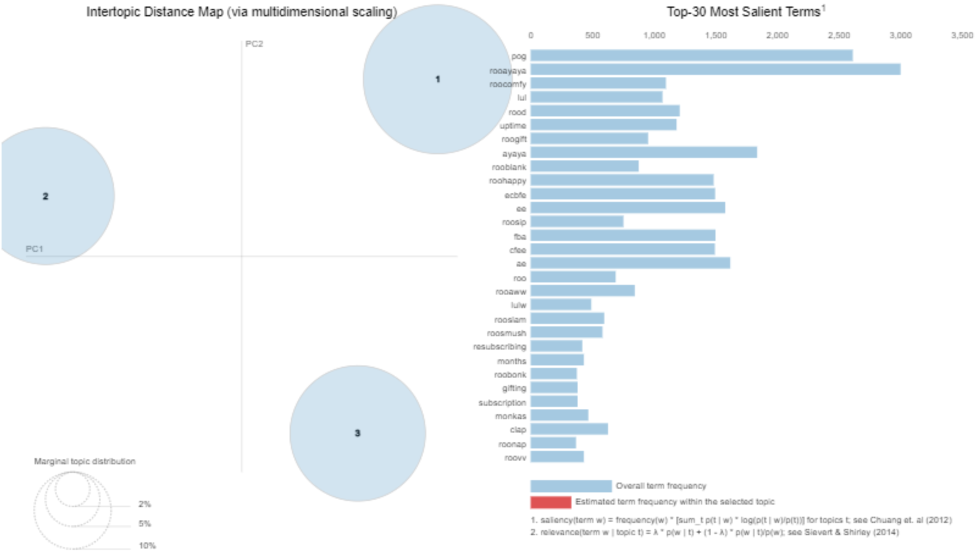
Default result for Call of Duty: Black Ops 4 stream:



Default result for Dead by Daylight stream:



Default result for World of Warcraft stream:



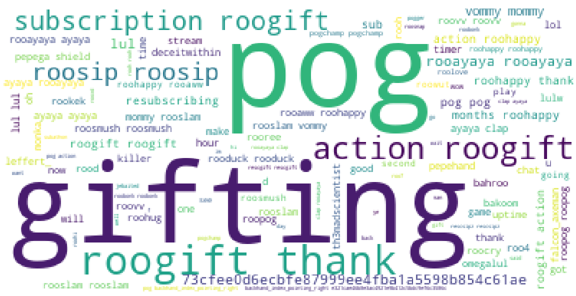
Appendix B – Word clouds

This appendix showcases the word clouds associated to the chats mentioned in Section 5. This was not used in the experiment but could be used as brief exploratory analysis, to get a quick visualisation of the whole chat. It was key in ensuring that I was on the right track in terms of data pre-processing, as any unwanted terms would be clearly visible.

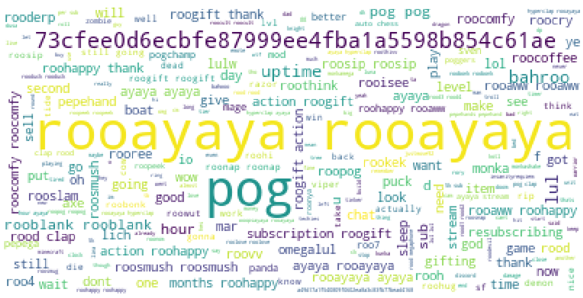
The word cloud was also used as a means of predicting which topics would be most prevalent, so I could quickly choose the 'best' chats to analyse without having to go through the whole process of applying LDA.

<u>Word cloud for Call of Duty: Black Ops 4 stream:</u>

<u>Word cloud for Dead by Daylight stream:</u>



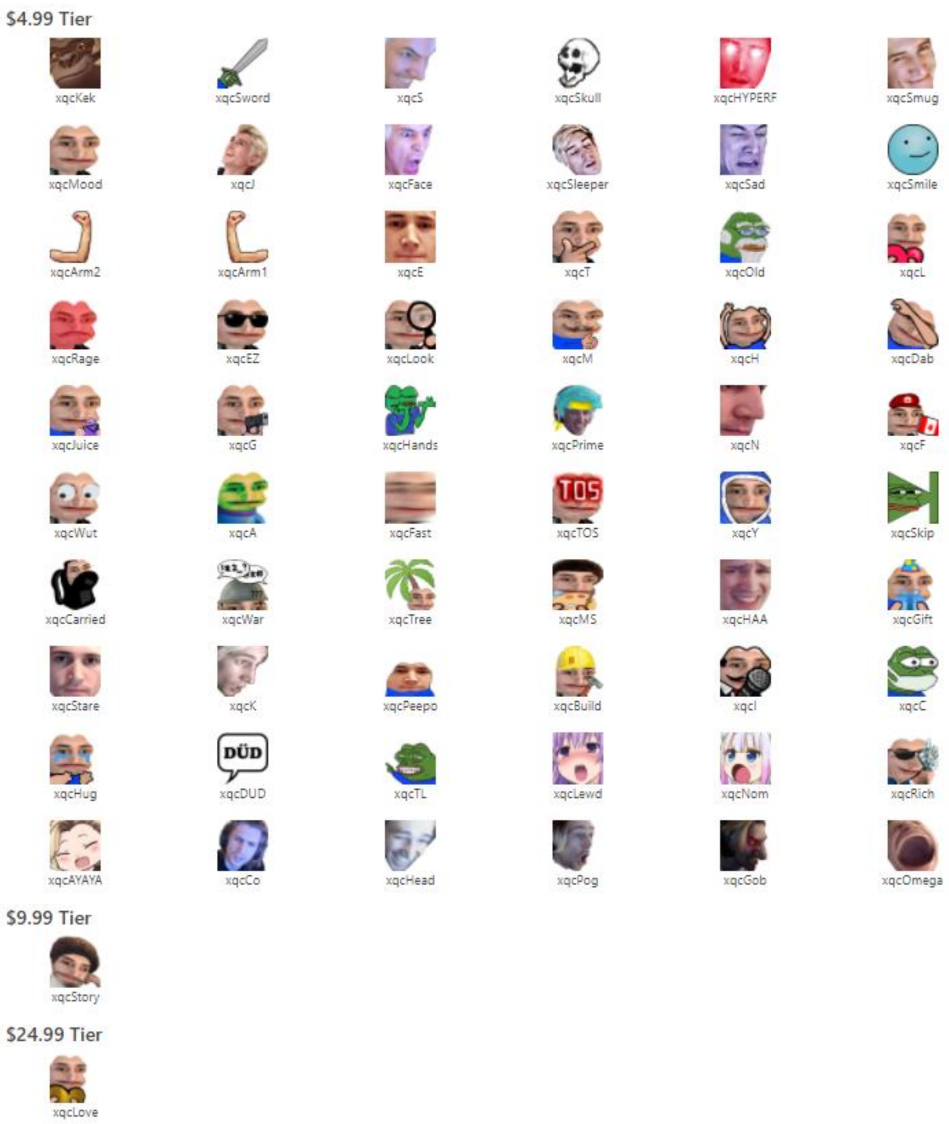
<u>Word cloud for World of Warcraft stream:</u>



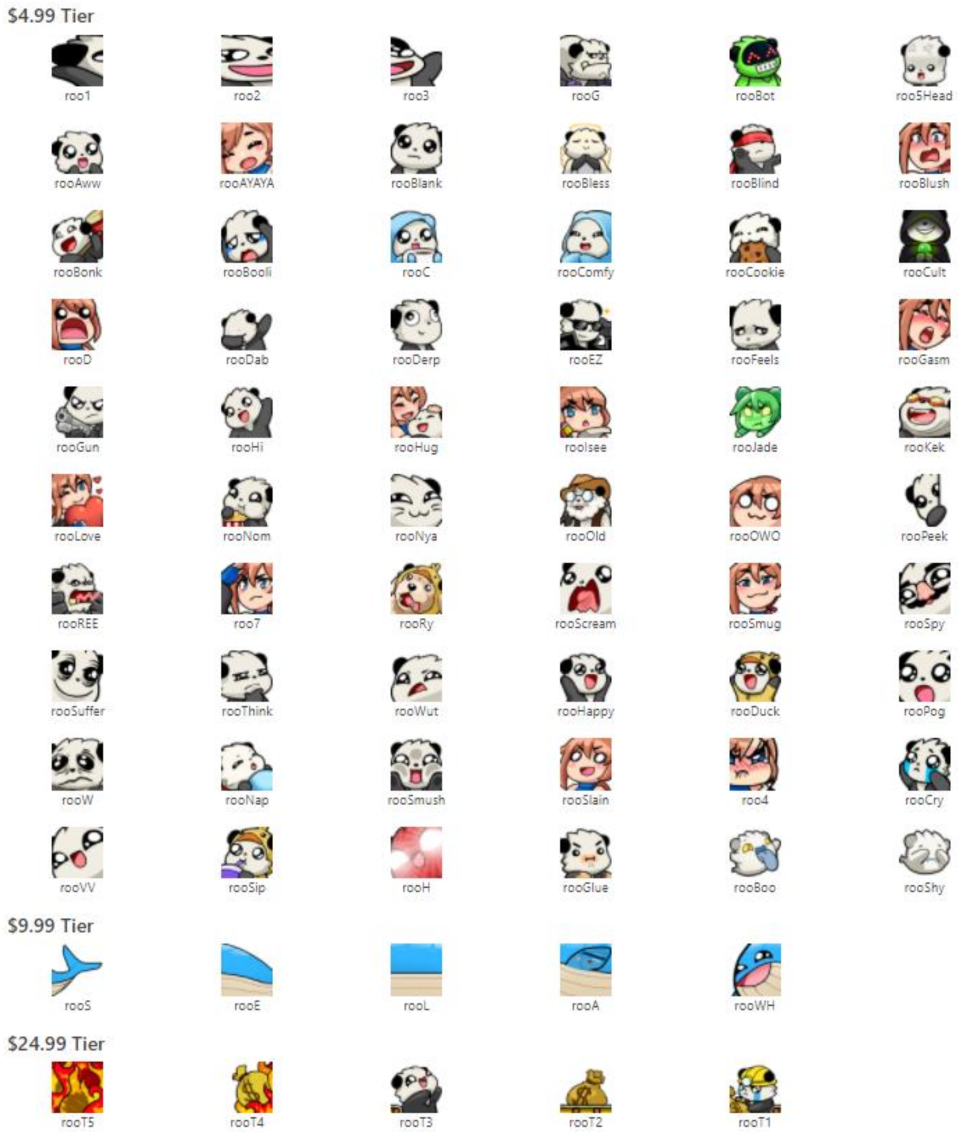
Appendix C – Emotes

Appendix C is to give more context to the emotes mentioned in this report. While I have given examples of the way the emotes mentioned are used across this report, it would be useful to be able to visualise the emotes as a convenience.

<u>xQc emotes:</u>



Admiralbahroo emotes:



'pog' emote:

Here you can clearly see this emote is used for when something exciting happens on stream.

'monkaS' emote:

An emote used in times of nervousness, sometimes used when waiting in suspense for some event to occur.

Bibliography

[1] YouTube. "YouTube," *YouTube.com.* [Online] Available: https://www.youtube.com/ [Accessed January 2021]

[2] Twitch. "Twitch.TV," *Twitch.tv.* [Online] Available: https://www.twitch.tv/ [Accessed January 2021]

[3] W. C. Partin. "Bit by (Twitch) Bit: "Platform Capture" and the Evolution of Digital Platforms," *Social Media + Society. SI: Platforms and Cultural Productions,* September 2020

[4] K. Blashki, S. Nichol. "Gamer geek's goss: linguistic creativity in young males within an online university forum (94/\3 933k'5 9055oneone)," *Australian journal of emerging technologies and society,* vol. 3, issue 2, 2005, pg. 71-80

[5] Twitch Emotes. "Twitch Emotes," *Twitch Emotes.* [Online] Available: https://twitchemotes.com/ [Accessed January 2021]

[6] H. Nakamura. "GMHikaru," *Twitch.tv.* [Online] Available: https://www.twitch.tv/gmhikaru [Accessed January 2021]

[7] C. Ringer, M. A. Nicolaou, J. A. Walker. "TwitchChat, A Dataset for Exploring Livestream Chat," *University of York and The Cyprus Institute*, 2020

[8] C. Ford, A.M. Tsaasan, D. Gardner, B. Nardi, L.E. Horgan, J. Rickman, C. Liu. "Chat Speed OP: Practices of Coherence in Massive Twitch Chat," *CHI 2017: Augmented Bodies and Interactions,* May 2017, pg. 858-871

[9] J. Seering, R. E. Kraut, L. Dabbish. "Shaping Pro and Anti-Social Behaviour on Twitch Through Moderation and Example-Setting," *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing.* ACM, 111-125

[10] D. Bulygin, Ilya Musabirov, A. Suvorova, K. Konstantinova, P. Okopnyi. "Between an Arena and a Sports Bar: Online Chats of eSports Spectators," *Human-Computer Interaction arXiv: 1801.02862v3.* 2020

[11] R. Poyane. "Toxic communication during streams on Twitch.tv. The case of Dota 2," *Mindtrek '18: Proceedings of the 22nd International Academic Mindtrek Conference,* October 2018, pg. 262-265

[12] W. A. Hamilton, O. Garretson, A. Kerne. “Streaming on Twitch: Fostering Participatory Communities of Play within Live Mixed Media,” *Proceedings of the 32nd annual ACM conference on Human factors in computing systems.* ACM, 1315-1324

[13] J. W. M. Verhaar. “Toward a reference grammar of Tok Pisin: an experiment in corpus linguistics,” *Honolulu: University of Hawaii Press.* 1995

[14] The Verge. “The Verge,” TheVerge.com. [Online] Available: https://www.theverge.com/ [Accessed February 2021]

[15] A. Murakami, P. Thompson, S. Hunston, D. Vajn. “‘What is this corpus about?’: using topic modelling to explore a specialised corpus,” *Corpora,* 2017, vol. 12, pg. 243-277

[16] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, R. Harshman. “Indexing by Latent Semantic Analysis,” *Journal of The American Society for Information Science,* 1990, vol.41, pg. 391-407

[17] T. Hofmann. “Probabilistic Latent Semantic Indexing,” *ACM SIGIR Forum,* vol. 51, no. 2, 2017

[18] D. M. Blei, A. Y. Ng, M. I. Jordan. “Latent Dirichlet Allocation,” *Journal of Machine Learning Research 3,* pg. 993-1022, 2003

[19] X. Cheng, X. Yan, Y. Lan, J. Guo. “BTM: Topic Modeling over Short Texts,” *IEEE Transactions on Knowledge and Data Engineering,* vol. 26, no. 12, 2014

[20] Python Software Foundation. “Python,” *python.org,* version 3.9, 2020 [Online] Available: https://www.python.org/downloads/ [Accessed April 2021]

[21] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research,* vol. 12, pg. 2825-2830, 2011

[22] R. Rehurek, P. Sojka. “Gensim – python framework for vector space modelling,” *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic,* vol. 3, no. 2, 2011

[23] B. Greer. “Sklearn LDA vs. GenSim LDA,” *medium.com* [Online] Available: https://medium.com/@benzgreer/sklearn-lda-vs-gensim-lda-691a9f2e9ab7 [Accessed April 2021]

[24] C. Sievert, K. E. Shirley. “LDAvis: A method for visualizing and interpreting topics,” *Proceedings of the Workshop on Interactive Learning, Visualization, and Interfaces,* pg. 63-70, 2014

[25] Affiliate Program. “Joining the Affiliate Program,” *Twitch.* [Online] Available: https://help.twitch.tv/s/article/joining-the-affiliate-program?language=en_US [Accessed May 2021]

[26] Twitch Support. “Chat Basics,” *Twitch.* [Online] Available: https://help.twitch.tv/s/article/chat-basics?language=en_US#:~:text=Chat%20is%20an%20essential%20part,in%20to%20your%20Twitch%20account.

[27] S. Teja. “Stop Words in NLP,” *Medium.* [Online] Available: https://medium.com/@saitejaponugoti/stop-words-in-nlp-5b248dadad47 [Accessed May 2021]

[28] F. Malik. “NLP: Text Processing Via Stemming And Lemmatisation In Data Science Projects,” *Medium.* [Online] Available: https://medium.com/fintechexplained/nlp-text-processing-via-stemming-and-lemmatisation-in-data-science-projects-ad4d5176060e [Accessed May 2021]

[29] F. Lengyel. “xQc,” *Twitch.tv.* [Online] Available: https://www.twitch.tv/xqcow [Accessed May 2021]

[30] P. Shimek. “AdmiralBahroo,” *Twitch.tv.* [Online] Available: https://www.twitch.tv/admiralbahroo [Accessed May 2021]