

School of Electronic Engineering
and Computer Science

Final Year
Undergraduate Project 2018/19



Final Report

Programme of study:

BSc Computer Science

Project Title:
**Dynamics of Twitch
chatrooms**

Supervisor: Dr Felix Cuadrado

Student Name:

Maksym Sokolenko

Student Id: 160482159

Date: 22.04.2019

Disclaimer

This report, with any accompanying documentation and/or implementation, is submitted as part requirement for the degree of BSc Computer Science at Queen Mary University of London. It is the product of my own labour except where indicated in the text. The report may be freely copied and distributed provided the source is acknowledged.

Acknowledgements

I would like to thank Dr Felix Cuadrado for the advice he gave me through the project and his supervision.

Thank you to my friends and relatives for their advice, moral support and motivation that helped me throughout this project.

Abstract

This project is centred around the streaming platform Twitch.tv and the dynamics of chatrooms of this website. On this platform there are many different channels with their own audience, the audience can interact with both the streamer himself and talk with each other via the chatroom. Each channel differs from one another, not only with their content but with their communities and hence chatrooms. This project aims to give a deep analysis of different chatrooms and analyse their behaviour, what kind of language they are using and what are the most common chat patterns used in both a single channel and across the website. The results

I also explain what risks are involved in this project.

Table of Contents

1.1 Aims and Objectives	7
1.2 Overview	8
Section 2.0	10
2.1 Background reading	10
2.1.0 Twitch.....	10
2.2 Variety Streamers.....	13
2.3 How Variety Streamers were chosen	13
2.4 Twitch definitions	14
2.4.0 Emotes	14
2.4.1 Emotes usage	14
2.4.2 Subscriber	15
2.4.3 Subscriber-only mode	16
2.5 Resources Consulted:.....	16
2.5.0 Chatty and chat log gathering.....	18
2.5.1 The initial idea of how to gather chat logs.....	19
3.0 Tools used	20
3.0.0 Chatty.....	20
3.0.1 Chatty usage.....	20
3.0.2 Pandas.....	20
3.0.3 The Jupyter notebook	20
3.0.4 Anaconda	21
3.0.5 Data analysis	21
3.1 Software:.....	21
3.2 Knowledge:	21
4.0 Requirements	22
4.1 Functional Requirements	22
4.2 Non-Functional Requirements.....	22
5.1 Preliminary Results	23
Vods emotes gathering	23
5.1.0 Clip emotes gathering	24
5.1.1 Summary	24
5.2 Graph Design	24
5.3 Graph Implementation	25

5.3.0 Analysis of channels	25
CohhCarnage.....	26
DrDisrespect.....	27
Forsen.....	29
Shroud	30
Sodapoppin	32
XQCow.....	33
5.2 Top Twitch, BTTV emotes.....	35
5.3 FFZ extension emotes	36
5.4 Most used emote	37
Omegalul.....	38
LulW	38
Summary	38
5.5 How spammy chat is in different channels	39
5.6 Further chat analysis.....	43
Summary	47
6.0 Final Results	50
6.1 Challenges.....	50
6.1.0 Twitch API and coding.....	50
6.1.1 Chatty application	50
Section 6.2 Project Risks	50
Risks Faced.....	51
6.3 What I Would change.....	52
Chosen Twitch channels.....	52
7.0 Conclusions	54
Legacy of the Project	54
Section X.X – References.....	55
Table of figures.....	56
Appendix	57
Revised time plan.....	57
Risk Assessment.....	60

1.0 Introduction and Objectives

1.1 Aims and Objectives

This project aims to show how numerous and unique chatrooms are on the platform and how they form individual communities. Originally, Twitch was a streaming platform mainly used for streaming games and game tournaments. Twitch can be compared to ordinary television where each streamer is viewed as an individual channel. During the growth of the platform, channels of each streamer started to grow along with it. Subsequently, this led to the formation of the communities around each channel. Currently, there are over two million registered streamers and over ten million daily active users. Some channels share viewers and play similar games. Channels that do not stick long to one specific game are called variety streamers. Variety streamers are the ones I am concentrating my attention on in this project, as they do not stick to one game. For example, variety streamer Moonmoon_ow in the duration of the week played several different games, like Welcome to the game 2, Pokemon Lets go Pikachu, One hour one life, Red Dead Redemption 2. Hence, a community formed in a channel consists of viewers that joined the stream while the streamer played different games, thus the community is a mix of viewers. Such communities represent not only one group of viewers but a wide variety of them and patterns created in these chatrooms will be used not only inside the specific channel but across the whole platform. (figure 1) Variety streamers can be called the face of twitch not only due to their popularity but also because newly introduced people to the Twitch will encounter one of these channels promptly. Streamers of these channels are also the ones that engage with their chat the most. Consequently, this leads to very active chatrooms which create patterns and spams that then become popular on the platform. My primary objective is to see that the majority of chatroom patterns and spams come from such channels. My further actions will be to analyse them and produce a document that supports this statement. This will be done by analysing data collected from the chatroom of specified channels. With these patterns being used but the viewers outside the website, it is self-evident how popular Twitch is and with its continuous growing can become a substitute for the usual television.

1.2 Overview

In this report, I will discuss my research carried out before beginning data analysis, the implementation which I chose to go with and what result I have obtained. Section 2 presents the background for this project and summarises the most relevant research of the topic. This section also contains all the definitions required in order to understand several aspects of the project. In Section 3, I present requirements of this project and knowledge that I needed in order to complete this project.

Preliminary results are shown in Section 4 with a small summary of the concluded work. Section 5 talks about design and implementation carried out in the project, I also show results concluded from the analysis of the gathered information. The next section talks about final results and the legacy of this project with possible improvements that could've been done. The following two sections show a time plan that has been followed and possible risks faced in this project. Furthermore, I conclude which of the risks had a significant impact on the project and what other risks that were not considered at the beginning I faced. The last section is dedicated to references to information that were used in this project. This report assumes a basic understanding of programming principles and little to no knowledge of Twitch.tv.



Figure 1 - Chatroom reaction to a joke that was told by the streamer (Lirik)

Section 2.0

In this section, I will discuss my initial research, background reading and programs I required to understand and learn during my project.

2.1 Background reading

2.1.0 Twitch

In order to understand why I decided to pick a variety streamers, it is crucial to look at the whole website first. Twitch is a popular online service for watching and streaming digital video broadcasts. When it was founded in 2011, Twitch originally focused almost entirely on video games but has since expanded to include streams dedicated to artwork creation, music, talk shows, and the occasional TV series. [8] On this website anyone can create an account and start streaming. While the user is streaming anyone can join their channel and watch them as well as talk to the streamer in chat or to other users who watch the stream as well. Each game or category has their viewer base who watch only specific channels as they are interested more in the game than the channel itself. For example, most game tournaments, events, presentations are all broadcasted on Twitch. If you visit the website for the first time without any specific idea of what to watch, the Twitch front page has a showroom of trending channels that are live and also has a small description of the channel. (figure 1) Under showroom, some sections have the most live viewers at the moment (figure 3) as well as Top live channels for Pc, ps4 or Xbox one. (figure 4). After watching one channel, if you want to come back and watch it again there is a "Follow" button that you can press and on the left side of the page "Followed channels" will be displayed (figure 2). Followed channels bar consists of all channels that you follow and is sorted by most viewers first. Channels that are currently offline will also be displayed in this bar underneath all channels that are live. Each channel can be seen as different TV channels with their own viewer bases and schedules when the stream is live. Each game or category has their viewer base who watch only specific channels as they are interested more in the game than the channel itself. For example, most game tournaments, events, presentations are all broadcasted on Twitch.

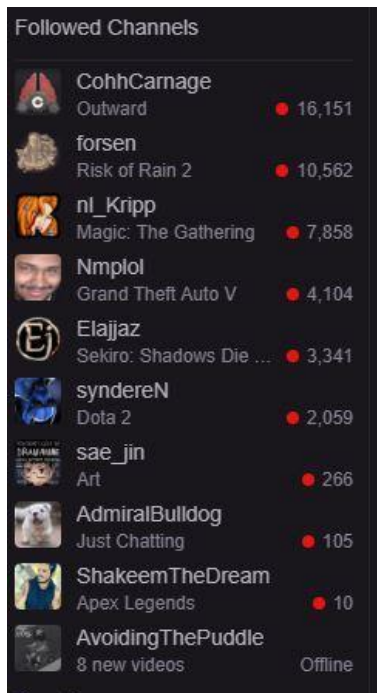


Figure 2 - Example of how Followed Channels bar looks like



Figure 3 - Trending channels on the main page of Twitch

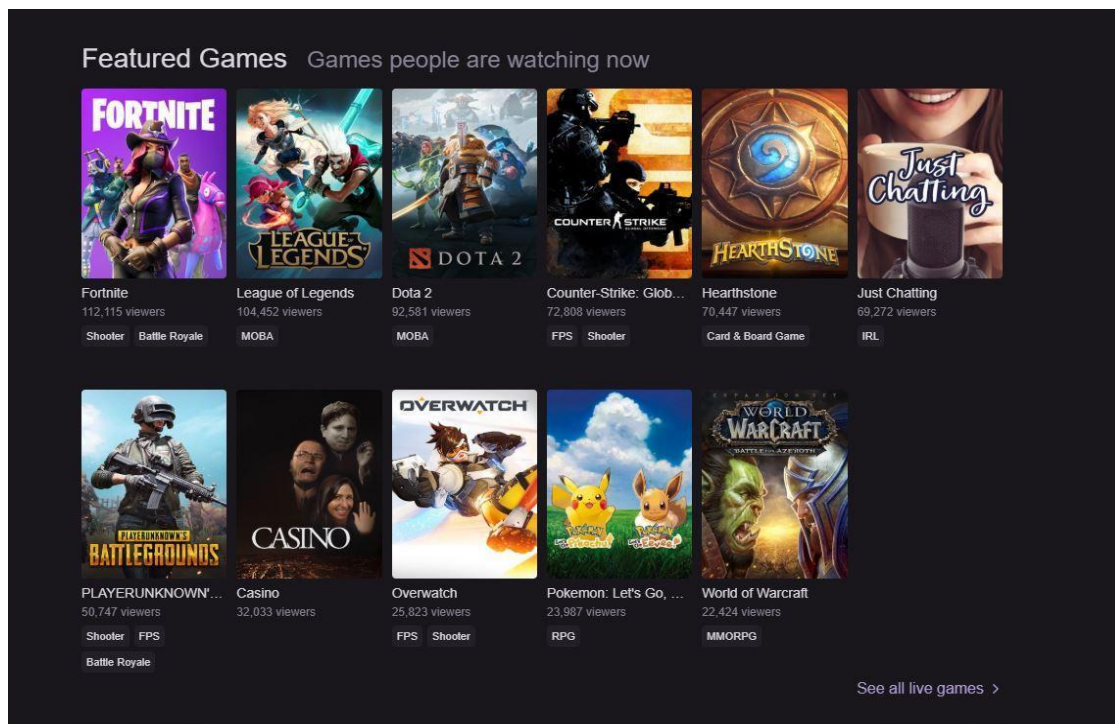


Figure 4 - Featured sections with the most viewers online

Each category shows all live channels filtered by most popular first. Games that are also played competitively will have most popular channels of streamers who are professional players, however, when a new game comes out most popular channels in that section would be the variety streamers. People tend to watch variety streamers not so much for the game they are playing, but because of the entertainment, that streamer himself provides. Such streamers already have a loyal fanbase who will watch them, not for the game they are playing which will boost the number of viewers they have to the top live channels, but as a new game is released. This is because people who are interested in the game itself tend to watch the most popular channels in that section. As a result, variety channels grow because they gather viewers from different sections of the website and continue doing it when they play another game. Since most sections will have variety streamers at the top, I decided to concentrate on them as they are not popular due to the game but famous as an entertainer who makes their popularity their achievement.

2.2 Variety Streamers

After deciding to concentrate my attention on variety streamers, I wanted to see how they are all connected and what kind of content do they provide. In order to do so, I look at the most popular variety channels on twitch and their most popular clips.

Twitch created a tool called twitch clip which lets anyone in the channel make a short clip of up to a minute during the live broadcast. Most of the time when people create a clip it is due to some information that the streamer said or if something entertaining happened. Top clips of all time from various channels have on average 100, 000 views and as they are popular in one channel other streamers also watch those clips on their channel. If such a clip had a joke that chatroom can turn into spam, they will do so and thus create a new chat pattern that it becomes popular will be used everywhere across the website. For my choice of streamers, I picked 20 of their most popular clips and analysed what emotes were used during the clip. I also used a program called “Chatty” to collect chat logs from live broadcasts to see what the most popular emotes that were used for the whole duration of the stream.

2.3 How Variety Streamers were chosen

In order for my research to consist of relevant information, I had to pick channels that would be suitable for this research. There are many variety channels that exist on the platform and an easy way to sort them by popularity is by using Twitch Metrics web site which has statistics on most popular channels. This statistic can be sorted by adding a filter “Variety Channel”. After sorting channels by variety channels, I picked channels from the top 50 most popular ones. However, I also applied additional filters myself while picking channels. All channels have to be consistently live; channels should not have any controversies surrounding them as this could affect the number of viewers this channel has as well as the behaviour of the chat room. All channels stream in the English language. Each channel is owned by an individual rather than a corporation and he is the only person who streams from this account. Web site information changes with time and most popular variety streamer that I picked in September 2018 could not be on the list in April 2019. That is why I should point out that channels picked for my research are most popular Variety streamers by the date September 2018.

2.4 Twitch definitions

Twitch has some specific words that require a general explanation.

2.4.0 Emotes

One crucial aspect of the web site is emotes. The audience has an option to express themselves not only with text but also with emotes. Emotes are like emojis on your phone but each one of them is originally designed by streamers themselves. There are emotes that are available to everyone – basic Twitch emotes and two free twitch browser extensions BTTV and FFZ. Each emote is typed using a specific word, for example by typing “Kappa” in chat it will be seen as an emotes rather than a word. There are also emotes that can be unlocked by connecting a twitch account to an amazon prime account which adds 39 more emotes to used. Furthermore, each partnered streamer has their own emotes slots for their channel with a maximum of 52 unlocked slots. These emotes can be accessed by subscribing to that channel. User is able to use these subscribers emotes in any channel on Twitch while their subscription is active.

2.4.1 Emotes usage

Every day in our life during conversations with people our face shows our reactions and how we feel regarding the topic we are talking about. With computers and phones, it is easier to express our emotions by using emojis. One of the first big software that used emojis was ICQ and later adopted by Skype. These applications can be seen as the main influence on why emojis became so popular. Today, modern phones have implemented emojis into the keyboard so that it is easier to access them. Similarly, as emojis used on the phone or Skype, Twitch.tv has its own emotes. Twitch provides numerous amounts of emotes that can be used to express nearly any real-life reaction. For example, if something funny happened on stream, instead of typing “haha” people tend to use an emote such as “Lul”, “Lulw”, “omegalul”. Most of the time it is easy to understand what an emote means by reading the text name of the emote or by looking at the picture of it. However, some emotes requires you to understand the context in which it has been used in order to understand what it means, a good example would be an emote “pepelaugh” which is commonly used when a streamer is clueless of the correct answer while viewers had figured out the answer already. Typical spam used during such event is “He doesn't

know pepelaugh” or “Don’t tell him pepelaugh”. Emotes became one of the core aspects of the website that Twitch is famous for. Similarly, how we use emojis on our phones viewers use emotes in chat. However, it is common to use fewer words and more emotes in chat as in popular channels where there are more than seven thousand concurrent viewers it will be hard to see the text message and easier an emote. In popular channels, the chat does not specifically talk to the stream and can be seen as a general reaction to what is happening on stream. They show reactions through the emotes for example: if something scary happened on stream - chat will spam “monkaS” or “monkaOmega”, if streamer is playing a song that chat enjoys, they will spam “forsenPls”, “SourPls”, “pepeJam”, “pepeD”. The chatroom for a big streamer is important as it can show whether viewers enjoy the stream or not which helps them to understand what will attract viewers and what won't.

Another important part that should be considered is that if one emotes becomes popular in a channel, other channels will also start using it. During my research, I found that subscriber emotes of one channel are not most commonly used in other channels, however, the majority of BTTV and FFZ most popular emotes are also one of the most popular emotes that are used in all channels that were used in this research. This shows that channels chatroom communicates indirectly as most commonly used emotes are the same.

2.4.2 Subscriber

Twitch channels that have a consistent viewer base for every stream have the chance to be partnered with Twitch. This gives the streamer a “subscribe button”. Viewers that want to directly support the streamer can purchase a subscription to the channel which gives them privileges of being able to use the chatroom if it is in “subscriber only” mode, chat room bots become more forgiving allowing subscribers to type more messages without being penalised for it. Non-subscribers have a limit of one message every thirty seconds. Any message that would be type faster will get deleted automatically. Subscribers also unlock subscriber only emotes that only they can use. These emotes can be used across the whole platform. Subscription is bought for every channel separately.

2.4.3 Subscriber-only mode

Subscriber-only mode is a setting that can be applied on the chatroom making it possible for the user to access chat only if they have a subscription to the channel.

2.5 Resources Consulted:

In order to create a database of twitch clips, I had to use the twitch API to get information about clips from the website itself (figure 4). In order to do so, I had to log in using my twitch account on the twitch developer dashboard, register an application and receive a client ID where the information that was requested would be sent. In my case all the requests were done on a local host. I did not have a broad knowledge of how to retrieve information from the website, or how to use git bash before this project. The main idea was to go through top 20 clips of all time for every streamer that I have picked look at the chat during the period of the clip and look at what emotes did they use. There is also a website called "Stream Elements" that shows the most popular emotes used on the website as well as emotes used in specific channels. Having both information about clips and emotes used in clips as well as most popular emotes I can create graphs with a visual representation of emotes that were used. This step required me to use git bash to collect information about clips (figure 5). An example of a GET call:

GET

'https://api.twitch.tv/kraken/clips/top?channel=NameOfTheChannel&period=all&trending=false&limit=20'

This call will return most popular clips of all time sorted by a number of views from the most to the least. The number of clips returned is limited by 20 but could be any number.

I stored the gathered information about clips in text files so that if any clip became out-of-date, this information would show which exactly clips were used for analysis in this project.


```
{
  "slug": "SilkyAmazingTruffleKie88",
  "tracking_id": "251964495",
  "url": "https://clips.twitch.tv/SilkyAmazingTruffleKie88?tt_medium=clips_api\u0026tt_content=url",
  "embed_url": "https://clips.twitch.tv/embed?clip=SilkyAmazingTruffleKie88\u0026tt_medium=clips_api\u0026tt_content=embed",
  "embed_html": "\u003ciframe src='https://clips.twitch.tv/embed?clip=SilkyAmazingTruffleKie88\u0026tt_medium=clips_api\u0026tt_content=embed' width='640' height='360' frameborder='0' scrolling='no' allowfullscreen='true'\u003e\u003c/iframe\u003e",
  "broadcaster": {
    "id": "96769016",
    "name": "tinthetatman",
    "display_name": "TinTheTatman",
    "channel_url": "https://www.twitch.tv/tinthetatman",
    "logo": "https://static-cdn.jtvnw.net/jtv_user_pictures/tinthetatman-profile_image-4cb867e7d8af3440-150x150.jpeg",
    "curator": {
      "id": "11859687",
      "name": "painsushi",
      "display_name": "PainSushi",
      "channel_url": "https://www.twitch.tv/painsushi",
      "logo": "https://static-cdn.jtvnw.net/jtv_user_pictures/painsushi-profile_image-382d5bea95de2a52-150x150.jpeg",
      "vod": null,
      "broadcast_id": "28984247856",
      "game": "IRL",
      "language": "en",
      "title": "Recognize",
      "views": 205588,
      "duration": 35,
      "created_at": "2018-06-06T14:45:10Z",
      "thumbnails": {
        "medium": "https://clips-media-assets2.twitch.tv/AT-cm7C251964495-preview-480x272.jpg",
        "small": "https://clips-media-assets2.twitch.tv/AT-cm7C251964495-preview-260x147.jpg",
        "tiny": "https://clips-media-assets2.twitch.tv/AT-cm7C251964495-preview-86x45.jpg"
      }
    }
  }
}
```

Figure 5 - An example of returned information about tiwtch clips from the website.























	TOP TWITCH EMOTES		TOP BTTV EMOTES
1	 Kappa 358,374,822	1	 FireSpeed 598,925,457
2	 TriHard 325,044,317	2	 LuL 164,121,185
3	 PogChamp 303,708,389	3	 haHAA 99,125,965
4	 4Head 236,483,408	4	 FeelsBadMan 57,915,232
5	 LUL 225,263,468	5	 Clap 54,291,439
6	 cmonBruh 171,752,488	6	 AlSmug 53,504,769
7	 gaminaKappa 161,097,188	7	 forsenPls 53,039,866
8	 119,272,364	8	 FeelsGoodMan 45,752,915
9	 BibleThump 117,070,666	9	 sumParrot 38,021,025
10	 90,186,958	10	 gachiGASM 34,398,315

Figure 6 - Most used emotes across the whole website.



Figure 7 - Most popular emotes used in a single channel

2.5.0 Chatty and chat log gathering

First, I wanted to collect data from the same broadcasts that the clips I gathered were from, however, many of these broadcasts are no longer available and can't be accessed. There is also no way that Twitch itself lets you collect information about past podcasts but there are bots created which connect to the live chat and get chat messages. I decided to collect information from live chats to compare whether the most popular emotes used in the chat are the ones used in the clips as well as are they are the ones that the most popular emotes in the channel and on twitch itself. In order to achieve this, I researched if it's possible to program a twitch chatbot to save chat message into a file.

2.5.1 The initial idea of how to gather chat logs

During my research, I tried to create a python chatbot that would connect to a specified chat room and collect each message in chat into a text file. However, Twitch.tv doesn't allow unauthorised bots to stay in the channel and they immediately disconnect them from chatroom as well as ban them. This is because bots in the channel are seen as viewbots which boost the number of live viewers a channel has and it is against the rules of the website to viewbot a channel.

3.0 Tools used

In this section, I will explain each tool that I used for this project.

3.0.0 Chatty

After numerous attempts to create a chatbot and further research, I found a program “Chatty”. Chatty is a Twitch chat client for desktop featuring many Twitch features. One of the features that was important for my project is that it connects to chatrooms and is able to store chat logs into a file. The difference between a chatbot and Chatty is that Chatty requires the user to log in using their Twitch account which it uses to connect to a chatroom thus viewbot issue is resolved. I used this program to connect to channels that I picked for this project and collect chat messages.

3.0.1 Chatty usage

In order to meet my requirements, I connected to each stream using chatty for five hours and saved all the messages that were typed during that period. This process was repeated 2 times in order to achieve requirements of analysing 2 broadcasts of each channel.

3.0.2 Pandas

What is Pandas? Pandas is an open source Data Analysis Library that provides many useful tools for developers in Python programming language.

In my project I used pandas in order to create graphs for every streamer to visually show the usage of emotes in their channel, emotes used across the whole web site and in order to compare whether emotes used both in clips and vods are the same as most used emotes in the channel and across the platform.

3.0.3 The Jupyter notebook

What is Jupyter notebook? The Jupyter Notebook is a web application that allows to create python documents that contain live code, create a table and visualise code.

Jupyter has many uses but the main features that I needed for this project are statistical modelling and data visualization.

3.0.4 Anaconda

What is Anaconda? Anaconda is an open-source distribution of the Python programming language for scientific computing with the main aim of simplifying package management.

The main advantage of Anaconda is that it has all the packages and libraries that are required for data analysis. (Pandas, Matplotlib and The Jupyter notebook)

3.0.5 Data analysis

After collecting all the data, I created graphs of data with different analytics. First, I analysed all the emotes that were used in clips and vods. Information on each channel was compared to each other in search of anomalies or similarities. Further analysis includes analysing which are the most popular twitch emotes used compared to those provided by the web-site Stream Elements, how many emotes are used per minute.

3.1 Software:

- Python 3.7 [<https://www.python.org/>]
- Chatty twitch chat client [<https://www.chatty.github.io/>]
- Anaconda [<https://www.anaconda.com/>]
- Git [<https://www.git-scm.com/>]
- Twitch developer console [<https://www.dev.twitch.tv/console/>]
- Jupyter notebook [<https://www.jupyter.org/>]
- Pandas [<https://pandas.pydata.org/>]

3.2 Knowledge:

- Python
- Basics of how Twitch API works
- Basics of Pandas
- Basics of Anaconda
- Basics of data analysis in general

4.0 Requirements

In this section, I will describe and identify requirements for this project. This section will consist of Functional and Non-Functional Requirements.

4.1 Functional Requirements

These are the core components of my project.

Analysis of all channels: Each channel will be analysed individually and in case of any anomalies being present, a detailed explanation will be provided.

Comparison of channels: After previous step completed, channels will be compared to each other to find similarities or differences between them. Important cases will be analysed and a detailed explanation will be provided.

Channels comparison to global emotes: Each channel will be compared to most popular global emotes in order to find additional similarities and possibly prove the hypothesis of this project.

Identify possible patterns: During the analysis, identify which emotes are used in all channels often and are there any possible patterns formed. Depending on the result do additional analysis to prove whether it's a pattern that is common for several channels or not.

4.2 Non-Functional Requirements

Re-usability: One of the main non-functional requirements of this project is the idea of this project paper being re-used possibly by students who will do a project on this topic next year. This document can be re-used again in order to compare the result of the year 2019 with their year. This will provide comparison whether Variety Channels that I used in my project are the same or changed, new patterns in the chatroom that appeared and what emotes are the most popular now.

Visibility: All the data gathered in this report should be visibly presented or be supported by a picture to give an example of a statement being true in the chatroom.

Data relevance: Collected data should be up-to-date and be relevant. If data changes fast, date of data relevancy should be specified.

Section 5.0 Design and Implementation

5.1 Preliminary Results

Vods emotes gathering

Application Chatty collects full messages from the chat and saves them into a text file. Each text file is for an individual stream. Chatty also records the time when the message was typed. Time stamps consist of hour and minutes when the message was sent. This feature significantly helped in making sure that all files were exactly 10 hours. The main objective was to join the chatroom when the streamer is already live. The reason I did not join the stream as soon as the streamer went live is due to the fact that all streamers used in this project do introduction pieces to their stream. The 'intro' is commonly just a screen with a message "Stream starts soon" and music playing in the background. This is common practise amongst streamers as it allows them to notify users that the stream is live and wait until the audience joins. I joined chatrooms as soon as the 'intro' was over and gameplay started. On average streamers stream for 6-8 hours. Therefore, I joined live streams several times in order to achieve my requirement of 10 hours of chat messages recorded for each streamer. In order to be sure that the data is exactly 10 hours. Since I collected information only from live broadcasts, any line that was saved from the offline chat was deleted.

While I required only the emotes that were used during a live broadcast. Text files consisted of approximately over 80000 lines on messages which would be very difficult to use in pandas as I would have to create a database for all messages, filter them to find emotes that were most commonly used and then plot a graph that consists of only emotes. In order to solve this problem, I wrote a python code that would go through the text file for every streamer and return most commonly used words. (figure8) The amount of words that are returned is specified in the code and can be changed to any number. Code output also required further filtering as it printed out both emotes and words that were used. After additional filtering, results were put in pandas using jupyter notebook. Each streamer has a separate graph of emotes used in vods and clips.

```
import re
from collections import Counter

words = open('words.txt').read().lower().split()
count = Counter(words).most_common(200)
print(count)
```

Figure 8 - Code for sorting text file to separate words from emotes

5.1.0 Clip emotes gathering

The initial idea was to analyse clips automatically, however, it is not possible to download clips and chat logs of old broadcasts from the website. Therefore, I manually watched every top clip of each broadcaster, copied all messages from the chatroom and put it into a text file which was then filtered the same way as the live broadcasts. After repeating this procedure for each streamer, I made a graph that consists of the most popular emotes used in 20 most popular clips of all time for each streamer. Emotes that were used less than 20 times were not included in the graphs.

5.1.1 Summary

After having all data gathered, which was a timely task at first, I was able to easily use obtained information in pandas to produce graphs. Further analysis required me to use the same information that was obtained for different graphs and different hypothesis.

An initial challenge was observed; Vods data could only be gathered when the broadcast was live. Several times live stream did not meet the requirement of duration and ended earlier as well as I had to wait for the broadcasts to go live on specific dates/times. This led to delayed data analysis. Furthermore, data collected from emotes used in clips was not automated and also required a lot of time to collect it.

5.2 Graph Design

Pandas provide many types of graphs that can be plotted. For my project, I picked bar graphs as visually it is easier to distinguish between data on a bar graph. First, I plotted graphs for each streamer separately for each bit of information that I have: graph for emotes used in clips, emotes used in vod, most popular emotes of Twitch

and extension BTTV and a separate graph of most popular emotes of extension FFZ. Twitch and BTTV graph contains twenty emotes: top ten used emotes for each one of them respectfully. I decided to separate FFZ to a separate graph because during my research I noticed that the majority of most popular emotes that are used across all channels in vods are FFZ emotes thus more emotes should be taken into consideration. FFZ graph consists of the top twenty most used emotes. Moreover, emotes used in clips consists of all emotes that were used throughout the duration of the clip and vary from channel to channel while most used emotes in vods consists of thirty most frequently used ones. In order to show how similarities and differences between vod emotes and clip emotes I combined both graphs. Vod graphs consist of a large amount of emotes used and it is hard to see on the combined graph how many emotes were used in top clips. An additional graph is provided for every channel that consists of only to clip emotes.

5.3 Graph Implementation

5.3.0 Analysis of channels

In this section, I will analyse some channels that - during my research - showed important anomalies or similarities. It is important to all levels of similarities that channels have between each other and what makes them different. Furthermore, I will show other statistical graphs in order to prove one of my hypotheses.

CohhCarnage

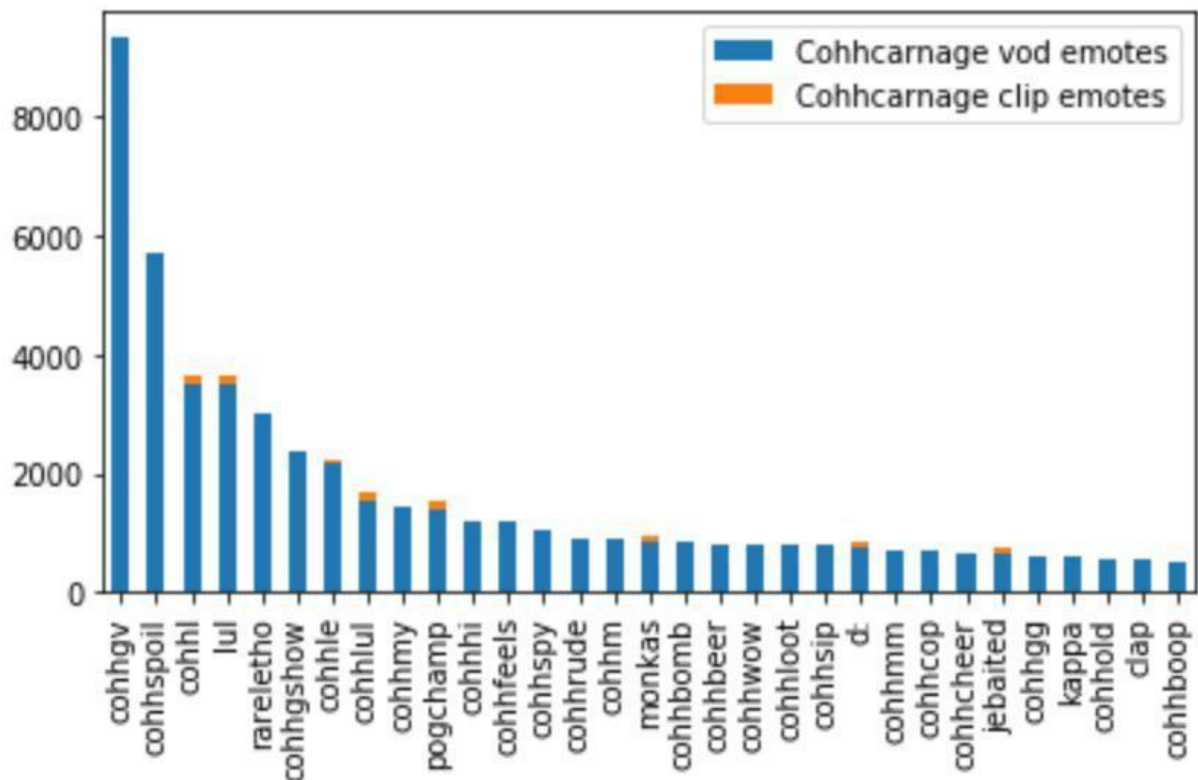


Figure 9 - CohhCarnage's graphs of top emotes used in vods.

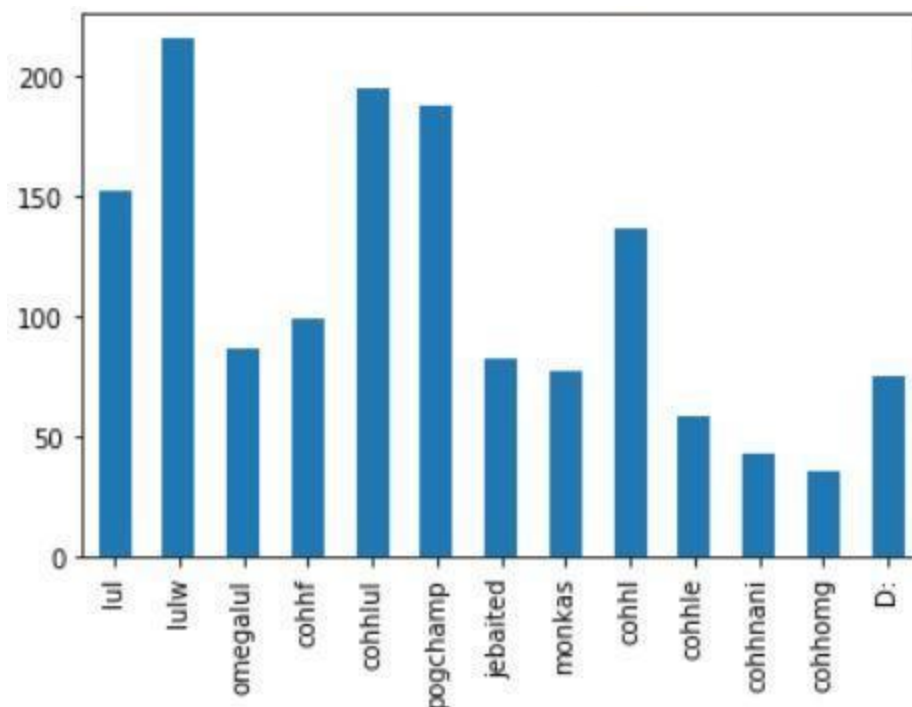


Figure 10 - CohhCarnage's graphs of top emotes in clips.

CohhCarnage's channel is a good example of a stream that rarely uses "sub-mode" and yet has a very active subscriber viewer base with the majority of all emotes used both in clips and vods being subscriber emotes. Any emote that has "cohh" in front of it is a subscriber emote. However, five of the emotes used in clips are not most used in vods ("lulw", "omegalul", "cohhf", "cohhnani", "cohhomg"), with the anomaly of most popular clips having emote "lulw" in first place while not being present in most commonly used emotes in the vod.

DrDisrespect

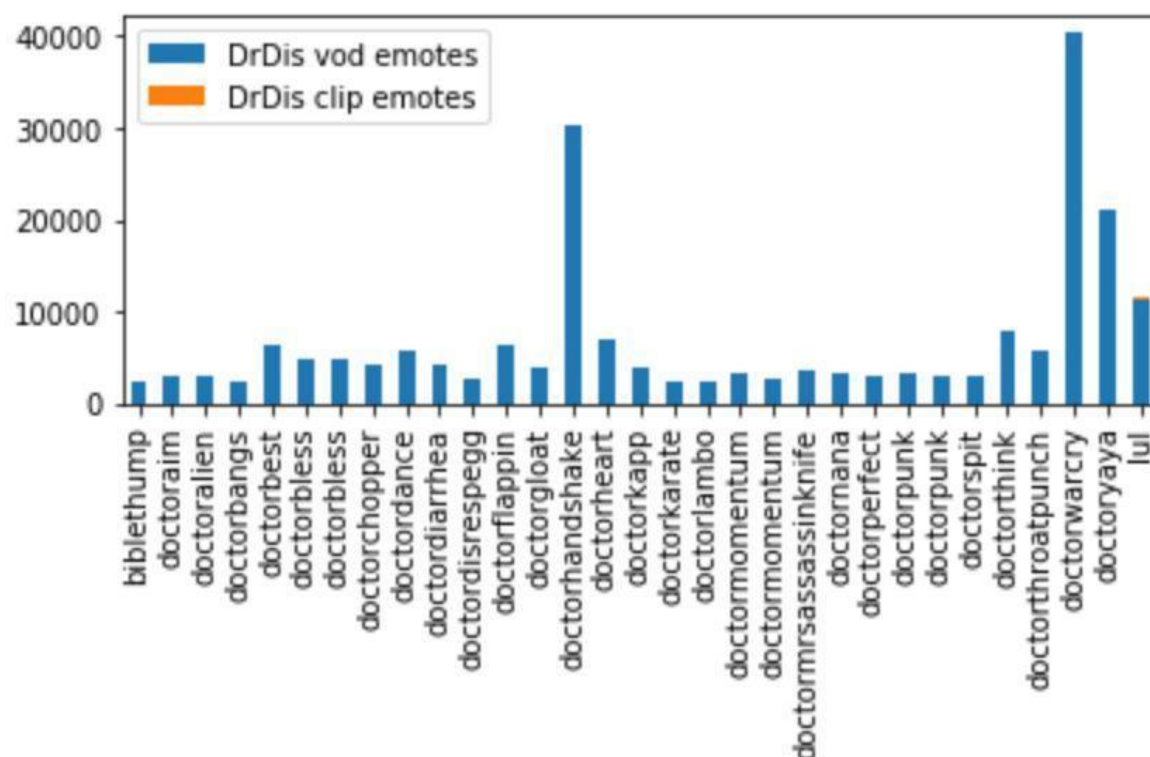


Figure 11 - DrDisrespect's channel graph of top emotes used in vods

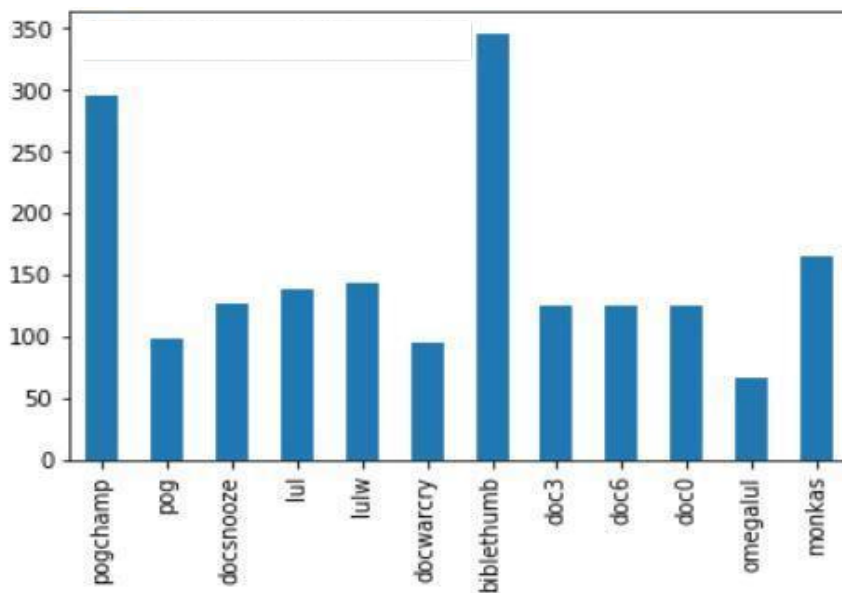


Figure 12 - DrDisrespect's channel graph of top emotes used in clips.

DrDisrespect's channel chat is always in subscriber-only chat which shows how active subscriber viewer base is with the amount of emotes used. Any emotes with "doc" or "doctor" in front is a subscriber emotes. In some cases, there are more emotes used than in other channels which don't have the limits of subscriber mode active. There are only two non-subscriber emotes on vods graph ("lul" and "biblethumb") which are also present in clips graph with "biblethumb" being the most used one". This channel's graphs also show that popular clip emotes were used during a specific moment that happened and are not commonly used during a live broadcast as in a long-term, subscriber emotes are used more commonly.

Forsen

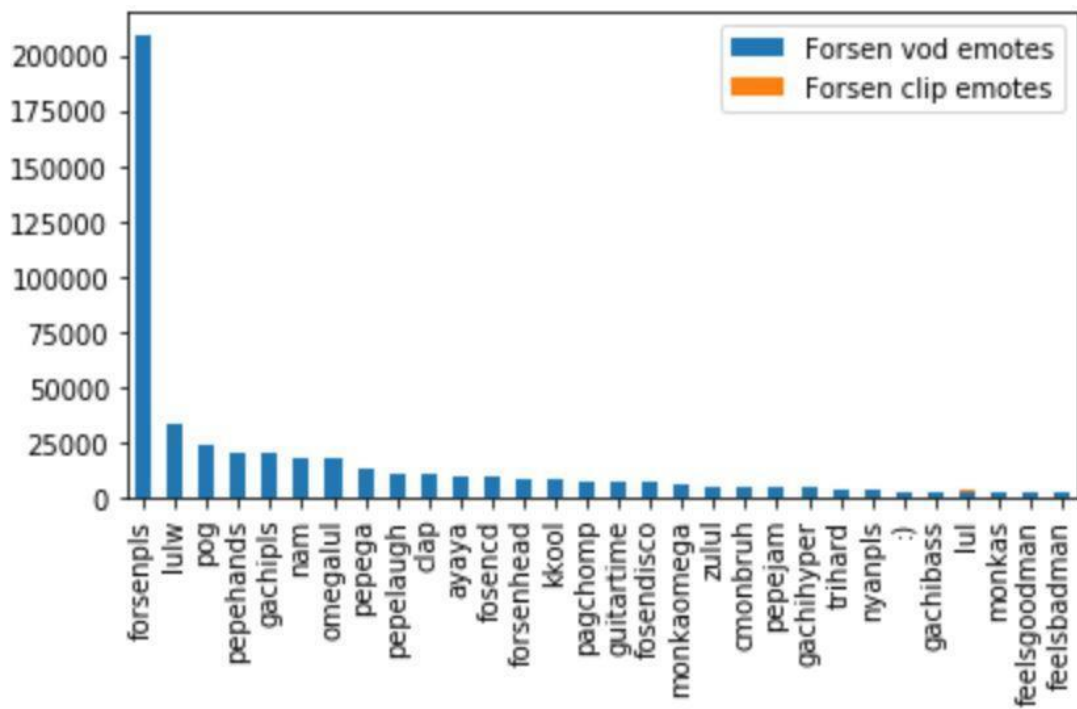


Figure 13 - Forsen's channel graph of top emotes used in vods.

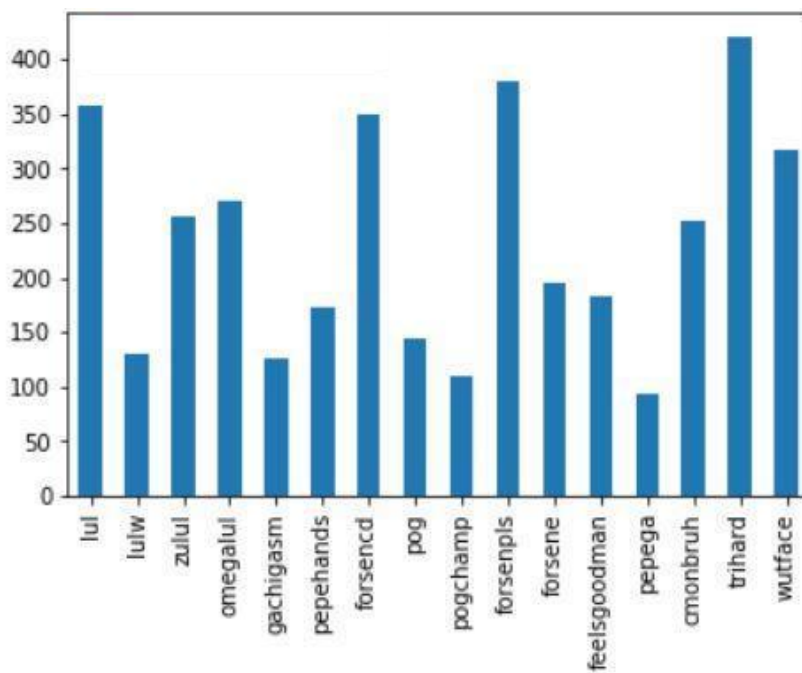


Figure 14 - Forsen's channel graph of top emotes used in clips.

Forsen's chatroom is a great example of an extremely active chat that uses an enormous amount of emotes in their spams with the most popular emote "forsenpls" that was used 209047 times. There are only two emotes that are popular only clips ("lulw" and "wutface") while other emotes are frequently used in both cases. In order to show how popular Forsen's emotes are we need to look at statistics provided by the website Stream Elements which shows that Forsen's emote "forsenT", a subscriber emote, is number 23 most used twitch emotes of all time, "forsenCD" is number 25 most used FFZ emotes and "forsenPls" is number 9 most used BTTV emotes. Although "forsenT" is a subscriber emote and other 2 emotes can be used by everyone, they first have to be added to the channel. In order to use any emote in extension FFZ or BTTV, emotes have to be approved by the streamer himself and added to the channel. This show that other channels that approve Forsens emotes share the viewer base and their chats follow similar patterns.

Shroud

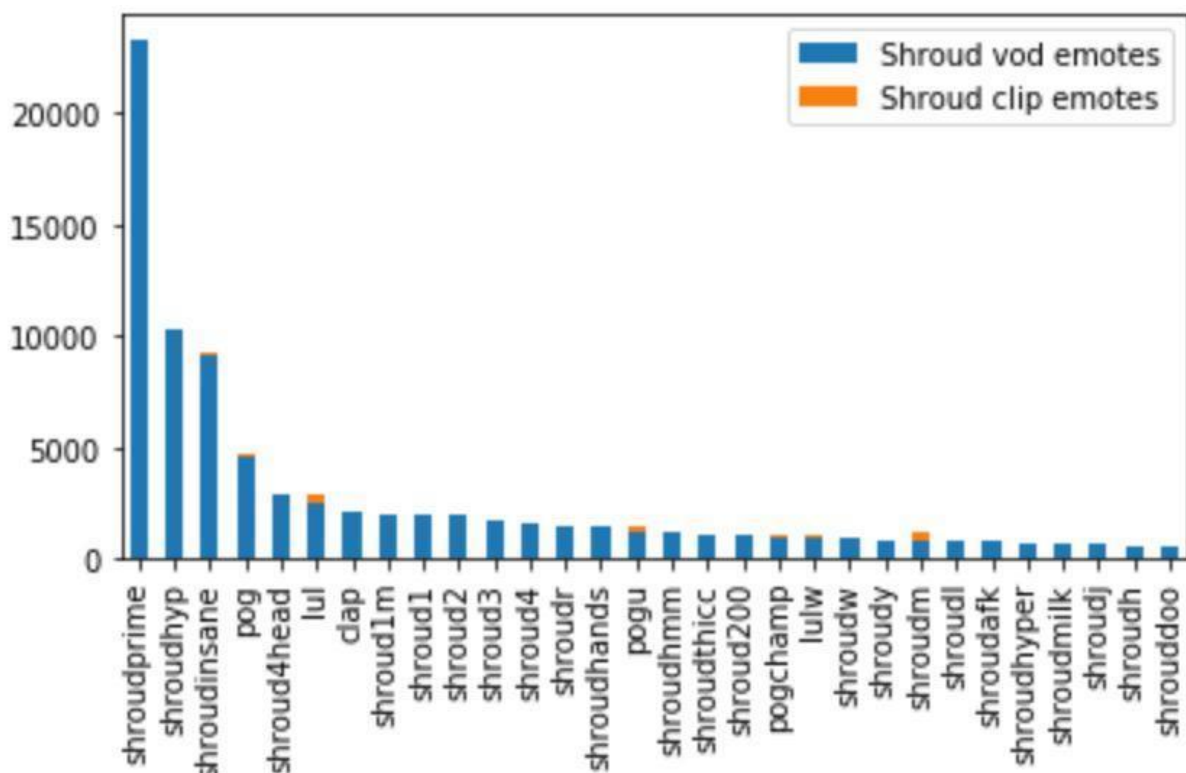


Figure 15 - Shroud's channel graph of top emotes used in vods.

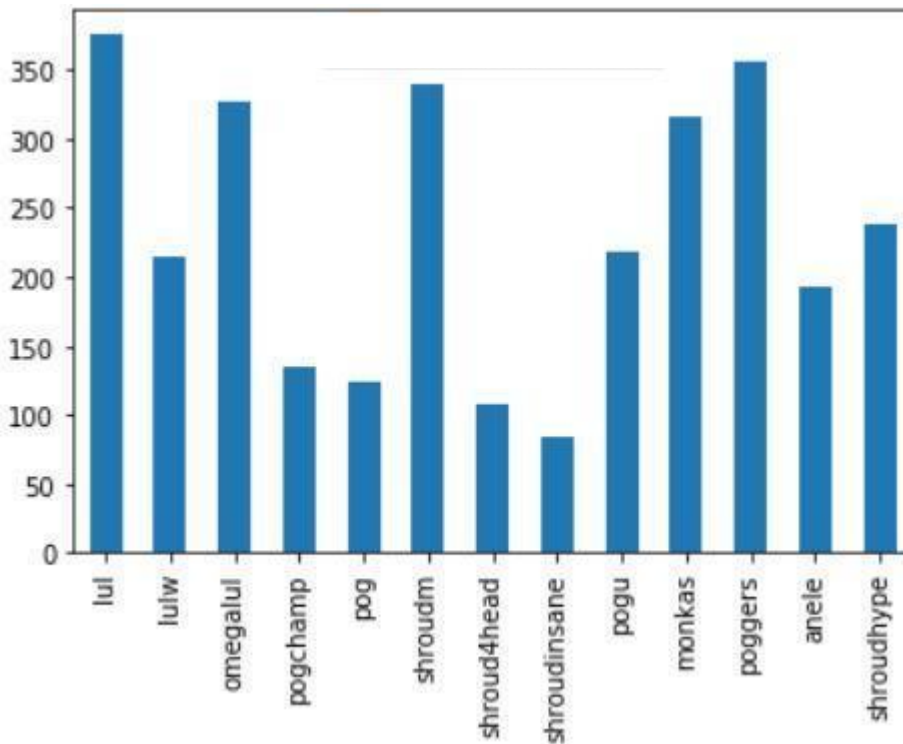


Figure 16 - Shroud's channel graph of top emotes used in clips.

Shroud is another example of a channel with an active subscriber viewer base as in vods there are only five non-subscriber emotes “pog”, “lul”, “clap”, “pogu”, “lulw”. Comparing this channel to channel DrDisrespect, this channel does not keep the chatroom in “subscriber only” mode but even so, the majority of the emotes are subscribers emotes. Statistics provided by Stream Elements also shows that Shrouds channel is currently the most subscribed channel on the platform. There is one pattern common to each streamer is to create their own emotes on a base of existing ones. For example, emote “4Head” has an alternative in this channel “Shroud4head”. Since the majority of emotes are replaced by subscriber emotes, subscribers will use them instead, however, there is no alternative to a few emotes thus their originals are used. These emotes are “pog”, “lul”, “clap”, “pogu”, “lulw”.

There are also emotes that popular in clips such as “omegalul”, “monkas”, “poggers”, “anele” that are not frequently used during the broadcasts. This shows that Shroud’s clips are capturing specific moments that happened on stream and thus people using emotes for a specific situation rather than spam common emotes that are shown to be popular in a long broadcast and not present in clips.

Sodapoppin

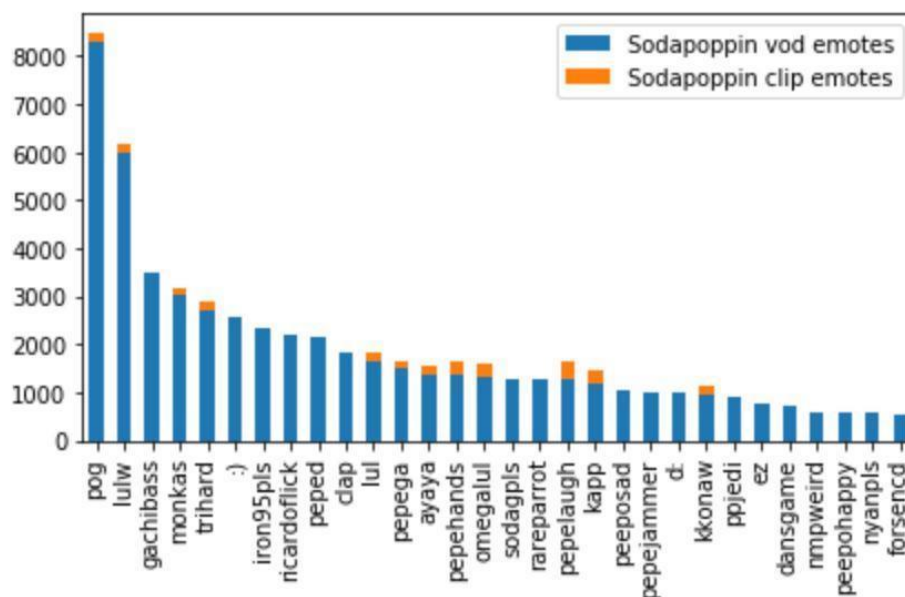


Figure 17 - Sodapoppin's channel graph of top emotes used in vods.

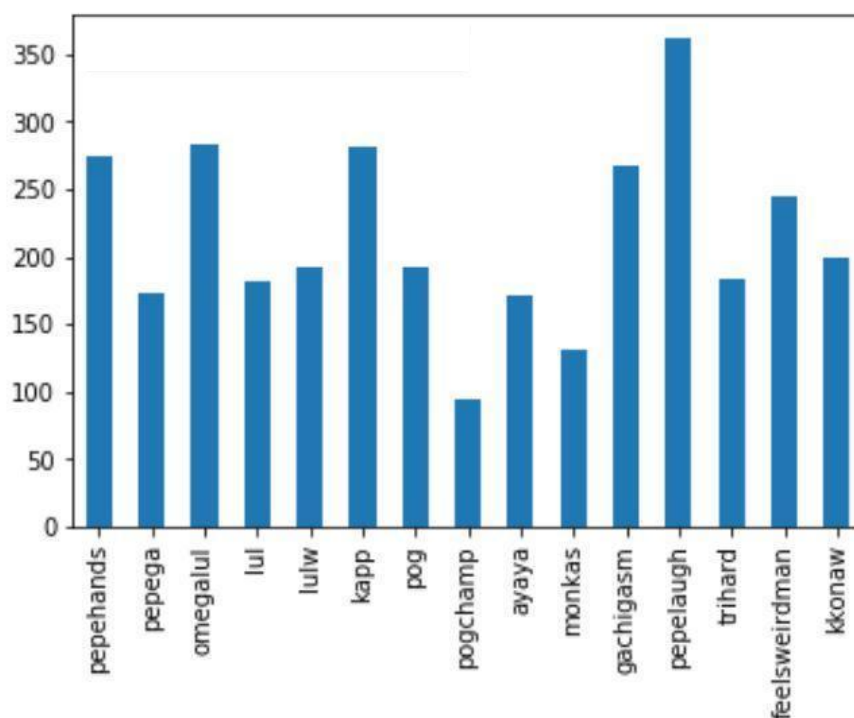


Figure 18 - Sodapoppin's channel graph of top emotes used in clips.

Emotes gathered in this channel's top clips are also all used often in vods. The anomaly of Sodapoppin's channel is that none of the most popular emotes in vods are subscriber emotes although it is one of the most popular variety channels. Moreover, as stated above, because streamers have to approve all BTTV and FFZ emotes that are going to be available in the channel, it shows that Sodapoppin

added an emote “forsenCd” that is an emote connected to another channel. Furthermore, it also shows that it is in top 30 top emotes used in the vods meaning the community uses this emote quite a lot thus proving that Forsen and Sodapoppin share quite a lot of viewers.

XQCow

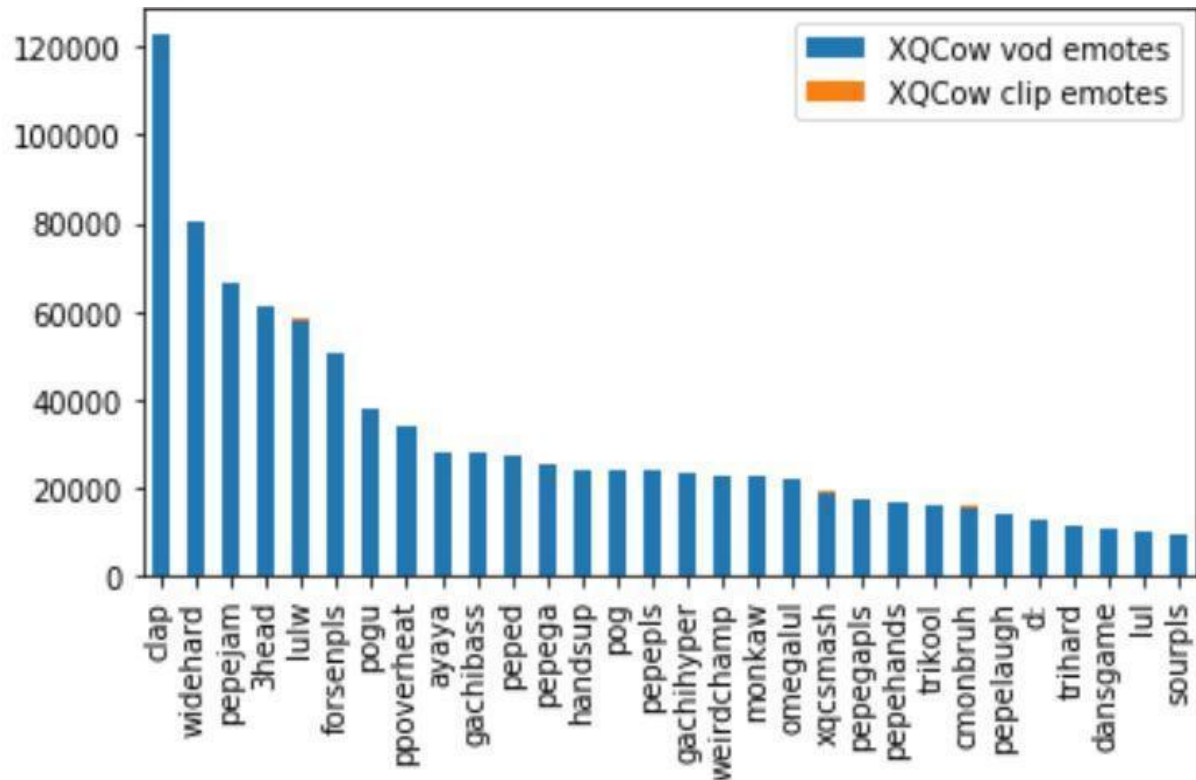


Figure 19 - XQCow's channel graph of top emotes used in vods.

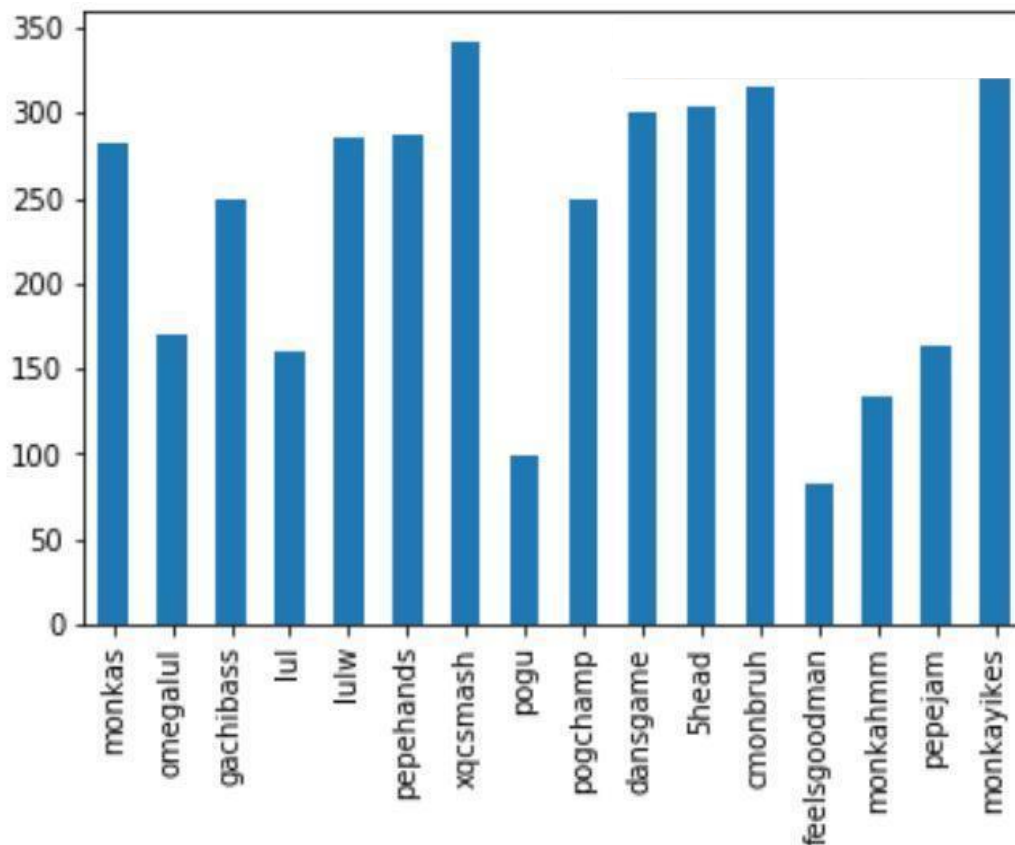


Figure 20 - XQCow's channel graph of top emotes used in clips.

XQC's chatroom, similarly to Forsen's chatroom, is submerged in messages with most used emote in vod reaching over 120000 times used. There is also a similarity between emotes used in clips and vods with most popular emotes being the same. The anomaly of this channel, similarly to Sodapoppin's channel, is that no subscriber emotes being among most popular in vods which show that even though chatroom is very active it is active as a whole rather than subscriber active. Another important aspect of this chatroom is to show how many emotes have been used in 10 hours compared to other channels. XQCow is one of the biggest streams, however, on average his viewer count is around 30,000 concurrent viewers. This number of viewers is also reached by the majority of streamers that I picked for analysis showing that this community is much more active compared to others.

5.2 Top Twitch, BTTV emotes

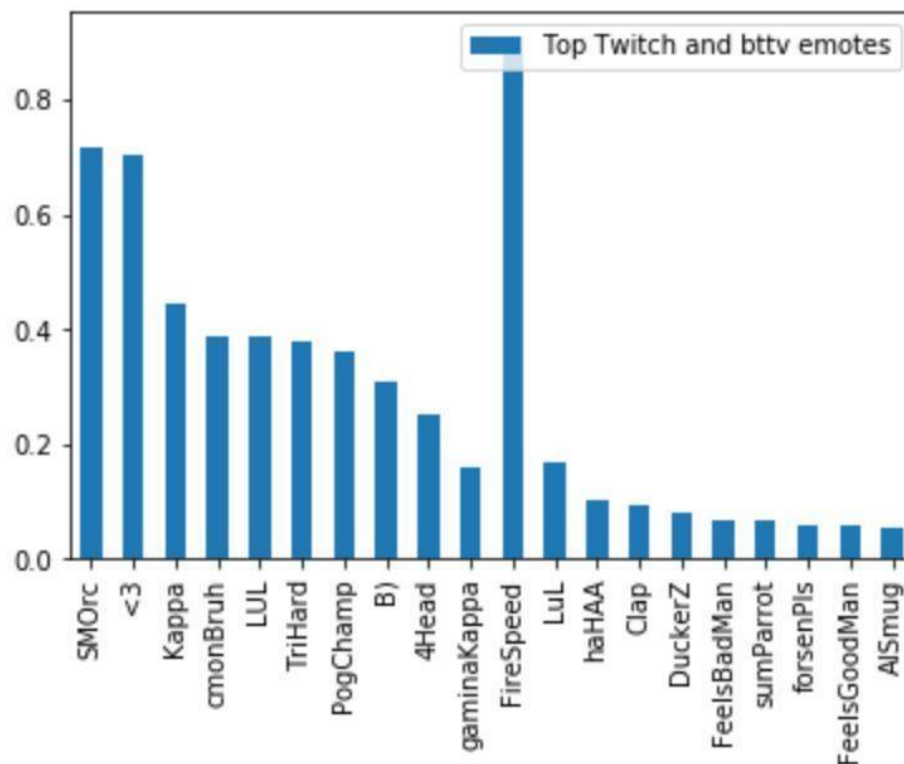


Figure 21 - Globally most used emotes of Twitch and extension BTTV.

The majority of all emotes presented on the graph have been observed in all channels with some of them being on the lists of most used emotes in vods or clips, however, there are two anomalies that I have encountered. Emote “FireSpeed” that has been used approximately 909331280 times was not encountered a single time in any channel during my research even though it is the most used emote of BTTV extension.

5.3 FFZ extension emotes

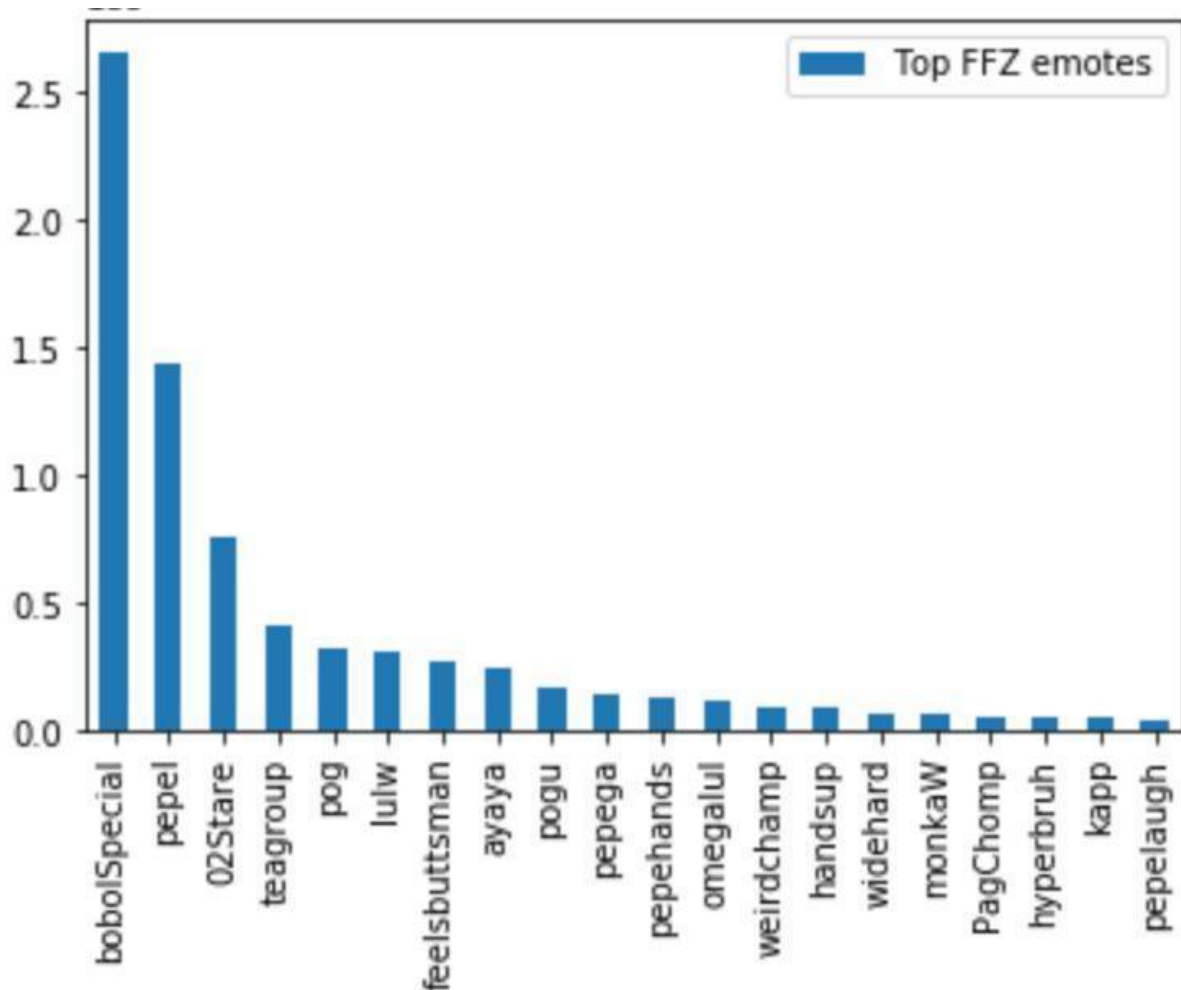


Figure 22 - Most used emotes of the extension FFZ.

I separated FFZ extension emotes into a separate graph due to the fact that top twenty emotes of this extension were more frequently used in variety channels chatrooms rather than top twitch global emotes or BTTV emotes. Nonetheless, there are anomalies with emotes: “boboSpecial”, “02Stare”, “teagroup” that were not used a single time in across all chatrooms. Apart from the anomalies of some emotes not being present in these channels, all other emotes are present in all channels. They might be not in top 30 emotes used in the channel but they were all used. This suggests that currently FFZ extension is more popular by viewers in these channels than emotes from BTTV extension or Twitch emotes. The only exception would be subscriber emotes; however, they are not available to everyone in the first place and have to be unlocked while these emotes can be used by anyone.

5.4 Most used emote

In this section, I wanted to show the similarity between different channels clips and vods by connecting them via most used emote. These graphs show which channels share the same top emote. I did not use graphs for two as the numbers were easier to represent as a table.

Lul

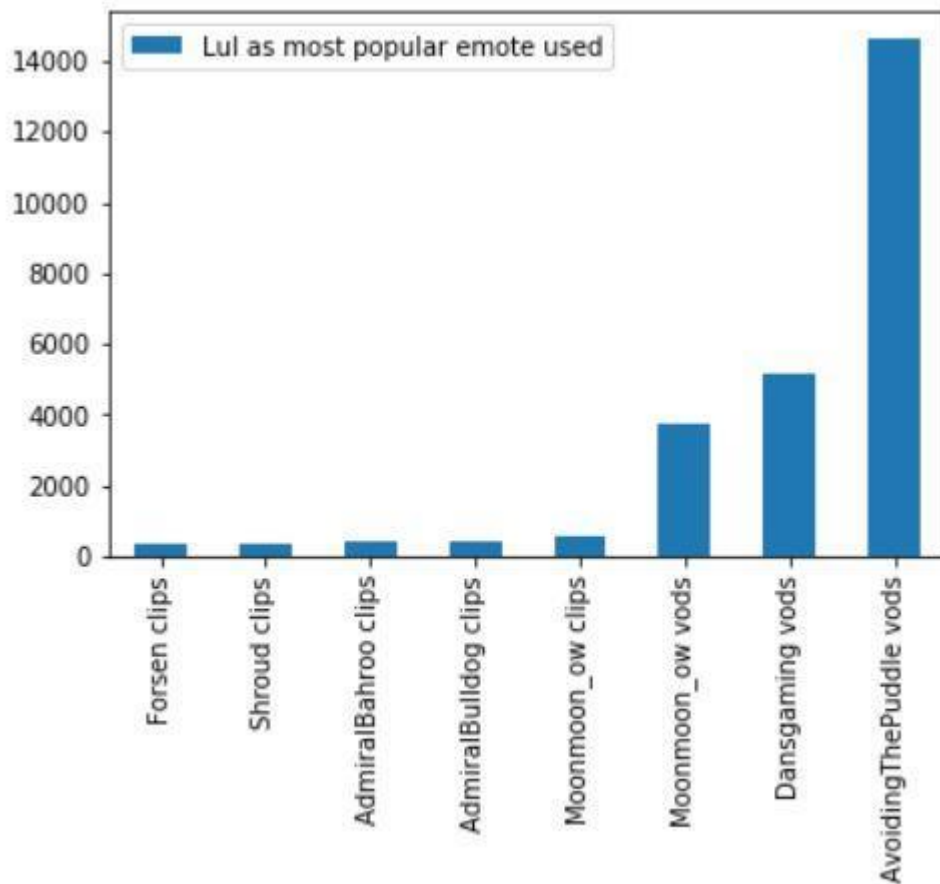


Figure 23 - Graph of cases when emote "LUL" was the most popular emote.

Omgalul

Omgalul as most popular emote used	
Destiny clips	326
Asmongold clips	372
Lirik clips	375
AvoidingThePuddle clips	396

LulW

LulW as most popular emote used	
CohhCarnage clips	215
Dansgaming clips	286
Destiny vods	3411
Asmongold vods	32872

Summary

Although originally, I expected to see more subscriber emotes to appear more in different channels to show how popular variety streamers share their viewer bases, I was only able to see this to be true only in one case of Sodapoppin's channel has one of the most used emote in vods from Forsen's channel. For the rest of the channels, I had to take a different approach of showing that some of the channels share most popular emote.

5.5 How spammy chat is in different channels

In this section, I want to show how different some channels are from others with previously shown graphs covering how many emotes were used in 10 hours during live broadcasts. However, not every channel graph is present due to no anomalies appearing in those chatrooms. The more viewers watch the stream, the more messages appear in the chatroom. Since it is impossible to see long messages in fast chatrooms, there is a pattern that can be seen in this analysis. The more viewers streamer has – the more emotes are used and messages typed in chat are shorter.

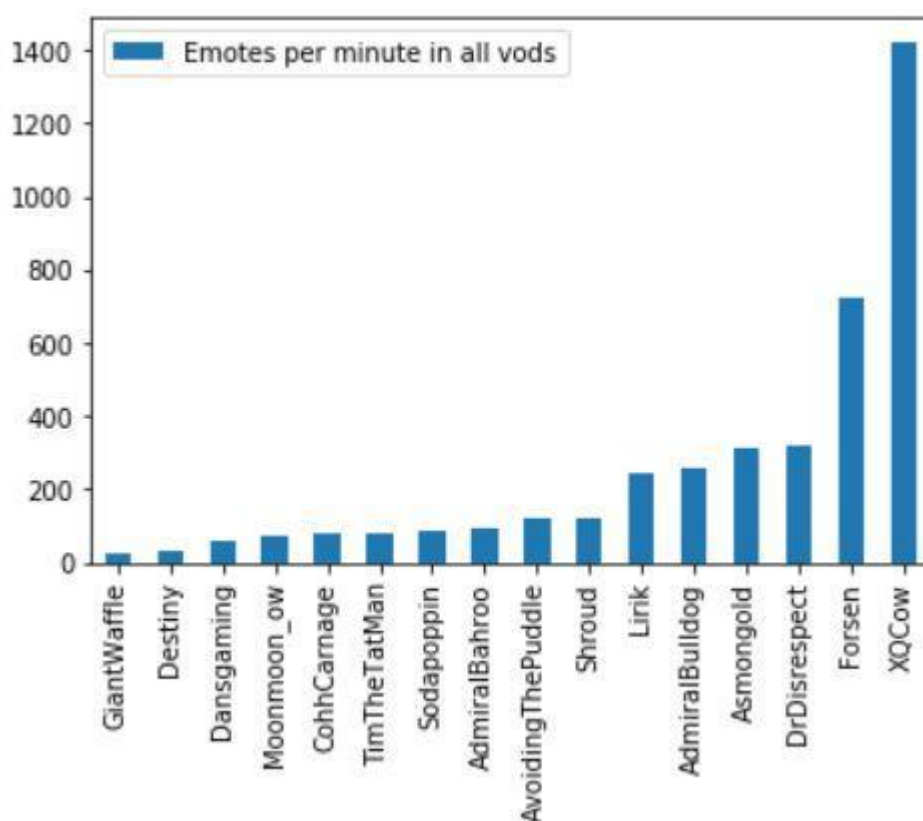


Figure 24 - Graph of how many emotes were used in vods per minutes.

This graph shows how many emotes approximately are typed in vod chat per minute. The number varies from Giantwaffle's chat with 23 emotes per minute to Forsen and XQCow chatrooms with 726 and 1418 emotes per minute respectively. One of the results that are expected was that less spam filled chat shows longer messages and higher interaction between the streamer and chat while more spam filled chats will behave on their own and show just a reflection of whether viewers enjoy the stream or not. Furthermore, a more spam filled chat will have less lucid messages compared to channels that show less emotes per minute. As an example, figure9 and figure10

shows two moments of streams. Giantwaffle is having a discussion with chat regarding a purchase of a new chair with Forsens chat reacting to gameplay that they do not enjoy.

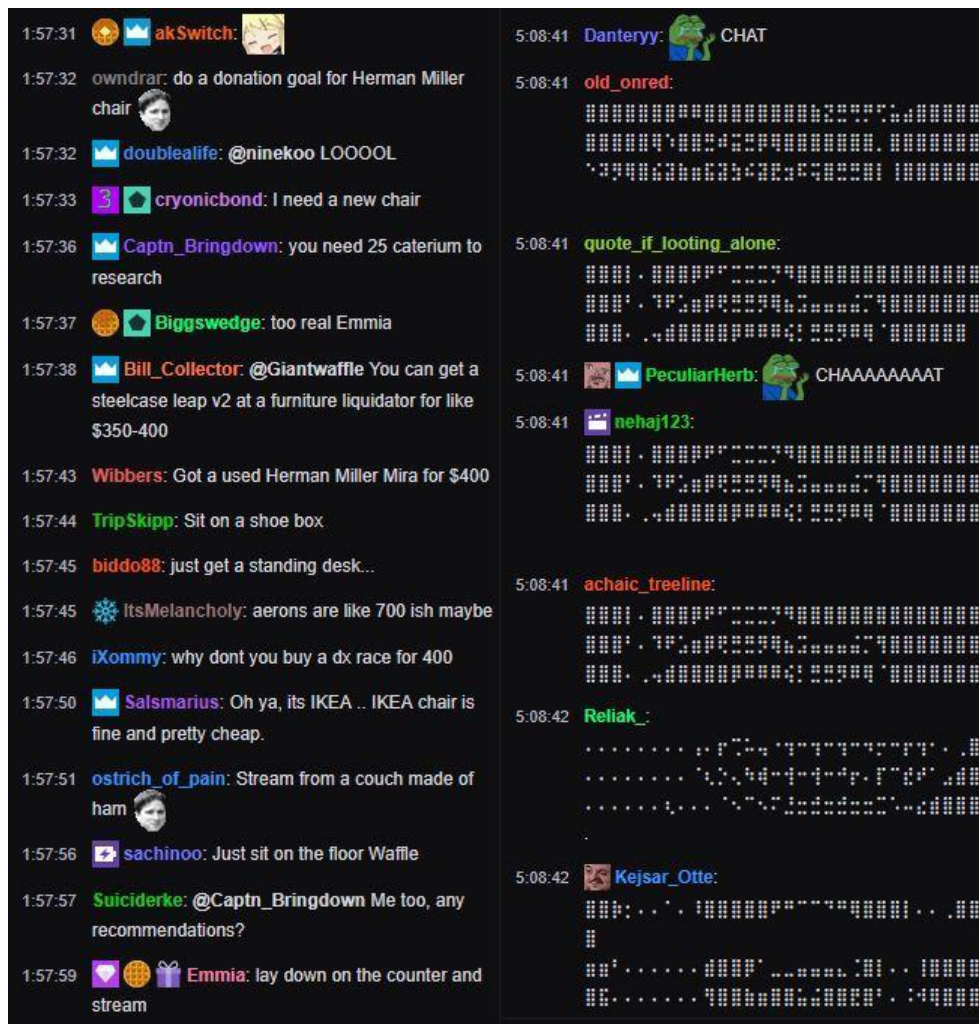


Figure 25 - Examples of a non-spammy and a spammy chats.

Channels with a spam filled chatroom have different interactions compared to other channels. In order to show how a spam filled chatroom reacts to something happening on stream, I will use an example from XQCow streams figure11. The streamer was saying number 2 repeatedly for a duration of 5 seconds and as a result whole chatroom started spamming 2. This is a good example of interaction as the majority of all interactions are similar. For example, if the streamer asks a yes or no question, chat will spam what they believe the correct answer is and the streamer will follow the advice of the majority of the answers that he sees.



Figure 26 - Chatroom repeats the message after streamer.

To support the argument that spam filled channels do not necessarily talk to the streamer as individuals and more as a group of viewers, figure12 shows how chat reacts to a song played by the streamer.



Figure 27 - Chatroom shows that they enjoy the song played by the streamer.

Chatroom shows that they enjoy the song that is playing by consistently spamming two emotes (“TriKool hyperclap”) compared to if they didn’t enjoy it, they would ask streamer to skip it and spam an emote “Dansgame”. This show that channels with a high number of concurrent viewers do not interact directly with each other. This puts a challenge on viewers of how to make their message visible. The solution is to create a chat pattern that the majority of other viewers will like and spam it as well so that the streamer will see it. These patterns are later going to appear in other channels as well due to its popularity. This hypothesis can be proven by looking at the figure12 which is taken from the channel” Pokelawls” which is a channel that was not used in this project for analysis. These emotes can be compared to most used emotes of XQCow vods which are “widehard” and “clap”. These emotes are the same as “TriKool” and “hyperclap” and used in the same way with the only difference that these emotes are just variations the same emotes. This shows that this chat spam pattern has been adopted by another channel’s chatroom. Although it is possible to perform statistical analysis on how many and which emotes are used, it

would not give a correct representation of why these emotes are the most popular. This is because one of the main, important aspects of the emote patterns is the context in which they are used. Depending on the situation or what the streamer is talking about different emotes are going to be used. Therefore, the above analysis is justifiable as it shows an example of emotes being used in different contexts.

5.6 Further chat analysis

It is important to note that the number of viewers does not directly mean that the chatroom will use more emotes and spam more messages.

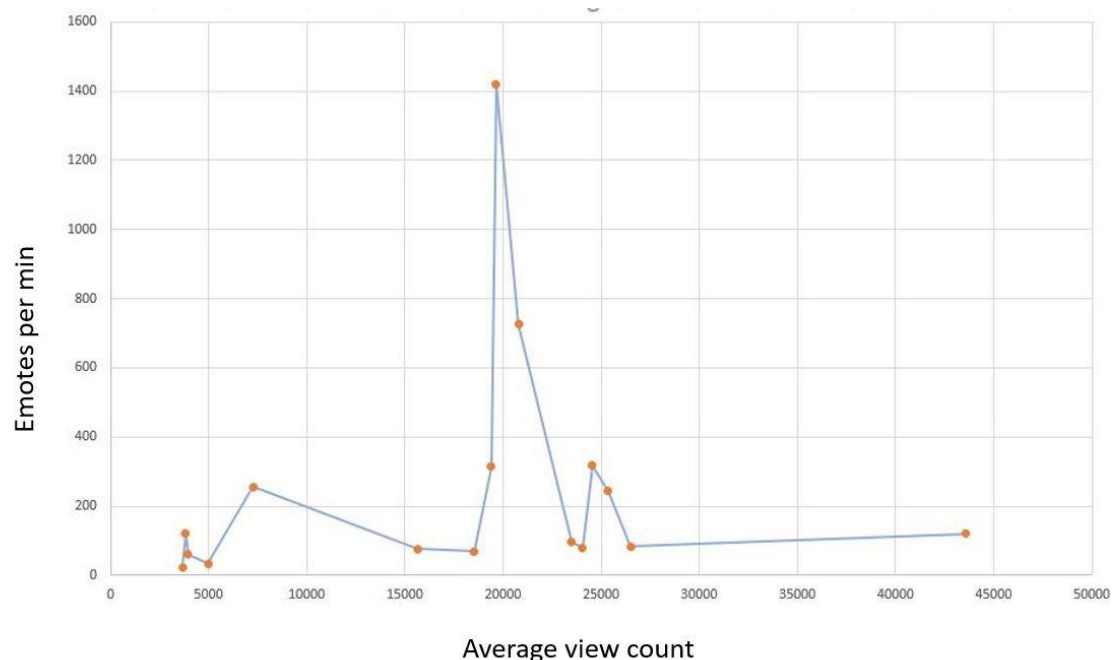


Figure 28 - Average view count/emotes per min graph with all streamers included.

The graph above shows how many emotes per minute are used in gathered vods with an approximate number of average viewers channels have. Channels from left to right: “GiantWaffle”, “AvoidingThePuddle”, “DansGaming”, “Destiny”, “AdmiralBulldog”, “CohhCarnage”, “Moonmoon_ow”, “Asmongold”, “XQCow”, “Forsen”, “AdmiralBahroo”, “TimTheTatMan”, “DrDisrespect”, “Lirik”, “Sodapoppin” and “Shroud” respectively. The graph shows that chat spam depends on the community more than the number of viewers. This can be seen as some channels with a low number of average viewers have a small amount of emotes used per minute as well as the channels with high numbers of viewers. Another example is

AdmiralBulldog's channel that has more emotes used per minute than channels with higher view count. There is a standard limit of messages each viewer can send in order to avoid unwanted spam. The limit is that the user can send the message and send exactly the same message only 30 seconds after the first one. This is because the majority of the time, spam consists of the same emote or phrase and emote viewers have to be highly active to increase the amount of emotes used per minute. (figure 13) There are longtime viewers who chat a lot and thus increase the amount of emotes used, however, with limitation of chat messages sent it shows that emotes high usage cannot be done a small number of active users.

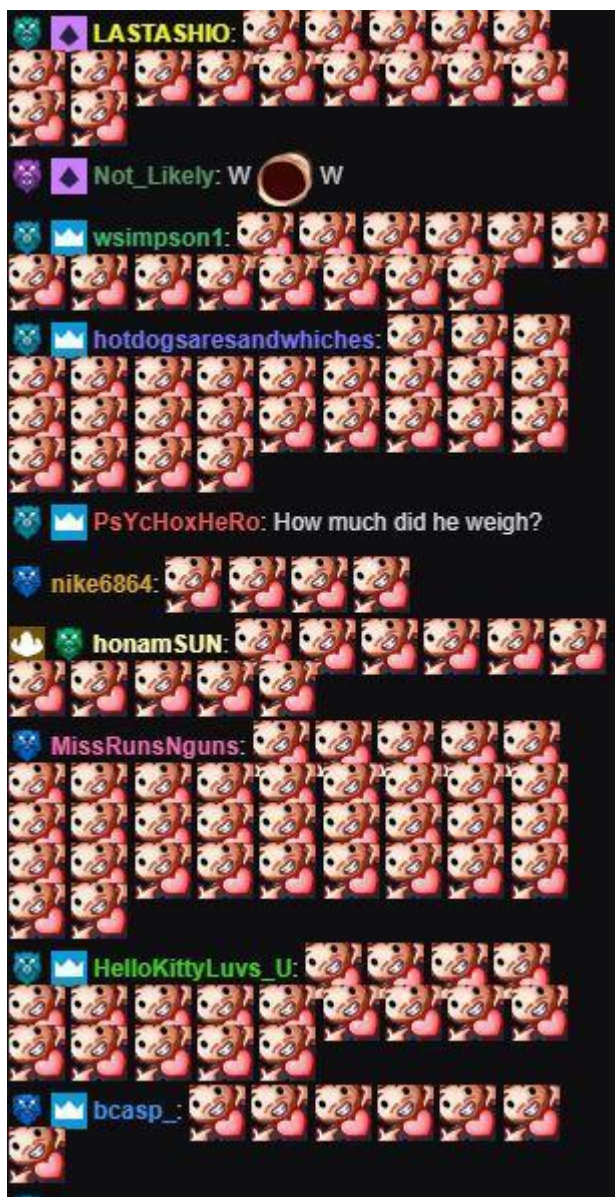


Figure 29 - Community together spam the same emote rather than one person being responsible for it.



Figure 30 - top 5 most active users of each channel.

Figure 14 represents the top 5 most active users of each channel and shows the differences in communities' participation with the streamer. As you can see from figure 14, Forsen's channel is one of the most spam active variety channels, with most active users having a colossal number of messages sent. On the other hand, channel TimTheTatMan has a low number of emotes used per minute even though the number of average viewers is higher than Forsens. This further supports the argument that it is the community of separate channels that support a high amount of spam messages rather than the average viewer count being responsible for it.

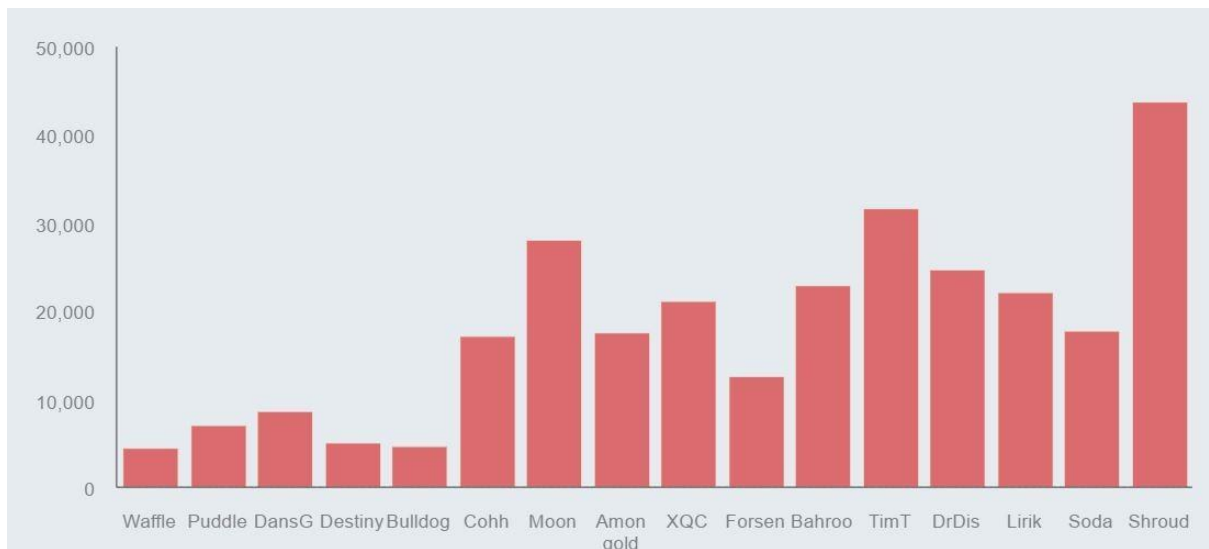


Figure 31 - Approximate number of subscribers each channel has.

The graph above represents the number of subscribers each channel has. There are examples of active subscriber viewer bases such as Shroud with 42,199 subscribers and average viewer count 43,574 and DrDisrespect with 28,307 subscribers and 24,536 average viewers. The most popular emotes in vods gathered in both of these channels were subscriber emotes and this graph also shows that it is not just a small part of the whole community that spam subscriber emotes but as the majority of both communities consist of subscribers.

Viewers that can be considered part of channels' community are those who participate in the chatroom or subscribe to the channel. The main feature of being a subscriber is the use of subscriber emotes which suggests that subscribers are also active chatroom users. The graph shows that the number of viewers does not affect how many emotes are used. This shows that viewers that are not subscribers and do not participate in chat are not part of the community of the channel as they will not affect emote statistics. However, the main effect of the number of subscribers is on which emotes will be most commonly used and the average amount of viewers. The active subscriber viewer base will usually use subscriber emotes in the majority of situations instead of global emotes. This separates the community as regular viewers cannot use subscriber emotes and if spam is made out of subscriber emotes it will not be used in other channels. As a result, channels with higher subscriber activity create spam patterns that are used only inside the channel and do not play a major role in creating global spam patterns. Channels that still have a high number of subscribers, where the most common emotes used are global emotes, are the ones that use global spams that can be adapted in any chatroom and does not require the user to have a connection to the channel from which spam originated. This draws the community closer together and they come up with spams that are used by everyone.

An example of a community with a high number of subscribers and average viewers, while most common emotes used in vods being global emotes, are Asmongold and XQCow. Both channels have a nearly identical average view count and a high number of subscribers compared to other channels, yet the most commonly used emotes in vods and clips are global emotes. This shows that their community is more driven to the usage of spams that can be used in any channel rather than spams that are popular locally in a single channel. On the other hand, DrDisrespect and Shroud are

streamers with most emotes used being subscriber emotes which means streams have higher subscriber activity in chat than non-subscribers. This can be seen from the vod gathered information as the majority of all emotes used are subscriber emotes. In conclusion, not all variety channels can be considered responsible for spam patterns as in some cases they have unique spam patterns that are only popular inside the channel itself even if the number of emotes used per minute is high. Channels with a smaller number of emotes used per minute cannot be considered responsible for creating chat patterns. Additionally, only some of the channels with active communities and statistics of global emotes being the most popular in the channel are responsible for spam patterns that will be used across the whole platform.

Summary

During my research, I made 2 hypotheses. First, that variety streamers are responsible for creating chat patterns and the second being that the a more spam filled chatroom is, the shorter their messages are. The second hypothesis means shorter in a sense that it consists of fewer words and more emotes as well as it has little to no meaning in it. The first hypothesis is hard to prove for certain that exactly these channels are responsible for all the chat patterns that are present on the platform, however, the majority of all non-subscriber emotes that are present in one channel, are also present in other channels. The number of times these emotes were used differ from channel to channel but they are all present. From further analysis, it can be seen that channels with a very active subscriber viewer base do not follow this particular pattern and instead concentrate more on the community of the channel. They use subscriber emotes more frequently but some of the global emotes are still present in these channels showing that there is no better alternative emote for viewers to use. These channels do not show a contribution to creating chat patterns but can be seen adapting existing ones. Channels like Shroud, DrDisrespect and CohnCarnage are the ones that, during my research, didn't show any unique patterns that would be used outside their own channel. However, the most common BTTV and FFZ emotes that they have are the same as in other channels, showing that although their community is mostly based around their individual channels, they still have popular patterns present. The second hypothesis

was easier to show by comparing all channels chatrooms. Calculating how many emotes were used per minute we can determine how spammy the chatroom will be. Comparing chatroom of GiantWaffle to XQCow chatroom it is easy to see how the interaction between chat streamer changes. Chatroom that has higher number of emotes used have smaller interaction between themselves and a stream so in order for the message to be seen it has to be presented in a form of a pattern that will be also used by the majority of people in the chat thus making chat spam the same message over and over so that the streamer would read it. This shows that the more creative chat patterns come from streams with hearing emote usage which also affects the length of the message that is sent. Chatrooms with higher amount of emotes have smaller text messages. Channels that are in the middle of the graph are sometimes more spam filled, other times there are more long text messages. These channels are most affected by the situation that is currently happening on stream. For example, in figure 15 AdmiralBahroos' chatroom is very spam filled with messages being sent at a fast pace and the majority of them being emotes only. On the other hand, figure 16 shows chatroom having a conversation with long messages being sent. This also shows that channels that are in the middle have the most flexible chatroom.

Spammy's chatroom is always spam filled and rarely will you see long text messages in them while channels with less active chatroom cannot be spam filled. Channels in the middle can adapt to both depending on the situation. The conclusion that can be made is that channels with spam filled chatrooms are the most responsible for creating new chat patterns that are later used in other channels since they are the channels that use the most amount of emotes. Channels that are both spam filled and can't have long text messages can also create new chat patterns but are less likely to do so and channels that don't use emotes as often as the other two groups will just adapt existing patterns into their chatrooms but will not create new ones.

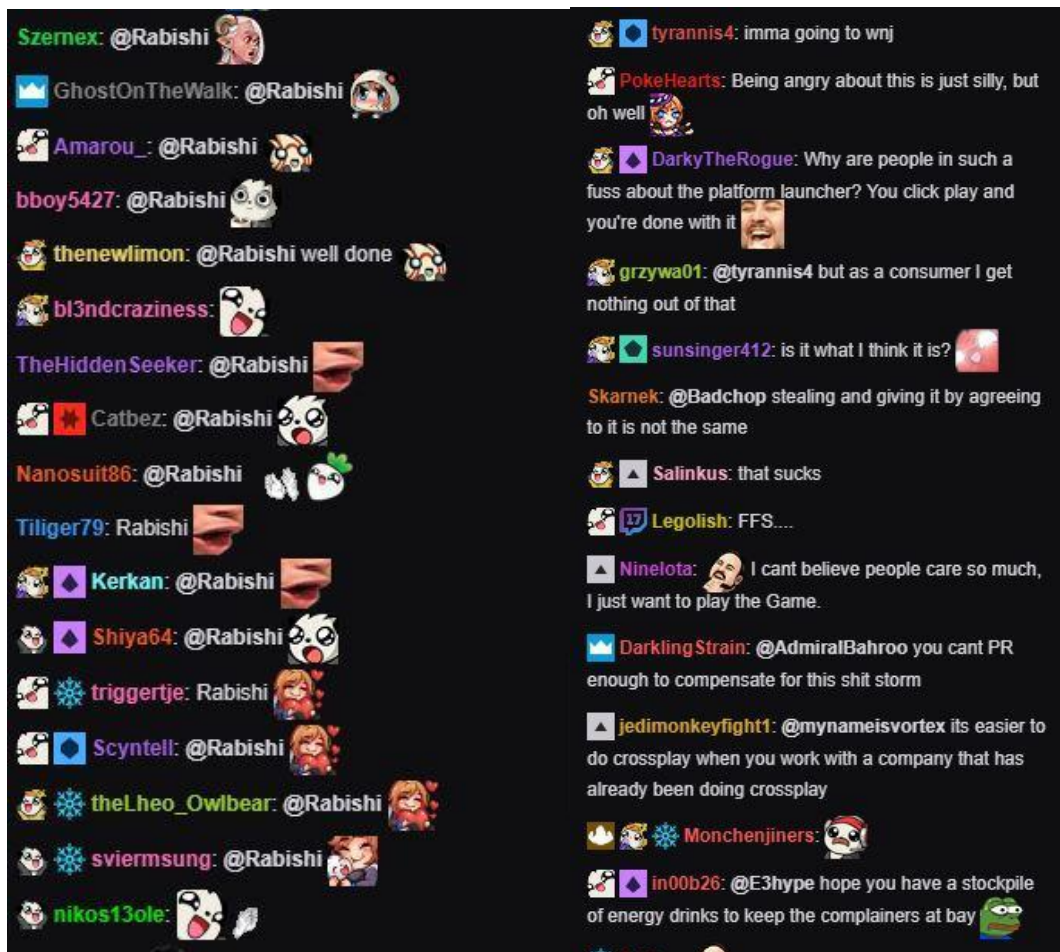


Figure 32 - Same chatroom being both spammy and not-spammy during the same stream.

6.0 Final Results

6.1 Challenges

6.1.0 Twitch API and coding

Before starting this project, I had very little knowledge of python and API. The first challenge for me was to learn how to use the Twitch API and register an application. The challenge with retrieving clips IDs was that some requests would not be accepted and an error would be returned. There was also an issue with information about clips as some clips would not contain detailed information regarding the clips compared to other ones. The other problem with requesting clips information is that it could be done only specific way and in case of an error, I had to reset git bash and retry the call. In some cases, it would require me to reset several times before the call was accepted.

6.1.1 Chatty application

I ran into problems with Chatty application several times during the data gathering. Chatty collects chat messages during the time that my account is connected to the chat room, however, text file where all the messages are saved only updates after my account disconnects from the chatroom. During the process of data gathering, Chatty crashed numerous times without saving the chat messages collected. This reset my progress back and I had to restart data gathering.

Section 6.2 Project Risks

This project has a few risks that must be taken into consideration. There are no risks regarding the software used or analysed, but the main risk comes from the information itself. Every twitch channel that is used in this project are relevant and broadcast consistently if not every day. This means that twitch clips collected earlier might not be the most popular one's next year.

Risks Faced

During the research I faced all the risks that I assumed would happen but as I progressed further in this project more problems arose. After the stream has ended the vod for it is stored for 2 weeks on the website and if the streamer himself did not save it, it would be automatically removed from the website. Furthermore, every streamer has a specific amount of vods that could be saved and when the new vod is saved the oldest one is removed. Some of the clips that were used in this project are still popular, however, since the vod that it is connected to is out-of-date, information regarding chat messages cannot be accessed. Due to this problem, some clips had to be replaced with a different one. Another problem faced during the project was Twitch policies regarding bots in chat. Twitch.tv has strict rules regarding the bots and any unauthorised bots in the chat room are automatically banned and disconnected from the chat. Due to this problem, I had postponed data analysis and concentrate on research of a different method to collect information from the chat. Lastly, I had to restart the clip data collection from the beginning as some of the clips were out-of-date. Since this problem is out of my control, I decided to connect my research to a specific date. All the information and data collected in this project is relevant by date of 8th of March 2019. There was also a risk that I encountered that affected this project. Some variety streamers that I decided to concentrate on did no stream or streamed very little thus I was not able to collect enough data to analyse. Due to this problem I had to choose different streamers as well as cut how many channels are going to be analysed. Original number of channels was 20 and the number was cut to 16.

Another problem I faced is regarding the privacy of each user. I used the authorised application that allows me to gather messages that were sent by different users. However, every single user did not give personal consent for their messages to be recorded. In order to keep each message anonymous, after chat data was gathered, I deleted every username from saving text files. The chatty application saves messages and usernames of who sent the message. Usernames were deleted from all files and in future, if students will use this report for a possible comparison of information, they will not have access to the data gathered as it will not be available publicly.

6.3 What I Would change

Chosen Twitch channels

I focused on variety channels and made analysis only of these channels. One of the main things I would change is that I would add more channels to data analysis adding channels that broadcast tournaments and channels that stream in one section (streamers that are known for streaming only one game). I would take several channels from each group and proceed to make the same analysis on each individual channel and a group overall. This would give results for individual streams that could be compared within its group and with other channels. Moreover, I would also concentrate on finding how many subscribers emotes of other channels were used across the whole platform and try to connect whether subscriber emotes that were used in other channels attract more people to the channel. I also would expand this project by analysing observing a new pattern appearing in the channel, trace it back to the channel where it was used originally and see how fast this pattern would spread and how many channels would start using it.

Furthermore, analysed patterns in this project are only emote patterns, however, there are known patterns that consist of both words and emotes. With the limit of time I had to do this project, it was impossible to add these patterns to data analysis. Two of “emote and word” spams that were noticed to be extremely popular are phrase “true LuW”(figure17) and using any word and replacing letter “o” with an emote “omegalul” (figure18) with extremely high popularity of both emotes across all channels it could be said with certainty that all channels most of the channels that are used in this project are regular users of these spams. However, in order to give solid evidence of all the possible patterns that are used in variety channels and which patterns are they responsible for, gathered data should be expanded further. Furthermore, the context in which emotes were used should be analysed as some of the emotes that are spammed could be used in the same way as spams with emote and words. More channels should be taken into consideration and analysis should be timed. For example, analysis should start when at least two channels add a new emotes that can be used by chatroom and data should be gathered from the first usage of the emote to the point when spam pattern appears. After a spam pattern is created, further data gathering should begin with monitoring other channels that will add the same emote to their channels and observe when will similar spam will be

used for the first time in the channel. User's account that used this spam first time in a different chatroom should also be analysed to find a possible connection between these channels and support both arguments that these channels share their communities and that channel which added the emote first was responsible for creating that pattern. Additionally, there are only text patterns called "Copy pastas" which are used only in very specific moments. The analysis could be done on what causes them to appear and how often.



Figure 33 - Spam patterns that consist of both emote and a word.

7.0 Conclusions

In this section, I will discuss the legacy of the project and how I have personally developed myself during the project.

Legacy of the Project

All the documents, code and links to applications used are available to the public via GitHub and any person interested in this research can access it. All the information is as stated relevant by the date 8th of March 2019 and even if in the future this information will be out-of-date it can be used to compare what twitch patterns were popular and see which patterns remained the same and which changed. This project leaves a solid foundation for further improvements of the research as well as information that can be used in order to program a more advanced chatbot.

This project concentrated on research and analytics of chat room patterns. However, I believe that it can be useful not only for further research. As a future addition, I came up with a concept for an advance chat room bot. There are many python's programmed bots available to use on Twitch Developer website but their limitation is that they are only programmed to perform a specific task. One of such bots responds to a specific emote when somebody uses it. I thought that a possible improvement could be done to such bot by implementing machine learning. Chatbot would gather chat messages and analyse possible patterns that are used in chat or a new one created. It would also keep track of how many times these patterns were typed allowing researchers to save time searching for patterns as all popular ones are already recorded. Furthermore, if this bot is used by several different channels it can provide comparative analysis and possibly detect when was the first usage of the particular spam pattern and in which channel it appeared. I do not have knowledge and experience of machine learning to program it myself but I believe that it potentially could be a great project idea.

Section X.X – References

1. Twitch Developers. (2019). *Twitch API v5*. [online] Available at: <https://dev.twitch.tv/docs/v5> [Accessed 3 Apr. 2019].
<https://dev.twitch.tv/docs/v5/reference/clips/> [accessed 13/09/2018]
2. Twitch Developers. (2019). *Chatbots & IRC*. [online] Available at: <https://dev.twitch.tv/docs/irc/> [Accessed 3 Apr. 2019].
3. Twitch. (2019). *Twitch*. [online] Available at: <https://www.twitch.tv> [Accessed 3 Apr. 2019].
4. Twitchmetrics.net. (2019). *The Most Popular English Variety Twitch Streamers, April 2019*. [online] Available at: <https://www.twitchmetrics.net/channels/popularity?game=Variety&lang=en&page=2> [Accessed 3 Apr. 2019].
5. StreamElements. (2019). *Chat stats*. [online] Available at: <https://stats.streamelements.com/> [Accessed 3 Apr. 2019].
6. Ford, C., Gardner, D., Horgan, L., Liu, C., Nardi, B. and Rickman, J. (2017). *Chat Speed OP PogChamp: Practices of Coherence in Massive Twitch Chat*. [online] Denver, Colorado, USA: ACM New York, pp.859-863, 865-867. Available at: <https://dl.acm.org/citation.cfm?id=3052765> [Accessed 3 Apr. 2019].
7. Community, T. (2019). *Twitch Emotes - Bringing a little Kappa to you everyday*. [online] Twitchemotes.com. Available at: <https://twitchemotes.com/> [Accessed 3 Apr. 2019].
8. Stephenson, B. (2018). *Twitch: Everything You Need to Know*. [online] Lifewire. Available at: <https://www.lifewire.com/what-is-twitch-4143337> [Accessed 3 Apr. 2019].
9. TwitchTracker. (2019). *Twitch Subs Count Statistics*. [online] Available at: <https://twitchtracker.com/subscribers> [Accessed 3 Apr. 2019].

Table of figures

Figure 1 - Chatroom reaction to a joke that was told by the streamer (Lirik)	9
Figure 2 - Example of how Followed Channels bar looks like	11
Figure 3 - Trending channels on the main page of Twitch.....	11
Figure 4 - Featured sections with the most viewers online	12
Figure 5 - An example of returned information about tiwtch clips from the website..	17
Figure 6 - Most used emotes across the whole website.....	17
Figure 7 - Most popular emotes used in a single channel	18
Figure 8 - Code for sorting text file to separate words from emotes	24
Figure 9 - CohhCarnage's graphs of top emotes used in vods.....	26
Figure 10 - CohhCarnage's graphs of top emotes in clips.....	26
Figure 11 - DrDisrespect's channel graph of top emotes used in vods	27
Figure 12 - DrDisrespect's channel graph of top emotes used in clips.....	28
Figure 13 - Forsen's channel graph of top emotes used in vods.....	29
Figure 14 - Forsen's channel graph of top emotes used in clips.	29
Figure 15 - Shroud's channel graph of top emotes used in vods.....	30
Figure 16 - Shroud's channel graph of top emotes used in clips.....	31
Figure 17 - Sodapoppin's channel graph of top emotes used in vods.....	32
Figure 18 - Sodapoppin's channel graph of top emotes used in clips.	32
Figure 19 - XQCow's channel graph of top emotes used in vods.....	33
Figure 20 - XQCow's channel graph of top emotes used in clips.	34
Figure 21 - Globally most used emotes of Twitch and extension BTTV.	35
Figure 22 - Most used emotes of the extension FFZ.....	36
Figure 23 - Graph of cases when emote "LUL" was the most popular emote.	37
Figure 24 - Graph of how many emotes were used in vods per minutes.....	39
Figure 25 - Examples of a non-spammy and a spammy chats.....	40
Figure 26 - Chatroom repeats the message after streamer.....	41
Figure 27 - Chatroom shows that they enjoy the song played by the streamer.	42
Figure 28 - Average view count/emotes per min graph with all streamers included.	43
Figure 29 - Community together spam the same emote rather than one person being responsible for it.....	44
Figure 30 - top 5 most active users of each channel.....	45
Figure 31 - Approximate number of subscribers each channel has.	45
Figure 32 - Same chatroom being both spammy and not-spammy during the same stream.	49
Figure 33 - Spam patterns that consist of both emote and a word.	53

Appendix

Revised time plan

Phase	Task Name	Start Date	End Date	Notes
1	Interim Report and Risk Assessment	29/10/18	07/12/18	
1.1	Identify requirements	22/10/18	22/10/18	
1.2	Write Project specification	21/10/18	01/12/18	
1.3	Improve Project specification based on feedback	05/10/18	05/11/2018	Original Project specification was submitted before getting feedback and was corrected after the deadline.
1.5	Start working on the interim report	24/11/18	07/12/18	I began working on the interim report later than I originally planned due to constant bot crashes which I ended up spending more time on.
1.6	Identify risks and begin the risk assessment	24/11/18	07/12/18	Before starting writing interim report risks assessment was done.
1.7	Finalise the interim report and risk assessment	06/12/18	07/12/18	Last check of the report before submission
2	Coding and Design	20/10/18	05/04/19	The coding part of the project began earlier due to the simplicity of the commands used at the beginning. This gives me an opportunity to concentrate more time on coding the twitch bot rather than gathering data.
2.1	Researching on twitch API, curl, git bash and twitch chatbot by watching/reading tutorials	23/10/18	06/11/19	

2.2	Draft data analysis	07/11/18	05/03/19	This is step is the first part of data analysis that is done from the beginning to give some idea of what to expect from a fully done data analysis that would be completed.
2.3	Start adding bits of data into the database	07/12/18	31/12/18	
2.4	Start sorting the database	25/01/19	25/02/19	This step has been postponed since chat bot constantly crashes and doesn't collect enough information.
2.5	Add more channels to the database if necessary (this step can be skipped)	02/02/19	02/03/19	
3	Begin analysis	05/03/19	29/03/19	
3.1	Start creating graphs and documents with the result of the analysis	05/03/19	07/03/19	
3.2	Verify that the information used is up to date	08/03/19	09/03/19	
3.3	Make minor changes in case some of the information is no longer relevant	09/03/19	16/03/19	
3.4	Verify that the project result followed the requirements that fulfil all of them	16/03/19	25/03/19	
3.5	Final corrections	26/03/19	27/03/19	
4	Final Report	27/11/18	23/04/19	
4.1	Log details throughout the creation of the project.	05/11/18	18/01/19	

4.2	Work on mistakes from interim report feedback	19/01/19	22/01/19	
4.3	Add information on additional hardware used after an interim report	23/01/19	27/02/19	
4.4	Add new references used, write the conclusion of the project	16/03/19	10/04/19	
4.5	Add information on possible future additions to the project	01/04/19	10/04/19	
4.6	Finalise the report	10/04/19	23/04/19	
5	Demonstration Day	24/04/19	24/04/19	
6	Viva	24/04/19	10/05/19	

Risk Assessment

Risk 1 – Not abiding by the time plan

Description of Risk	Description of impact	Likelihood rating	Impact rating	Preventative Actions
Fall behind on schedule	Work is done at the last minute	2	5	Have a detailed time plan that has to be followed thoroughly

Risk 2 – Project does not follow project specification

Description of Risk	Description of impact	Likelihood rating	Impact rating	Preventative Actions
Not following the specification	Project data collected is not relevant or incorrect	3	5	Information collected must be checked for its correctness

Risk 3 – Chatbot failures

Description of Risk	Description of impact	Likelihood rating	Impact rating	Preventative Actions
Chatbot stops working	Chatbot does not work. As a result, more time is spent fixing the bot than collecting or analysing data	3	5	Test the bot several times for its work and use online forums to get help to fix it faster in case of a crash.

Risk 4 – Unavoidable events

Description of Risk	Description of impact	Likelihood rating	Impact rating	Preventative Actions
Channels broadcast live and new clips appear that becomes more popular than the ones already collected	This will impact the time spend on getting new information regarding the clips and updating the database with new clips as well as removing the old ones	4	5	The primary way of avoiding this is to validate that the information collected is up to date before starting an analysis on it

Risk 5 – Unavoidable events

Description of Risk	Description of impact	Likelihood rating	Impact rating	Preventative Actions
Website/ Software updates	Software used in the project has a new version which might crash the existing database. Twitch update their policies and update the website making it impossible to use a chatbot or collect clip information.	1	4	Software used in the project has to be of the latest version available or version used must be stable for the whole duration of the project.