Analyzing Pokémon and Mario Streamers' Twitch Chat with LLM-based User Embeddings

Mika Hämäläinen[1], Jack Rueter[2] and Khalid Alnajjar[3]
[1]Metropolia University of Applied Sciences
[2]University of Helsinki
[3]F-Secure Oyj
`first.last@metropolia.fi/helsinki.fi/fsecure.com`

Abstract

We present a novel digital humanities method for representing our Twitch chatters as user embeddings created by a large language model (LLM). We cluster these embeddings automatically using affinity propagation and further narrow this clustering down through manual analysis. We analyze the chat of one stream by each Twitch streamer: SmallAnt, DougDoug and PointCrow. Our findings suggest that each streamer has their own type of chatters, however two categories emerge for all of the streamers: *supportive viewers* and *emoji and reaction senders*. *Repetitive message spammers* is a shared chatter category for two of the streamers.

1 Introduction

Streamers playing video games online have gained popularity over the past years (see Alvarez 2024). Platforms like Twitch[1] and YouTube[2] have enabled gamers to build massive audiences, with top streamers attracting millions of followers who tune in regularly to watch live gameplay and to engage in real-time chat (see Fernández-Robin et al., 2024).

This rise of live-streaming platforms has introduced a novel arena for examining the dynamics of online interaction and community formation. The communicative practices observed in these environments are rich with potential for analysis, offering insights into how digital communities emerge, how chatter identities are performed and negotiated in virtual spaces, and how technology shapes discourse and social interaction (see Speed et al. 2023).

Real-time chatters provide interesting data for research purposes. In this study, we aim to better understand what categories of chatters are there. We do this by building user embeddings for each

chatter using a large language model (LLM) and later on cluster these embeddings into chatter categories. By harnessing the capabilities of large language models (LLMs), we aim to construct a nuanced representation of individual chatters, encapsulating the multifaceted dimensions of their discourse.

We study two Pokémon streams, one by SmallAnt and one by PointCrow, and one Mario stream by DougDoug. We aim to see whether there are any mutually shared chatter categories, and if any streamers has their own unique chatter categories. Through this, we aspire to contribute to a deeper understanding of the sociocultural fabric of online gaming communities by shedding light on the complex interplay between content creators and their audiences in a rapidly evolving digital landscape.

Furthermore, we present a novel embeddings and clustering driven data-analysis method that is applicable as is in almost any digital humanities dataset. The code has been made available on Zenodo[3].

2 Related work

Twitch streams have been an object of study before our research as well. There is also a great body of literature on computational sociolinguistics (Säily et al., 2021; Tiihonen et al., 2023; Landert et al., 2023). However, in this section we focus on some of the prior literature relating to studying Twitch streamers.

In the context of online live streaming, prior research (Recktenwald, 2017) has explored the intricate dynamics between broadcasters and audiences, focusing on the interaction that occurs through both spoken language and written chat during live broadcasts of video games. The research identifies a novel communicative behavior termed 'pivoting,' where both broadcasters and audiences produce

---

[1]https://twitch.tv
[2]https://youtube.com
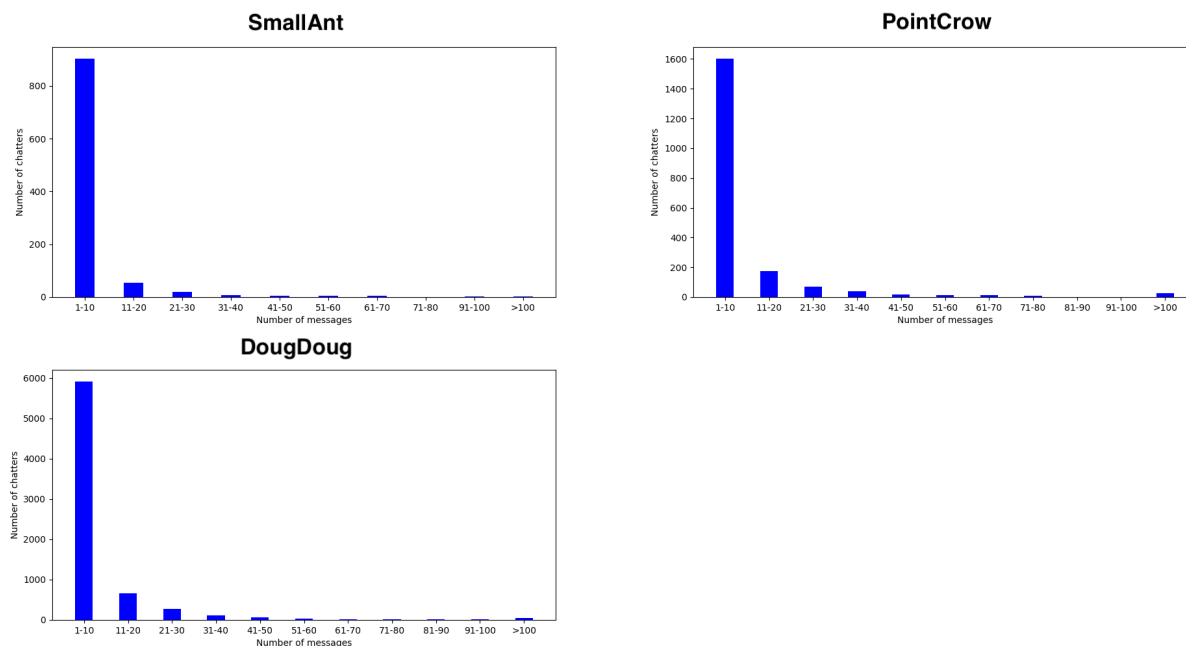[3]https://zenodo.org/records/13886601

Figure 1: Chatter engagement for each stream

context-dependent utterances in response to game events, demonstrating the highly interactive and situated nature of discourse in live streaming environments.

A mixed-methods study (Diwanji et al., 2020) explored the information behavior and perceptions of co-presence among Twitch users, leveraging human information theory and social identity theory as the theoretical framework. Through quantitative analysis using tools like LIWC and SPSS, alongside qualitative thematic analysis with Nvivo, the study identified that information reaction and production were prevalent behaviors across multiple streams, while also highlighting the strong sense of co-presence experienced by participants.

The study by Cabeza-Ramírez et al. (2022) investigates the complex interplay between video game usage, live streaming and the potential adverse effects associated with excessive engagement in these activities. Employing a multilayer perceptron model on a substantial sample of 970 video game users, the research identifies key factors influencing gaming behavior. Specifically, the study highlights that motivations related to a sense of belonging to gaming platforms and the positive social uses, such as making friends and pursuing gaming as a profession, significantly contribute to the risk of pathological gaming.

Building on the foundation of affordance theory, a recent study (Sjöblom et al., 2019) examined the practices of the most popular streamers, revealing

how these individuals leverage various communication modalities and social commerce elements to create compelling content. Their work helps in uncovering the emerging trends and common strategies employed by streamers, offering insights into the evolving landscape of online video streaming as a business model driven by individual content creators.

3 Twitch chat data

We study three streams: *HARDEST NUZLOCKE (interrupted by a board game w/ magicthenoah, failboat, captainkidd)*[4] by SmallAnt, *POKEMON FUSION JOHTO LEAGUE 150,000+ NEW FUSIONS | !tts on*[5] by PointCrow and *Can I beat 10 different Mario 64 speedruns simultaneously? !suck*[6] by DougDoug.

Obtaining chat data on Twitch is rather difficult programmatically. We build a small Python script that launches Chrome through Selenium[7]. This script monitors the chatbox and saves all chat messages to a file. This requires us to actually watch through the streams to collect the data. The additional benefit is that we do not need to write a crawler that might violate the terms of service of Twitch - this way we simply log out all messages that appear during the streams naturally.

---

[4]https://www.twitch.tv/videos/2220045551
[5]https://www.twitch.tv/videos/2211596823
[6]https://www.twitch.tv/videos/2217569664
[7]https://www.selenium.dev/

	SmallAnt	PointCrow	DougDoug
Messages	5088	20488	59207
Chatters	1001	1965	7182
Length	4h 31min	7h 22min	5h 26min

Table 1: Size of each stream dataset

Table 1 shows the key numbers of the dataset size for each streamer. SmallAnt had the lowest viewer engagement out of all the three streamers. DougDoug had the highest number of chatters in his stream.

SmallAnt was playing a Pokémon nuzlocke in his stream where he needed to do some calculations on Pokémon stats. The stream featured a section where he was playing a virtual board game with other streamers. PointCrow's stream was about continuing his playthrough of Pokémon Infinite Fusion game. During the stream, he took a break to eat dinner and watch online videos. DougDoug streamed Super Mario 64. The stream included a section of live coding and a message from his sponsor.

Figure 1 shows how active chatters were. Chatters are grouped in categories based on how many messages they sent. As we can see, most of the chatters on all streams sent only 1-10 messages during the entire stream.

4 User embeddings and clustering

In order to create a user embedding for each chatter, we concatenate the chat messages for each user and separate each message with a line break. To ensure that we get enough text to represent each chatter well enough, we exclude all chatters who have sent less than 20 messages. If a chatter has not sent enough messages, not enough can be known about them to support the further analysis.

We use the LLM called PaLM-2 (Anil et al., 2023) over Google's VertexAI[8] to embed each chatter's concatenated messages into a user embedding. These user embeddings serve as mathematical representation of the semantics of what each user was chatting about. In particular, we use *text-embedding-004* model for *SEMANTIC_SIMILARITY* task. We picked this particular task for the embedder because we intend to compare the semantic similarity of each embedding.

We cluster the user embeddings separately for each stream. For clustering, we use affinity propa-

gation (Frey and Dueck, 2007). It takes in an affinity matrix, which shows how close each embedding is to other embeddings, and it will automatically find an optimal number of clusters based on the affinities provided to the algorithm. We use cosine similarity to populate the affinity matrix. We use the methods provided in Scikit-learn (Pedregosa et al., 2011) for affinity propagation and cosine similarity. Similar clustering approaches have previously been used with word embeddings (Hämäläinen and Alnajjar, 2019; Stekel et al., 2022).

Some chatters end up clustered into their own clusters. We remove all clusters that have only one chatter, because we are more interested in the overall tendencies of chatter categories, not in individual deviant chatters.

5 Results

The clustering algorithm created 6 clusters for SmallAnt, 12 clusters for PointCrow and 31 clusters for DougDoug. On a closer inspection, we found that some of the clusters included mutually similar messages, so we proceeded to merge some clusters manually. This resulted in 5 cluster for SmallAnt, 4 for PointCrow and 6 for DougDoug.

Cluster name	Size	Characteristics
Supportive viewers	8	The chatters are engaged with the stream in a laid-back fashion. The messages are generally positive.
PartyKirby spammers	7	The chatters were mainly spamming PartyKirby emote
Emojis and reactions	5	The chatters mainly send emojis and short reactions such as "oh no" or "damn".
Strategic helpers	11	The chatters sent helpful messages to help the streamer plan their Pokémon party better.
Anime and gaming enthusiasts	15	These chatters were talking about anime shows and other games as well and not just what SmallAnt was playing.

Table 2: SmallAnt chatter clusters

Table 2 shows the chatter clusters for SmallAnt along with a short description that characterizes the chatters in this category in general terms. The largest category is *Anime and gaming enthusiasts*; this is the only cluster that was merged with a similar cluster that specialized in conversation about a game called *Guilty Gear*.

The results of our analysis on PointCrow's stream can be seen in Table 3. All of the clusters required merging except for *Newcomers* cluster. The largest clusters are people who are excited about the game being played, *Pokémon Infinite Fusion*, and *Supportive viewers*.

Cluster name	Size	Characteristics
Supportive viewers	70	The chatters send supportive messages to the streamer. The messages have emotive content and they may also be positive reactions to the stream.
Pokémon Infinite Fusion enthusiasts	111	The chatters talk about the game being played and react to different Pokémon fusions with anticipation.
Emojis and reactions	14	The chatters engage with the stream either by sending emojis or by sending different kinds of short reactions such as "let's go" or "LMAO"
Newcomers	6	These chatters are newcomers to the stream and are watching PointCrow for the first time. Their messages are generally positive

Table 3: PointCrow chatter clusters

Cluster name	Size	Characteristics
Emojis and reactions	114	The chatters use emojis and send short reactions that also frequently express confusion such as "huh?" or "D:"
Meta-level discussion	76	The chatters talk about meta-level things relating to DougDoug and his channel such as editors, subscriptions and sponsors.
Critical viewers	188	The chatters criticise DougDoug's Mario skills. Some of them send !suck command and tell him that the game is rigged.
Supportive viewers	185	The chatters express their support and how much they like DougDoug. Some of the messages may appear negative in tone, but have a clear positive intention.
Parrots	13	These chatters repeat messages with a similar content over and over again. Such as comments about poggies and cheese stream.
Random reactions	76	These chatters send plenty of emotive reactions that may be on a variety of different topics.

Table 4: DougDoug chatter clusters

Table 4 shows the clusters of DougDoug chatters. *Parrots* is the only cluster that did not require manual merging. The largest number of merged clusters are in *Critical Viewers* (11 clusters) and in *Supportive viewers* (9 clusters). Interestingly DougDoug has many chatters that send negative comments. In fact, there is an active Reddit r/wehatedougdoug[9], which indicates that hatered is a valid way of showing DougDoug fandom. Perhaps this polarity between the supporters and supposed haters is the reason why DougDoug has the highest number of active chatters out of all the streamers.

If we look at all the chatter cluster for all the streamers, we can see that a number of chatter groups emerge for all the streamers: *Supportive viewers* and *Emojis and reactions*. In addition,

SmallAnt and DougDoug have a cluster of people who spam repetitive messages (such as PartyKirby).

6 Conclusions

In conclusion, our study introduces a novel approach within digital humanities by utilizing a large language model (LLM) to create user embeddings for representing Twitch chatters. By employing affinity propagation for automatic clustering and refining the results through manual analysis, we were able to effectively categorize chat participants from streams by SmallAnt, DougDoug, and PointCrow.

The method is generic enough to be used with any kinds of documents. In our study, we found that a bit of manual merging of the clusters was needed. This can be automatized in the future by running multiple iterations of the clustering algorithm by clustering clusters. This could be achieved easily by calculating a centroid embedding for each cluster and recalculating the affinity matrix based on the cluster centroids.

Our analysis revealed that, while each streamer attracts a distinct type of chatters, there are common categories across the streams. Notably, all three streamers share categories of supportive viewers and emoji and reaction senders. Additionally, a category of repetitive message spammers was found to be common among two of the streamers. These insights highlight the potential of LLM-based embeddings for understanding and categorizing social interactions in digital environments.

In the future, it might be interesting to gather more chat messages on multiple streams of the same streamer to assess the stability of the chatter categories across different streams. Chat messages also include timestamps, which we did not take into consideration in this study. It might also be interesting to include the content of the actual stream in the study, as our current study was only limited to the chat messages.

7 Limitations

The sample size for this paper is relatively small as it only covers one stream from three streamers. This means that the findings will not necessarily hold for every Twitch streamer or other streams by the three streamers that were studied.

---

[9]https://www.reddit.com/r/wehatedougdoug/

References

Nicolás Alvarez. 2024. The twitch boom: The rise of the plataform due to the impact of the covid-19 pandemic. *Revista Científica" Visión de Futuro"*, 28(2):59–77.

Rohan Anil, Andrew M Dai, Orhan Firat, Melvin Johnson, Dmitry Lepikhin, Alexandre Passos, Siamak Shakeri, Emanuel Taropa, Paige Bailey, Zhifeng Chen, et al. 2023. Palm 2 technical report. *arXiv preprint arXiv:2305.10403*.

L Javier Cabeza-Ramírez, Francisco José Rey-Carmona, Mª del Carmen Cano-Vicente, and Miguel Ángel Solano-Sánchez. 2022. Analysis of the coexistence of gaming and viewing activities in twitch users and their relationship with pathological gaming: a multilayer perceptron approach. *Scientific reports*, 12(1):7904.

Vaibhav Diwanji, Abigail Reed, Arienne Ferchaud, Jonmichael Seibert, Victoria Weinbrecht, and Nicholas Sellers. 2020. Don't just watch, join in: Exploring information behavior and copresence on twitch. *Computers in Human Behavior*, 105:106221.

Cristóbal Fernández-Robin, Diego Yáñez, and Joaquina Miranda. 2024. Twitch consumer behavior and the importance of streaming habits. In *International Conference on Human-Computer Interaction*, pages 234–246. Springer.

Brendan J Frey and Delbert Dueck. 2007. Clustering by passing messages between data points. *science*, 315(5814):972–976.

Mika Hämäläinen and Khalid Alnajjar. 2019. Let's face it. finnish poetry generation with aesthetics and framing. In *Proceedings of the 12th International Conference on Natural Language Generation*, pages 290–300.

Daniela Landert, Tanja Säily, and Mika Hämäläinen. 2023. Tv series as disseminators of emerging vocabulary: Non-codified expressions in the tv corpus. *ICAME journal*, 47(1):63–79.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830.

Daniel Recktenwald. 2017. Toward a transcription and analysis of live streaming on twitch. *Journal of Pragmatics*, 115:68–81.

Tanja Säily, Eetu Mäkelä, and Mika Hämäläinen. 2021. From plenipotentiary to puddingless: Users and uses of new words in early english letters. In *Multilingual Facilitation*, pages 153–169. University of Helsinki.

Max Sjöblom, Maria Törhönen, Juho Hamari, and Joseph Macey. 2019. The ingredients of twitch streaming: Affordances of game streams. *Computers in Human Behavior*, 92:20–28.

Abbie Speed, Alycia Burnett, and Tom Robinson II. 2023. Beyond the game: Understanding why people enjoy viewing twitch. *Entertainment Computing*, 45:100545.

Moshe Stekel, Amos Azaria, and Shai Gordin. 2022. Word sense induction with attentive context clustering. *Journal of Data Mining & Digital Humanities*.

Iiro Tiihonen, Aatu Liimatta, Lidia Pivovarova, Tanja Säily, and Mikko Tolonen. 2023. Measuring the distribution of hume's scotticisms in the ecco collection. In *Proceedings of the Joint 3rd International Conference on Natural Language Processing for Digital Humanities and 8th International Workshop on Computational Linguistics for Uralic Languages*, pages 36–44.