

WSA project

Titans of Discourse

Alessandro Ghiotto - 513944

Alessandro Montini - 517344

Andrea Pan - 514555

Introduction

Attack on Titan

Attack on Titan (AoT) is Japanese **manga and anime series** created by Hajime Isayama. It is set in a dystopian world where humanity is on the brink of extinction due to giant humanoid creatures called Titans, the story follows protagonist **Eren Yeager** and his companions as they uncover the secrets behind the Titans, the walls protecting humanity, and the truth about their world.

Motivation

Attack on Titan sparked intense discussions across Reddit, particularly surrounding its **controversial ending**, making it an ideal case study for identifying and analyzing **distinct online communities**.



Roadmap

- 1 Data Gathering and Analysis
- 2 Topological Analysis
- 3 Community Detection
- 4 Sentiment Analysis
- 5 Topic Modeling
- 6 Named Entity Recognition

1. Data Gathering and Analysis

Our Data

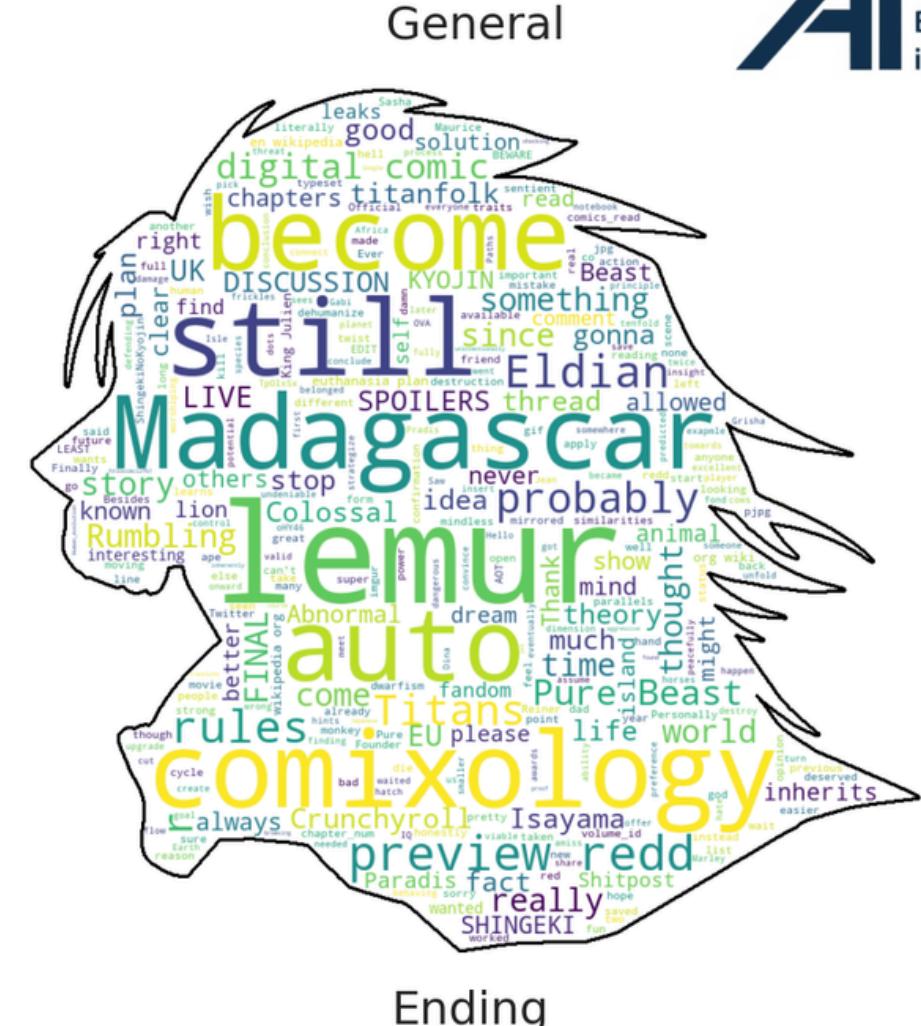
100 submissions from 3 different subreddits:

- r/attackontitan
 - r/titanfolk
 - r/AttackOnRetards

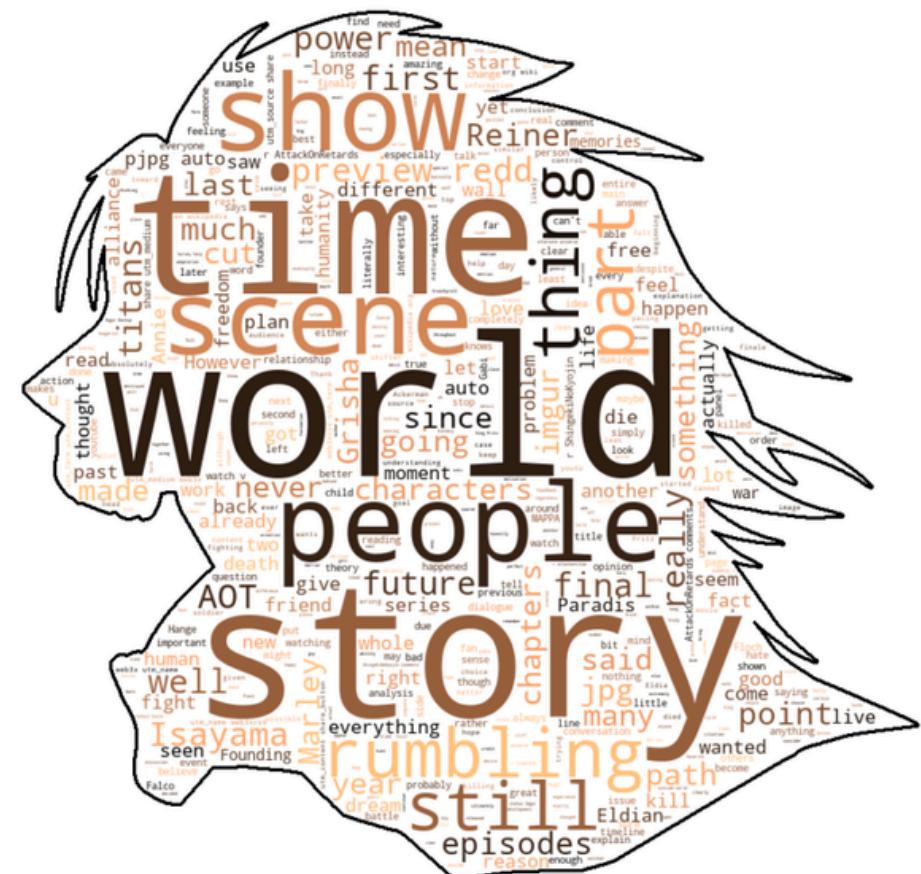
Two datasets:

- **General Posts:** top posts from each subreddit.
 - **Ending-related Posts:** top posts from each subreddit containing "ending," "finale," "final chapter," or "final episode."

Here we have the **WordCloud** for this two datasets. We added to the basic list of **stopwords** also the name of the characters and other domain specific words such as “anime”, “manga”, “episode”...



Ending

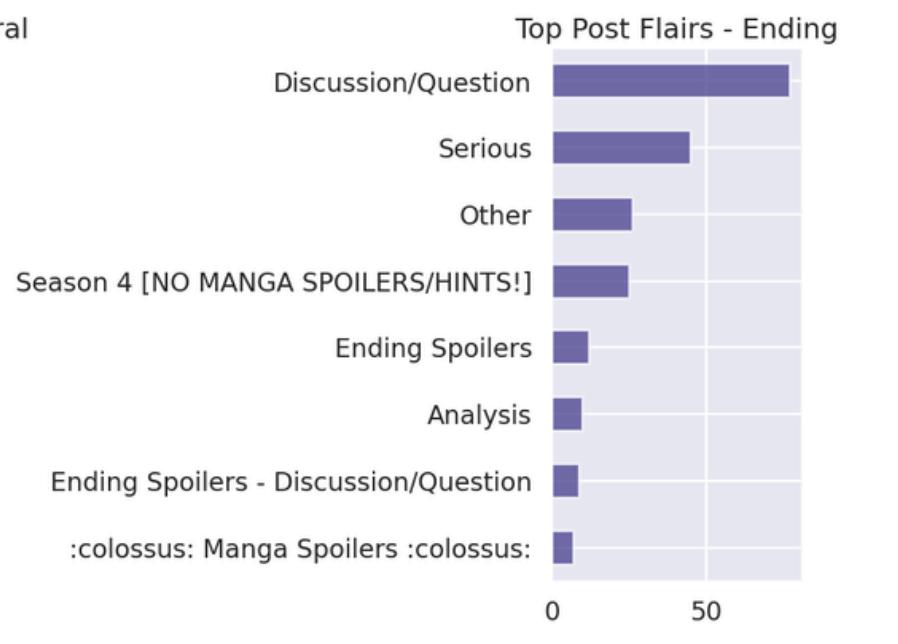
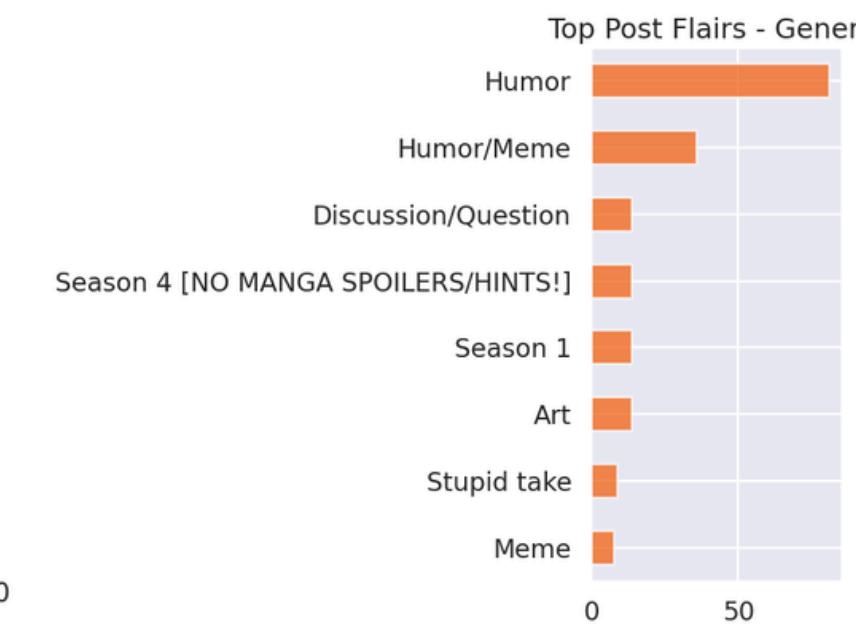
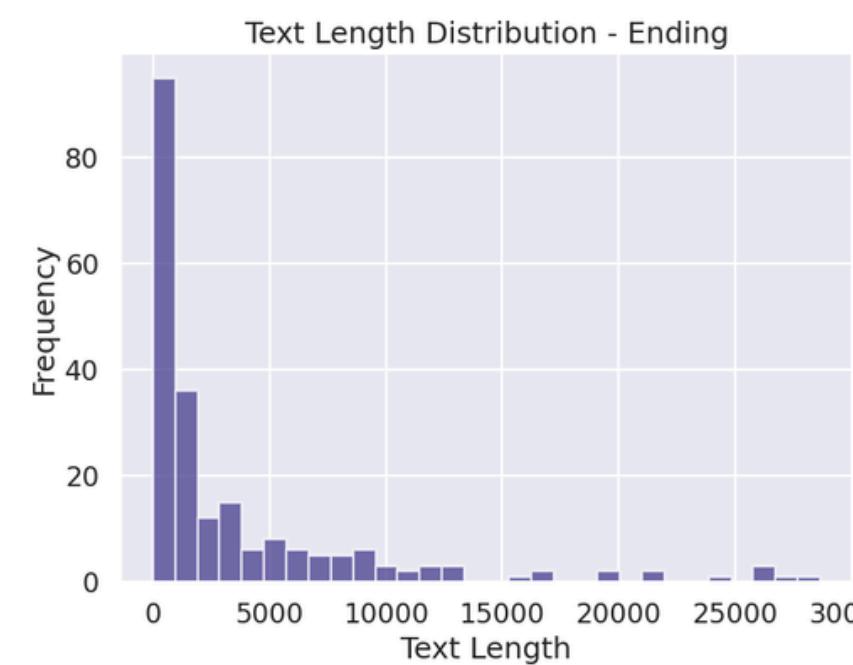
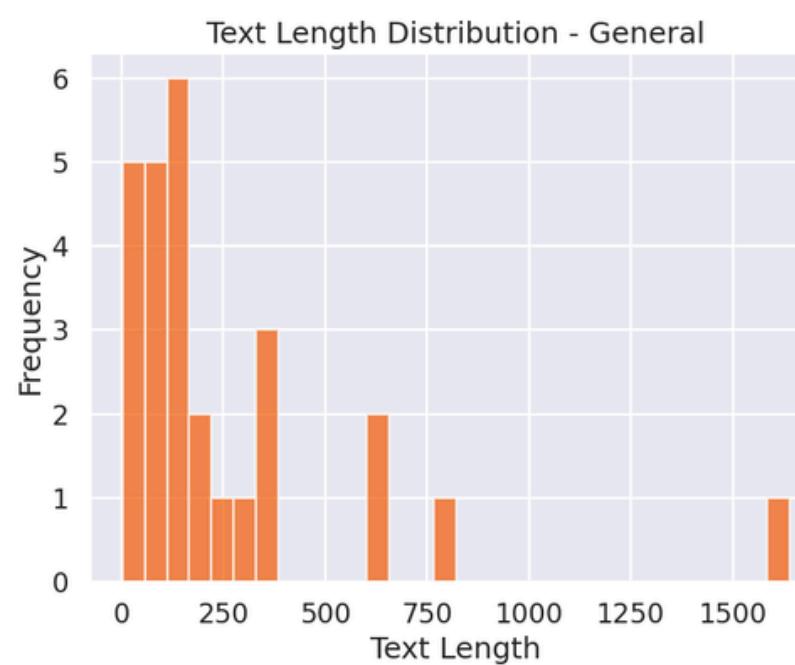
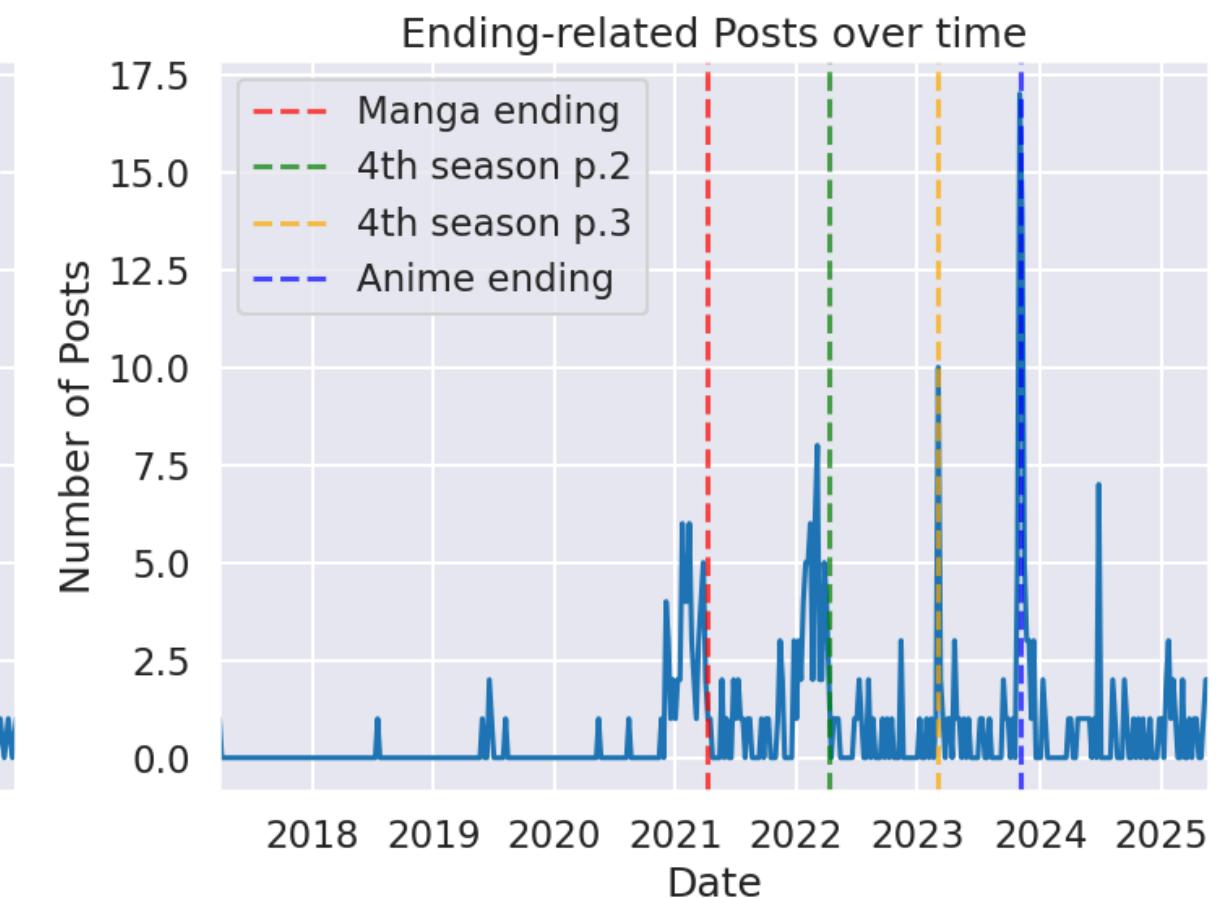
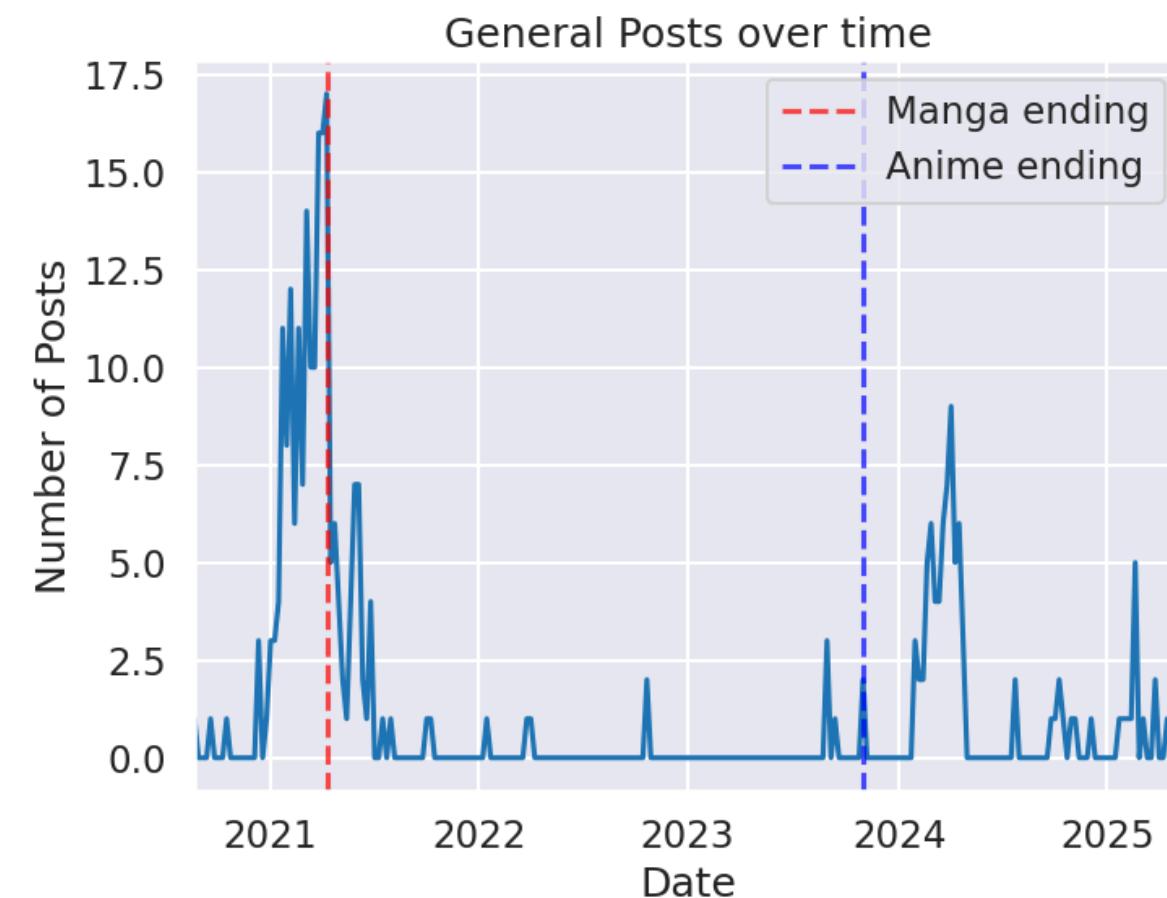


going under a landing moment - characters

Further analysis...

- The number of posts tends to peak at the end of the anime seasons.
- Ending related posts are in general much longer.
- General posts:
→ **humor** and memes
- Ending posts:
→ **discussion** and questions

Next we are going to focus on Ending posts only!



2. Topological Analysis

Build the graph of replies

$$G = (N, E)$$

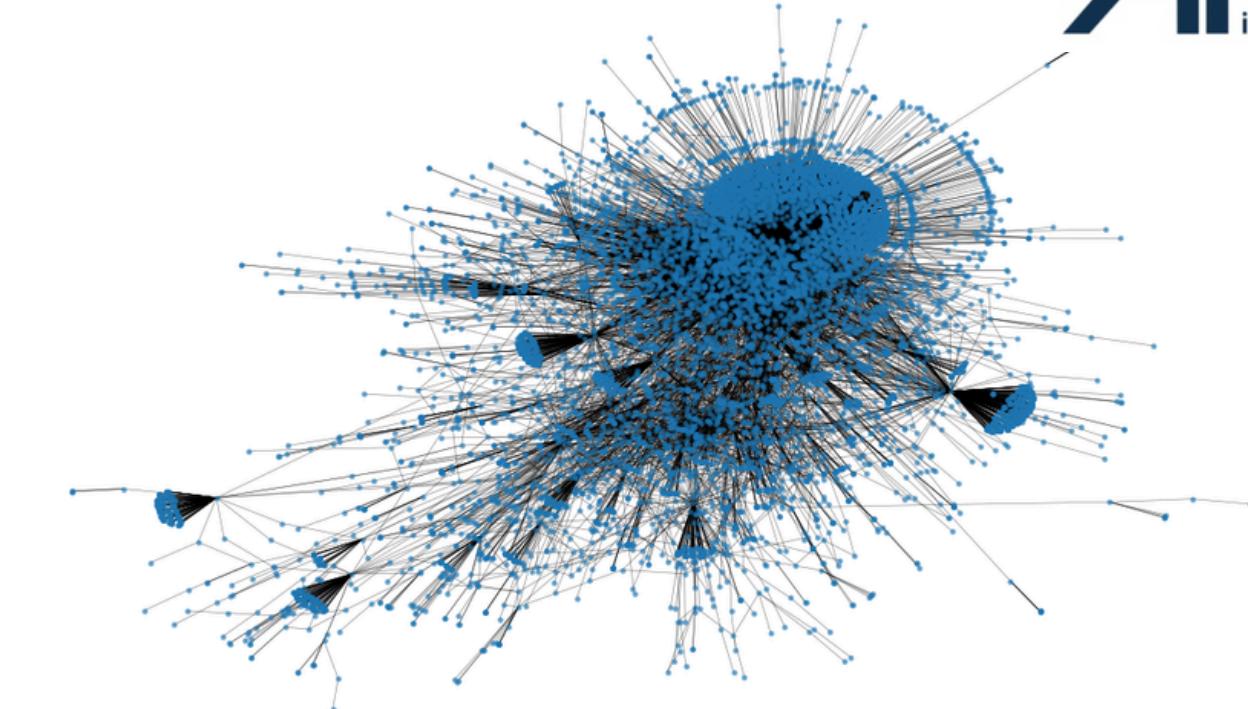
$$N = \{\text{users}\} = \{u_1, u_2, \dots, u_n\}$$

$$E = \{\text{comments}\} = \{(u_i, u_j) \mid u_j \text{ replied to } u_i\}$$

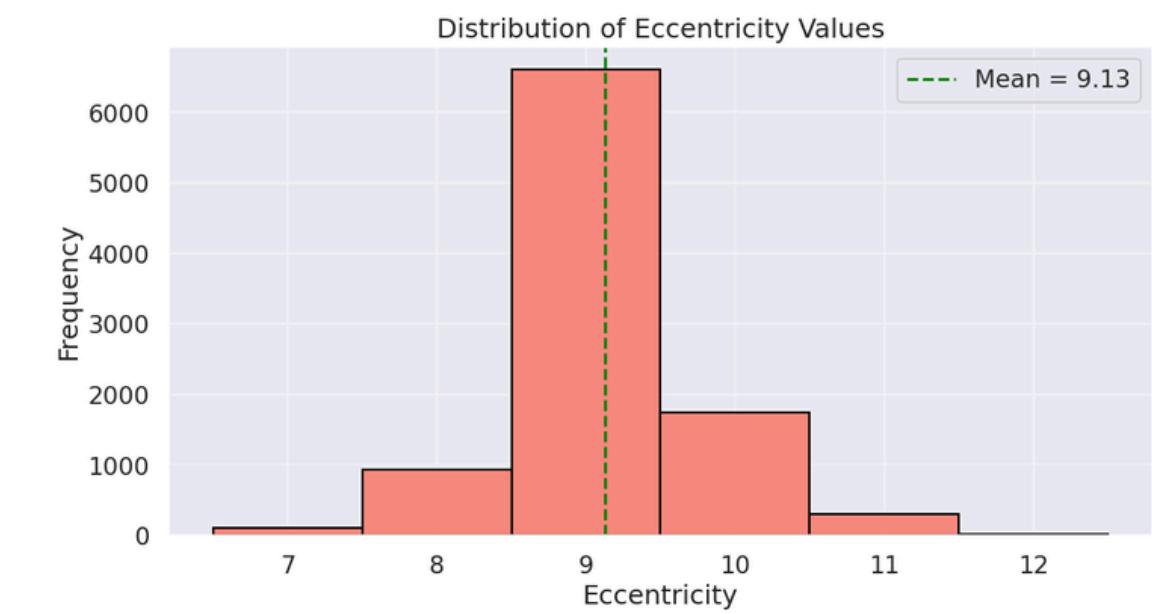
Even though the graph of replies is directed, we treat it as **undirected** so that we can focus on the presence of interaction rather than its direction.

- Even the most distant users are connected through relatively **short paths** (small radius - small eccentricity values)
- As expected for Reddit discussions, the **graph density is low**.
- **High-degree nodes tend to connect to low-degree ones**, in contrast to a complex network.

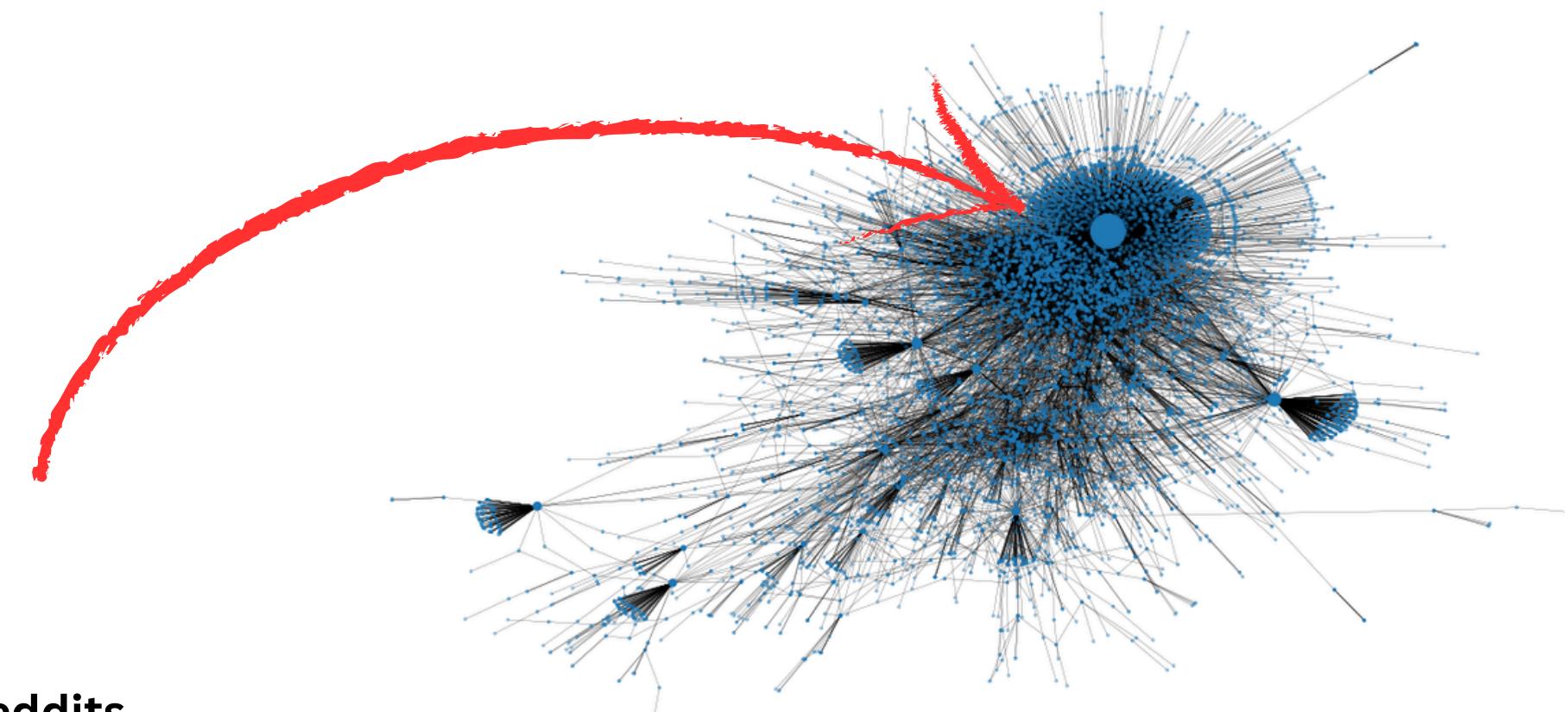
degree assortativity = $-0.2023 < 0$



Graph Statistics	
Radius	6
Diameter	11
Avg. Shortest Path	3.57847
Avg. Clustering Coeff.	0.10621
Density	0.00031



Degree Centrality



Centrality Measures

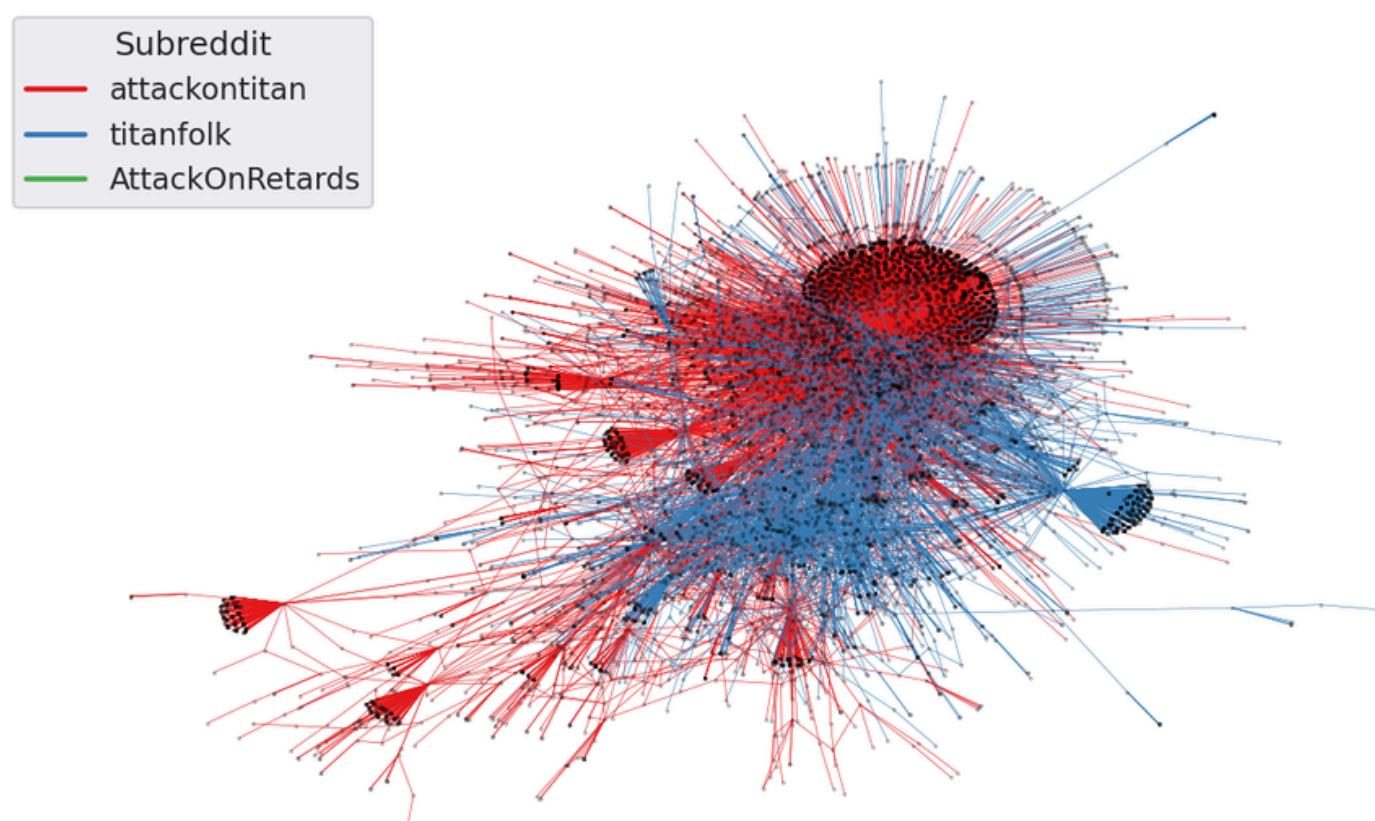
We used the centrality measure to scale the size of the nodes in the plotted graphs.

- **Degree** c.: highlight users who are either highly connected
- **Betweenness** c.: highlight users who act as bridges within the communication network.

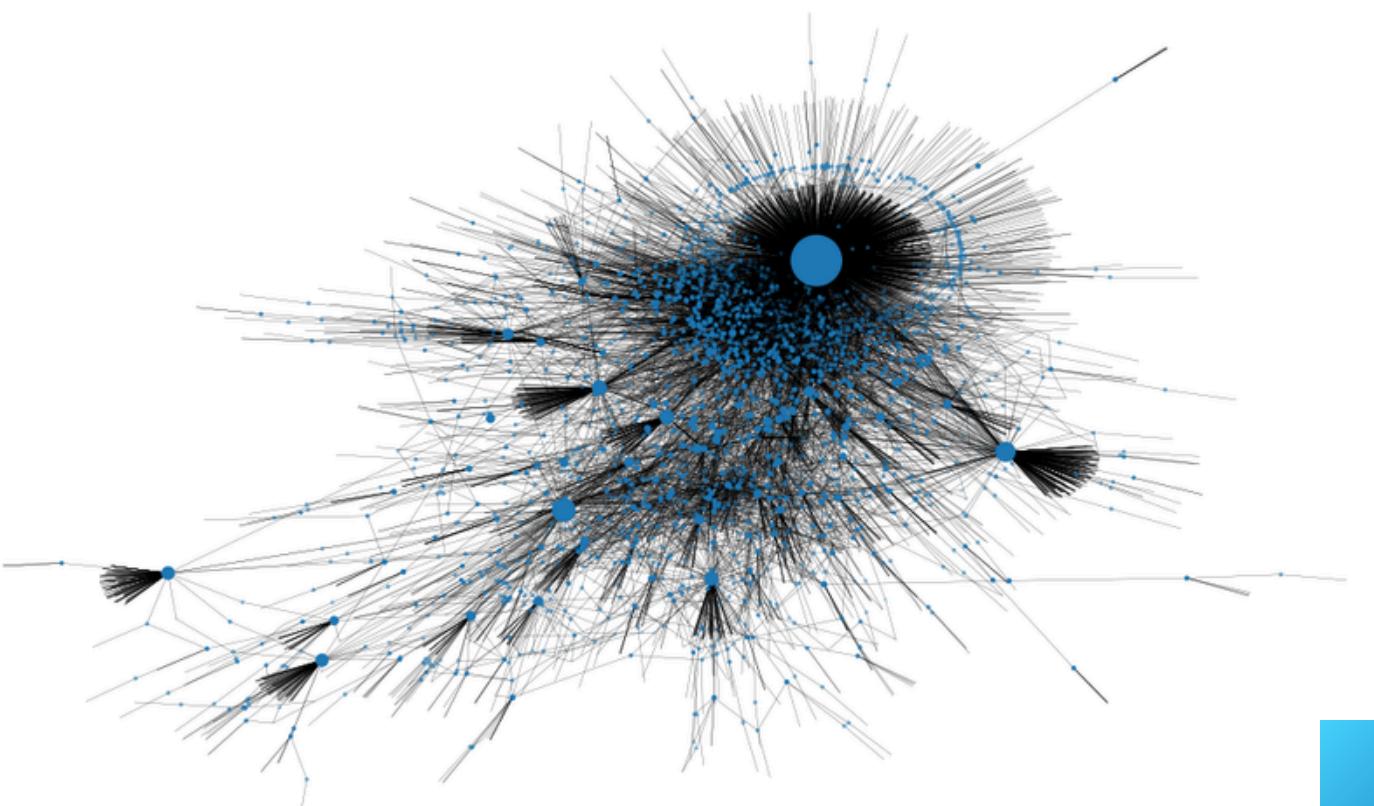
Mr. “Sane-Ni-Wa-To-Ri”, the most important user in our network

Down here: edges color coded by subreddit.

→ highlights the overlap and boundaries between **different subreddits**



Betweenness Centrality



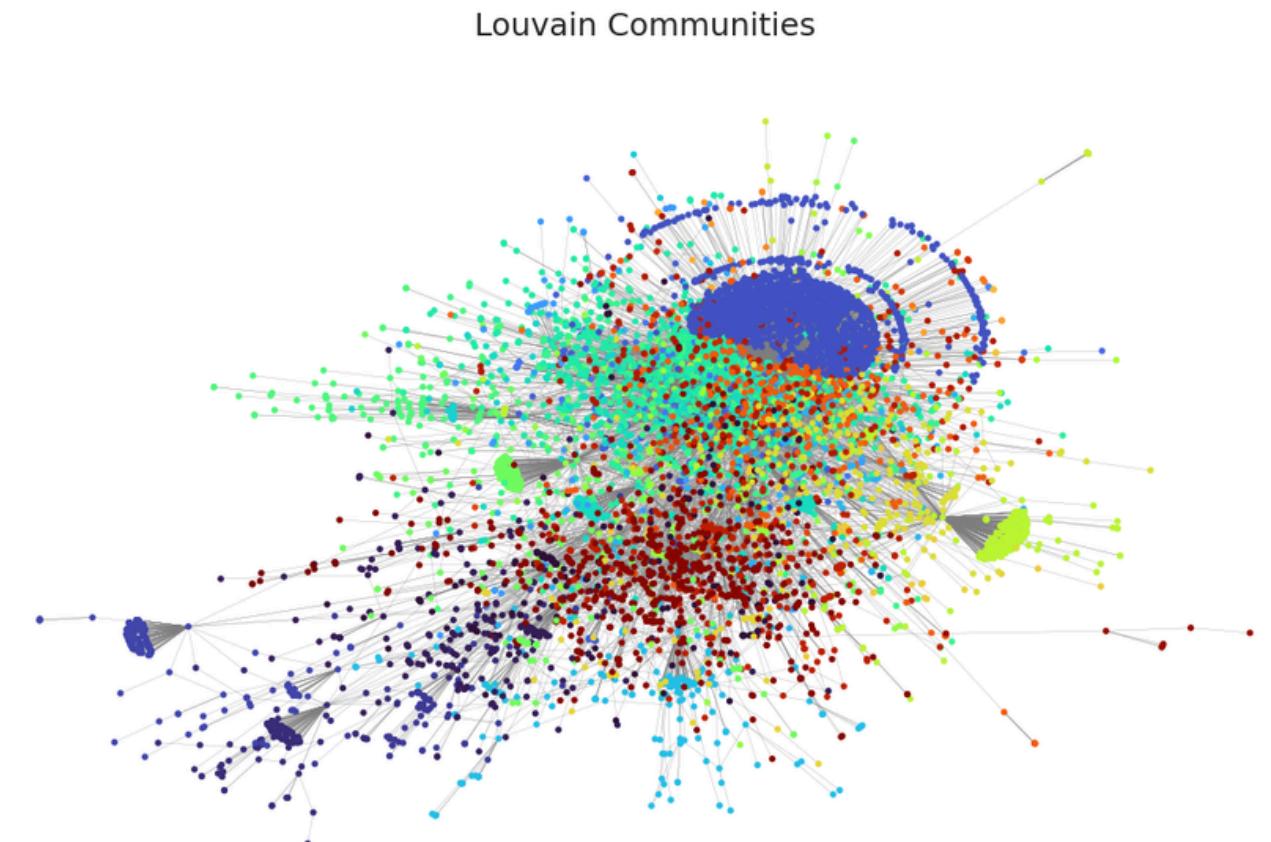
3. Community Detection

Community detection algorithms:

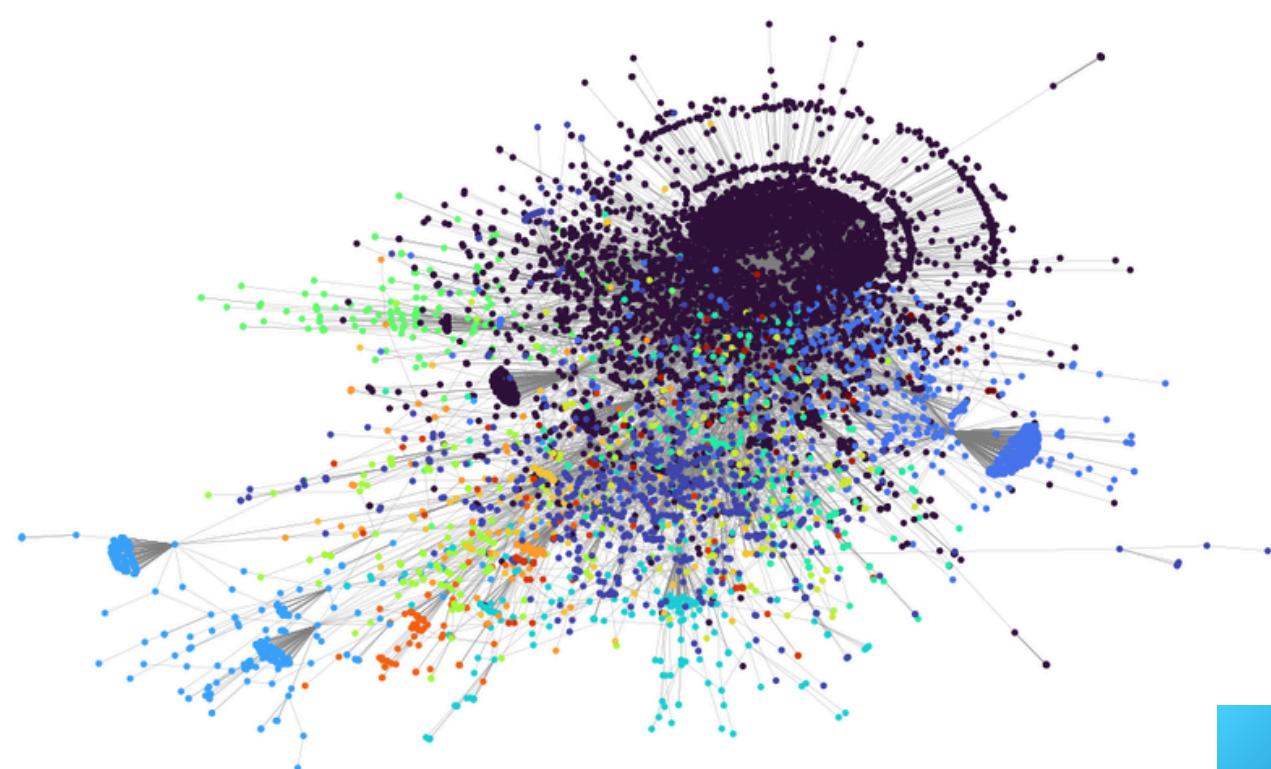
1. **Greedy modularity**: Iteratively merges communities to maximize modularity.
2. **Louvian**: Hierarchically optimizes modularity to detect multi-scale community structures.
3. **FluidC**: Simulates information flow to assign nodes to communities.
4. **Infomap**: Employs random walks and information theory to identify communities.

Metrics:

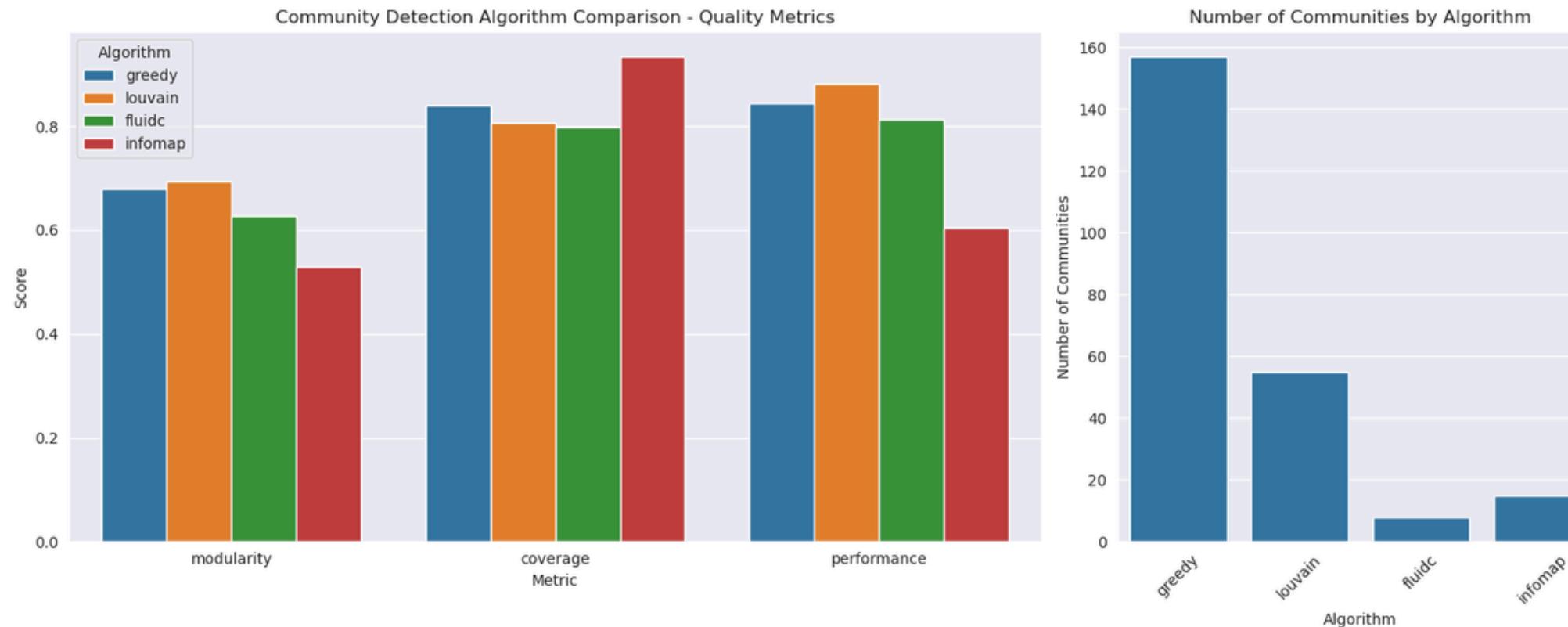
- **Modularity**: Measures the strength of division of a network communities.
- **Coverage**: The fraction of edges that fall within communities.
- **Performance**: Considers both intra-community edges and inter-community non-edges, rewarding well-separated communities.



Louvain Communities



Infomap Communities



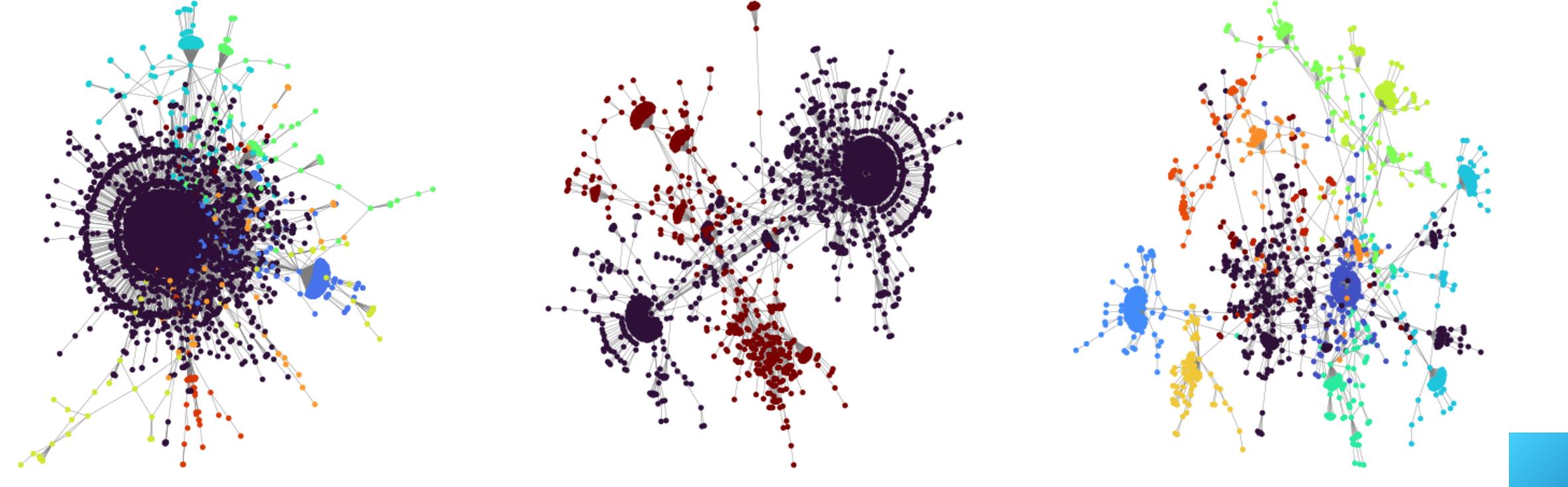
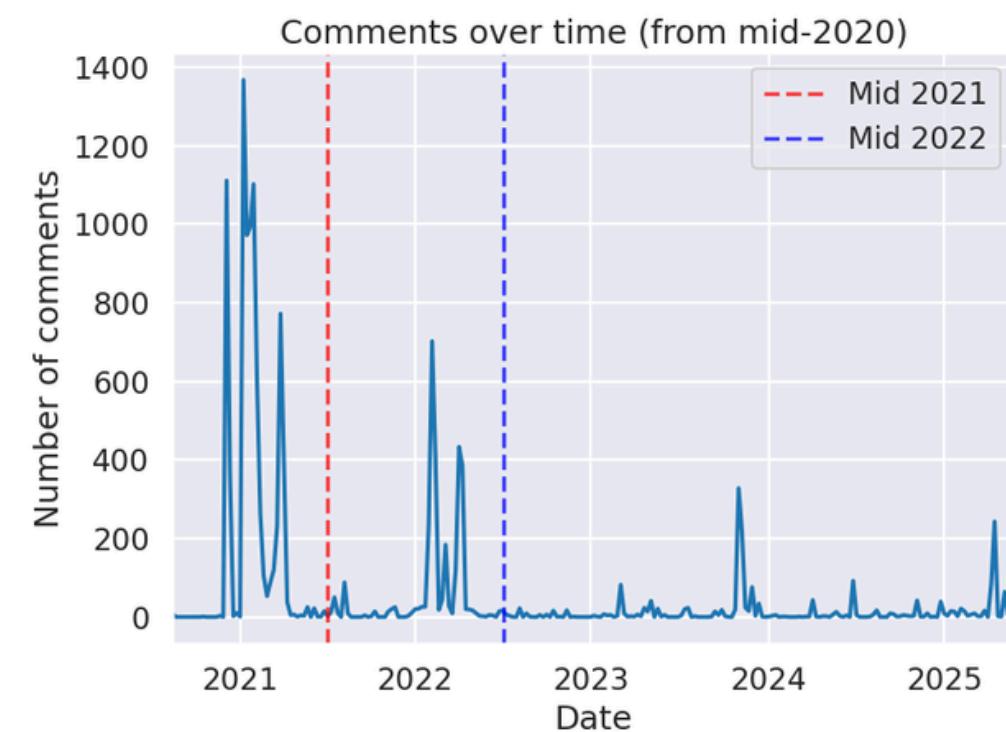
Winner: we chosen the **Infomap** partition as the best one, for its high coverage and reduced number of communities, and consequently higher simplicity and interpretability

Community detection over time

We split the graph into 3 subgraphs:

1. from 2020-06-30 to 2021-06-30
2. from 2021-06-30 to 2022-06-30
3. from 2022-06-30 till now

And run Infomap on them...



4. Sentiment Analysis

We employed four main sentiment analysis approaches:

- **AFINN**: A lexicon-based method that assigns integer scores to words.
- **NLTK Opinion Lexicon**: A lexicon-based approach using positive and negative word lists.
- **VADER**: A rule-based model specifically tuned for social media text.
- **Transformer-based Model**: A ML model that classifies sentiment on a five-point scale.

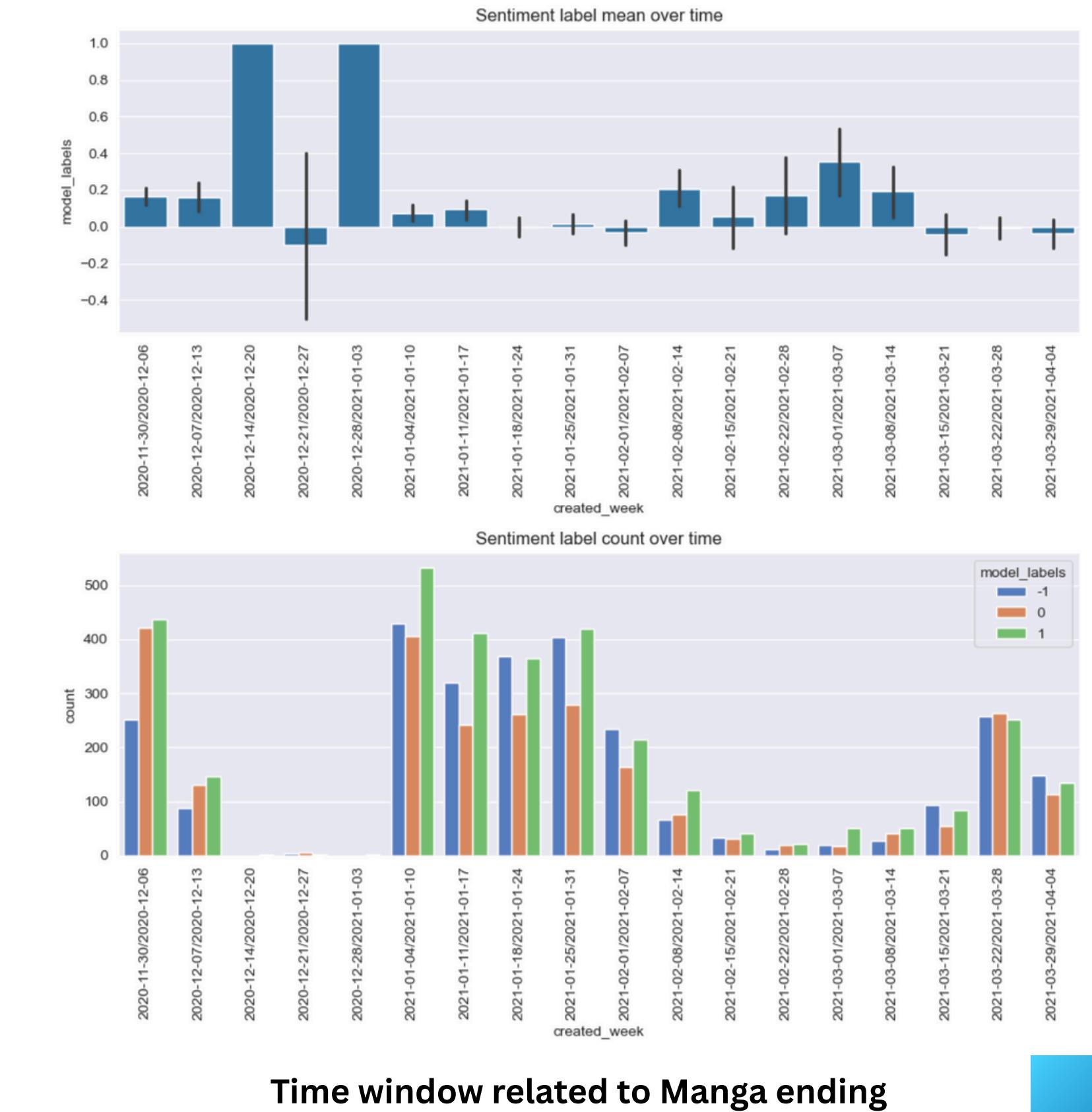
Transformer-based Model has **context-awareness** → this will be our preferred approach.

The most interesting time periods: The end of manga and anime.

As seen before, these are the periods when we receive the most comments.

Manga ending

- slightly **positive** sentiment
 - weeks with an **high count** of replies:
 - mean centered around zero
 - small variance.
- This indicates **polarized debates**: no dominant opinion that takes over the conversation.



Anime ending

- slightly **negative** sentiment
- **high variance**
- more **neutral** voices with respect to the Manga ending
→ the last two points suggest us that we have an ongoing debate

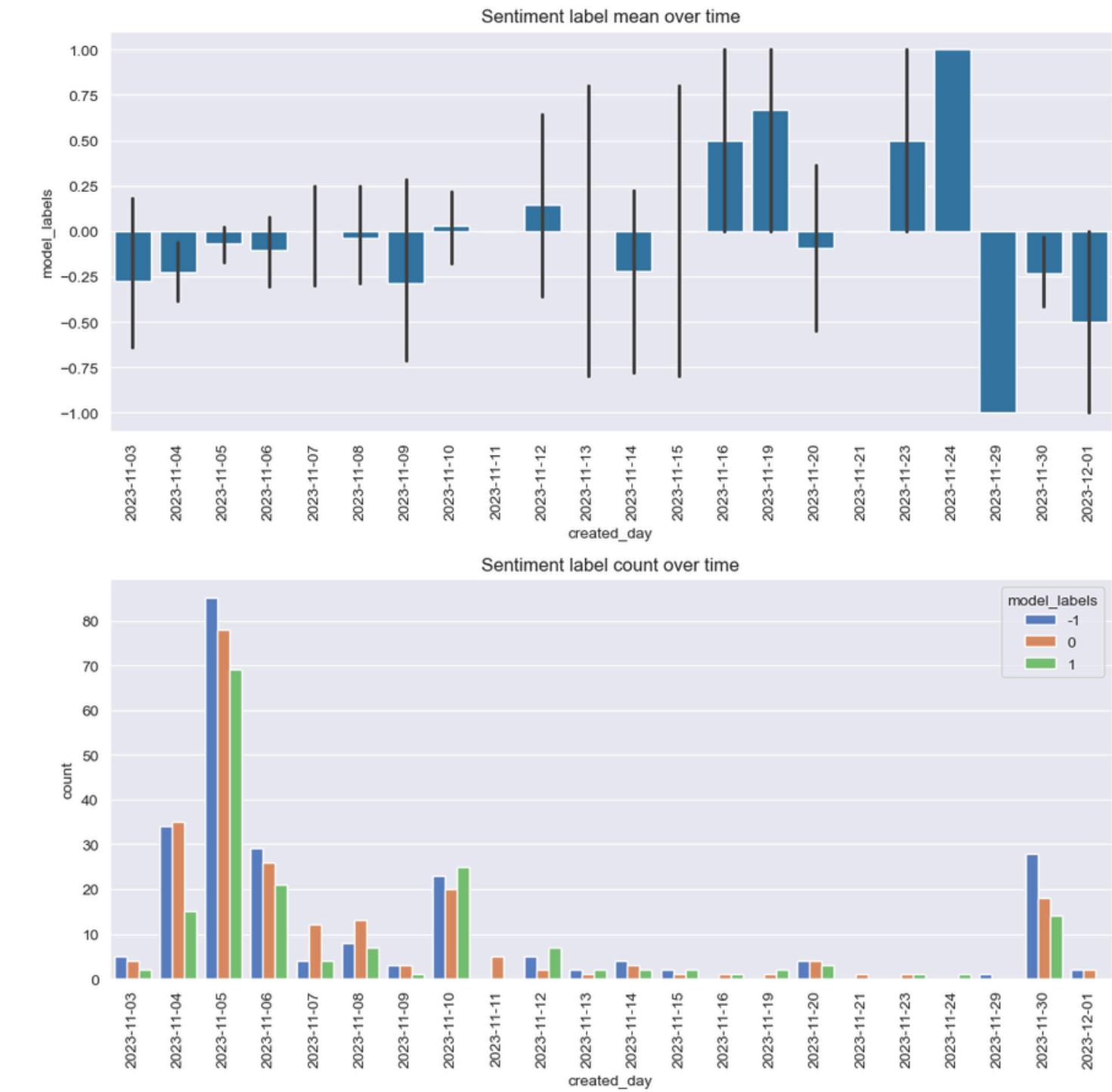
This can be a sign that the fans didn't appreciate the changes that the Anime did with respect to the Manga, like tighter pacing and a fragmented release schedule of the episodes.

Sentiment in the top 3 communities

Largest community: Sentiment skewed by Sane-Ni-Wa-To-Ri, otherwise leaning positive

Top 2 and 3: Neutral for the former, positive for the latter

Possible lore-masters or theory-crafters?



Time window related to Anime ending

5. Topic Modeling

Methods:

- **WordCloud** as simple and first approach
- **BERTopic** - a pipeline that leverages contextualised BERT embeddings.

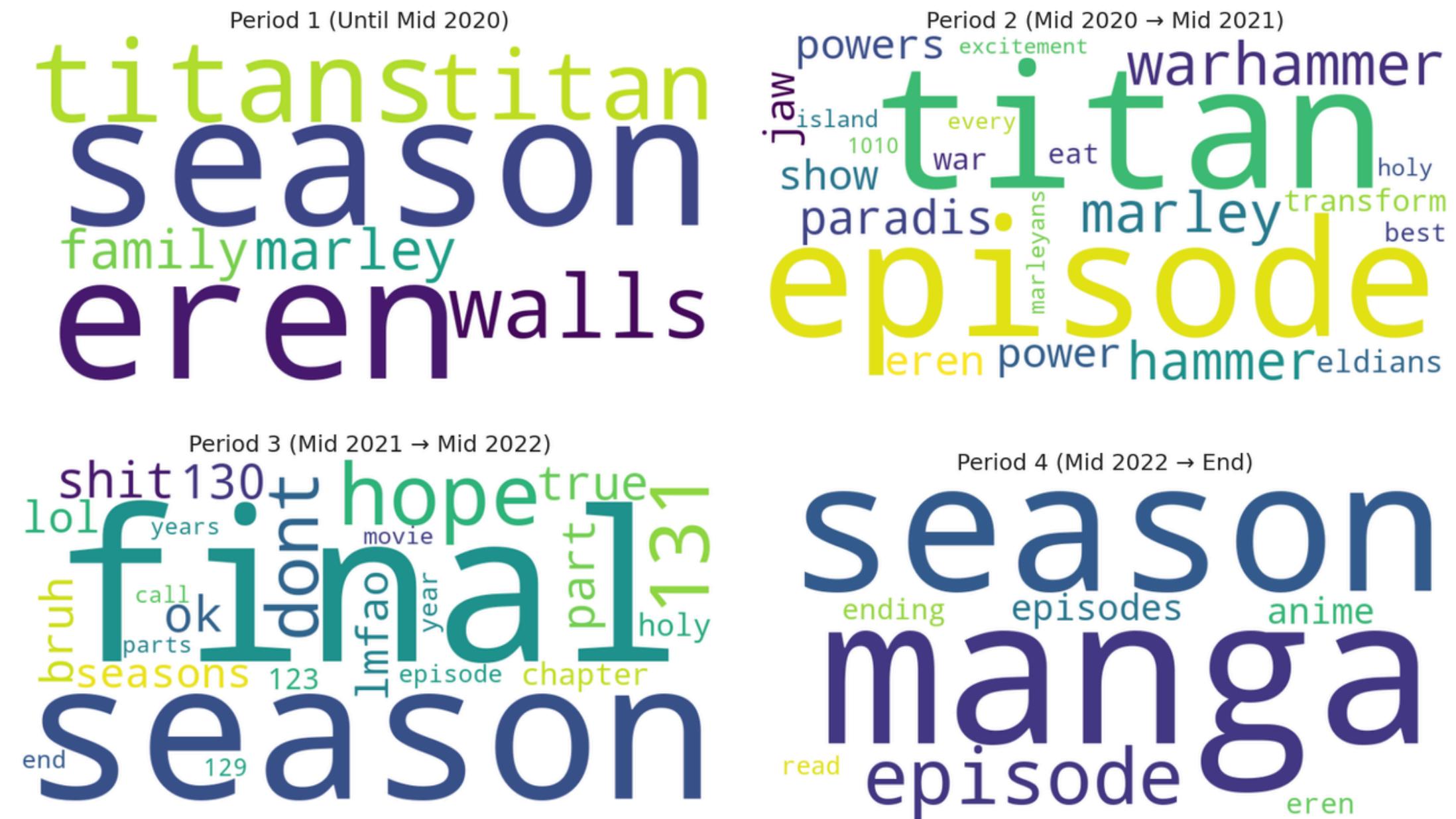
We applied it in the following settings:

- per community
- **per time window** → we will focus on this one, that produced the most interesting results.
- per subreddit

NOTE: Here the wordclouds are used just to visualize the results from BERTopic

Results

- **Start – mid 2020:** "Eren" and "titans"
- **mid 2020 – mid 2021:** War between Marley and Paradis



- **mid 2021 – mid 2022:** Emotionally charged and vulgar slangs (*shit, hope, lol, bruuh, lmfao*) → reactions to the ending.
- **mid 2022 – end:** discussion terms

6. Named Entity Recognition

Models (*Spacy* library):

- **en_core_web_sm**: lightweight, uses tok2vec
- **en_core_web_trf**: Transformers based (“*roberta-base*”)

Character co-occurrence graph

We use the PERSON tag to build the following graph:

$$G = (N, E)$$

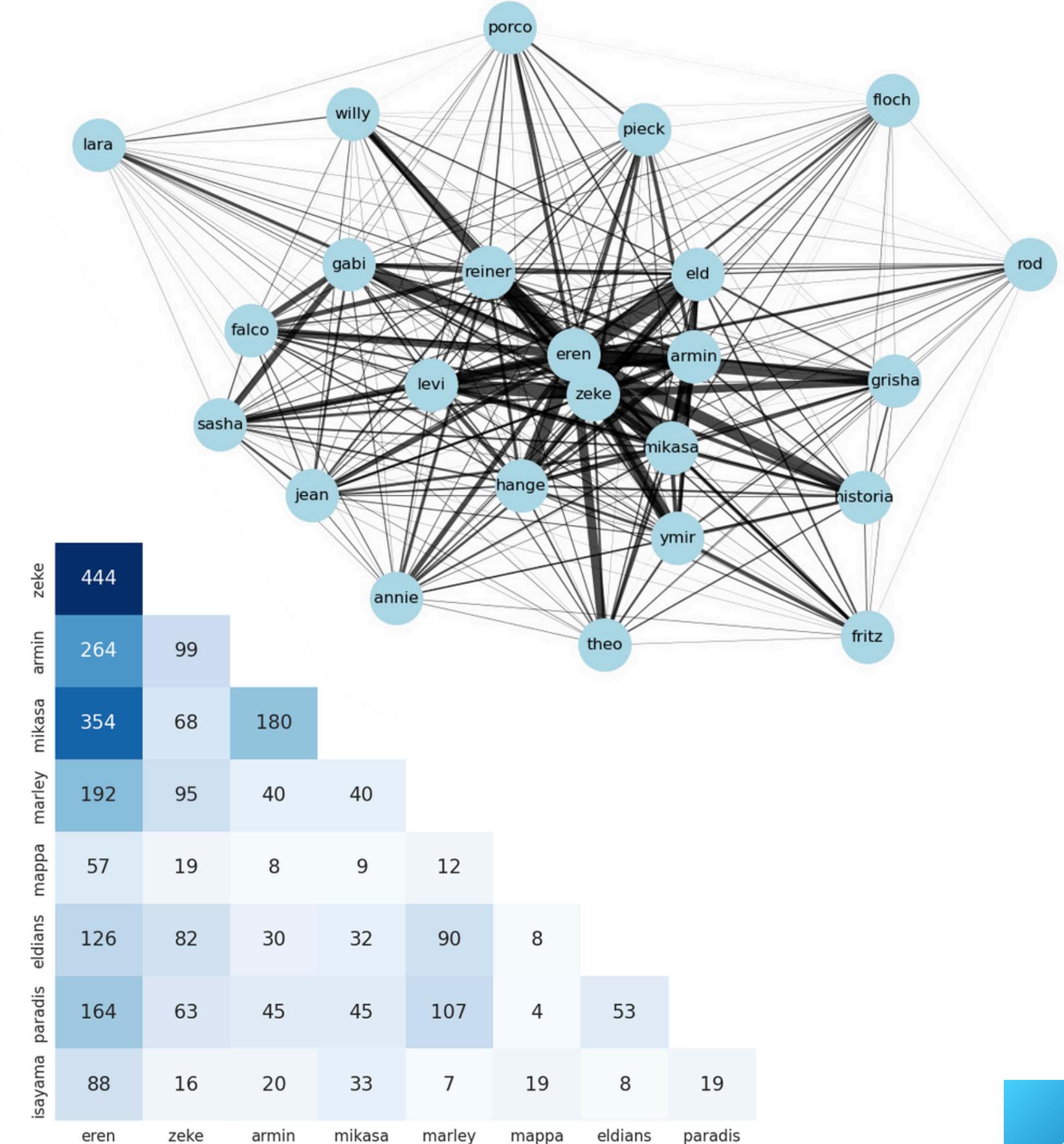
$$N = \{\text{characters}\} = \{c_1, c_2, \dots, c_n\}$$

$$E = \{\text{co-occurrences}\} = \{(c_i, c_j, w_{ij})\}$$

Where w_{ij} indicates the number of comments in which c_i and c_j are mentioned together.

NEL

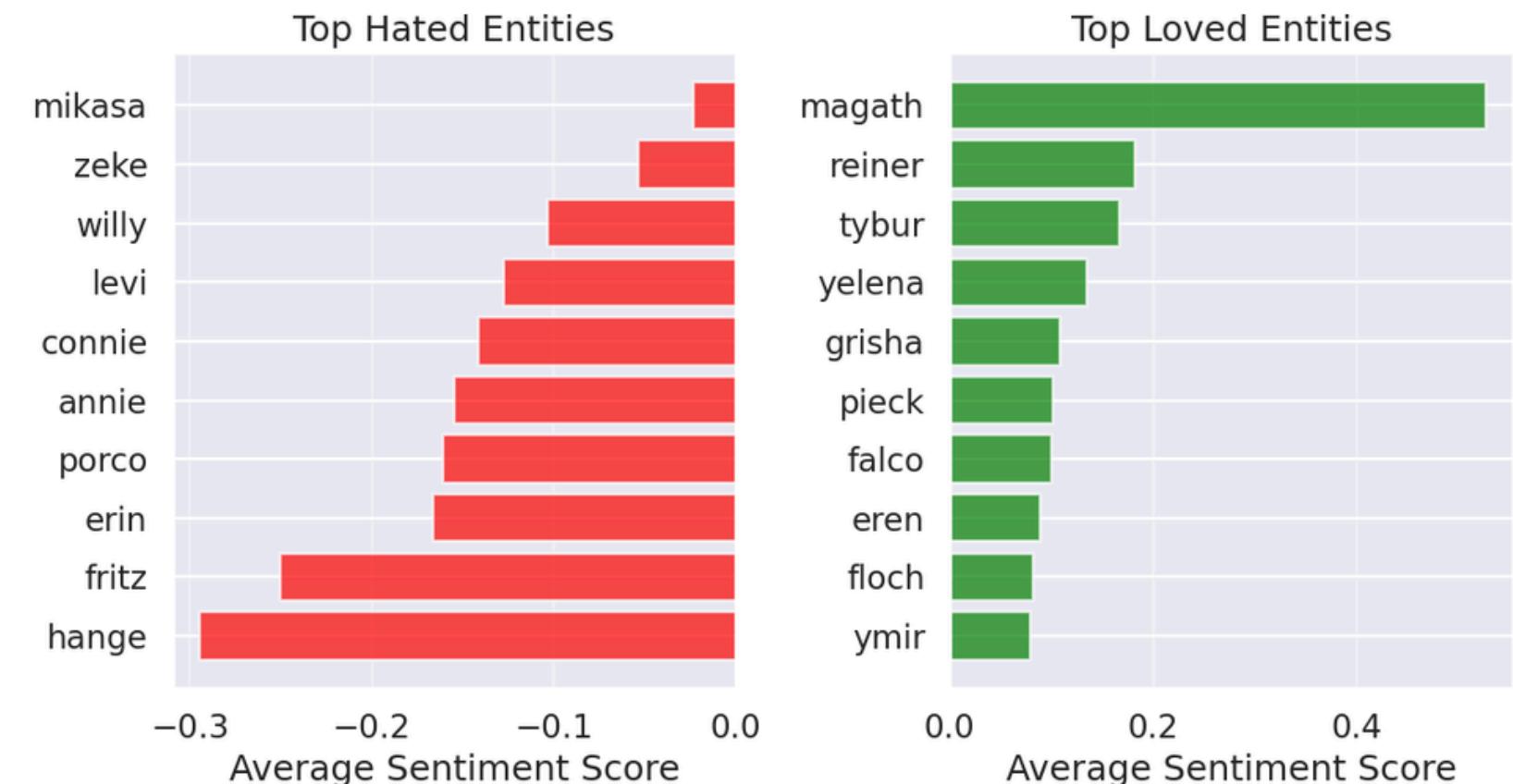
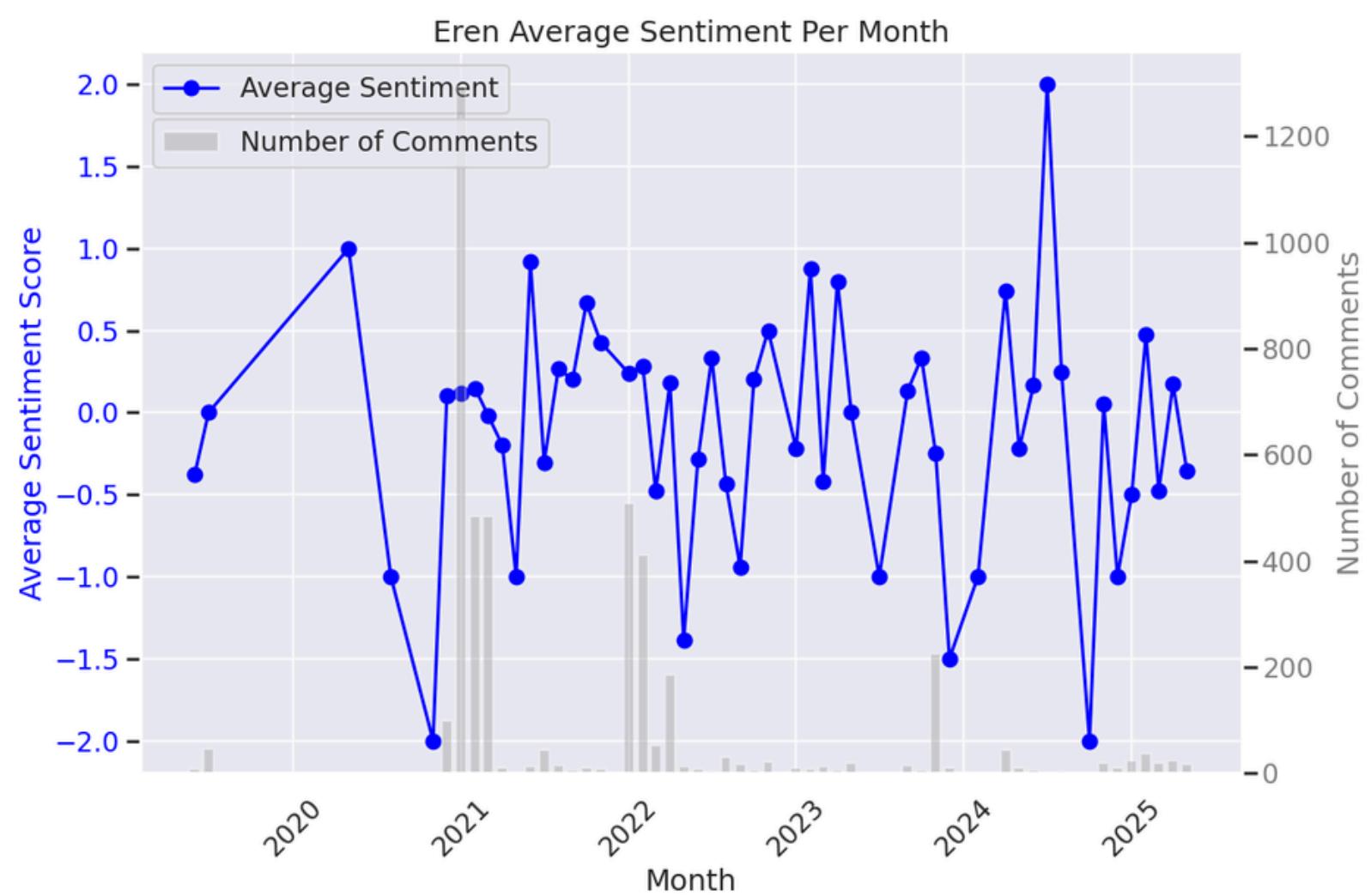
Then we linked such names to the entities in the **DBpedia** knowledge graph.



Character sentiment over time

How are the entities perceived by our communities?

Here the characters that appears in the most comments with the **highest/lowest sentiment**

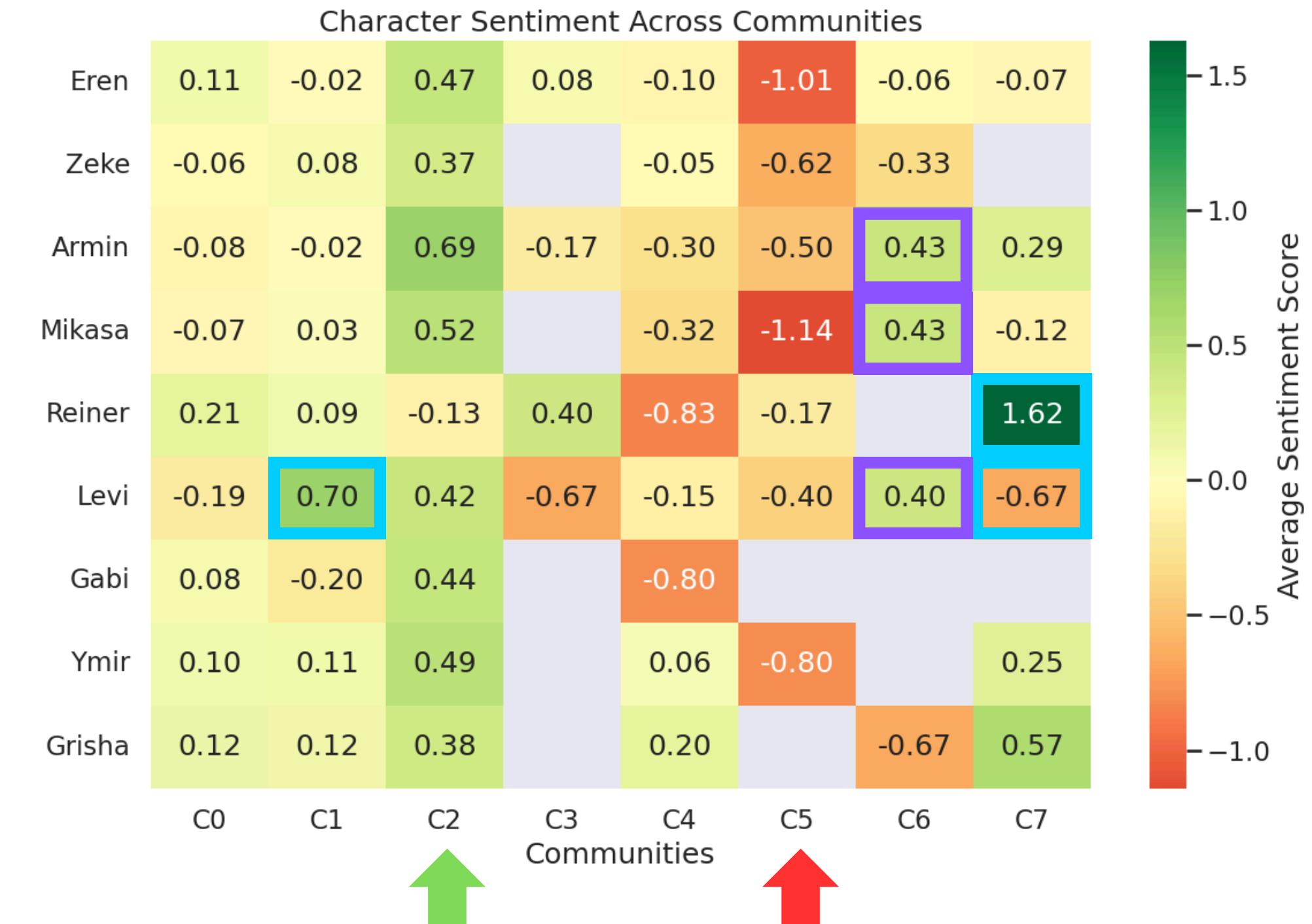


And here is the evolution of our protagonist, **Eren**. People's feelings towards Eren's character **vary greatly**. This is because it is only natural to have mixed feelings about what he did at the end of the story (he committed a mass murder.).

Characters sentiment per community

This final analysis on the sentiment of characters in the different communities is what really helped us to differentiate our communities.

- **C1:** **Levi's fanclub**, which is a ruthless, cold and efficient soldier - aka *humanity's strongest soldier*
- **C7:** roots for **Reiner** but hates Levi. Reiner is an infiltrated warrior of Marley, torn apart by his own guilt - the complete opposite of Levi.
- **C5:** **critical or salty** corner of the fandom during the ending.
- **C2:** **appreciative** group, with a preference towards Armin.
- **C6:** **Paradis Loyalists**. Not aligned with Eren's "end it all" vision, and blaming Zeke and Grisha for the conflict, they have a soft spot for Mikasa, Armin and Levi.



NOTE: the model analyze the sentiment of the comment, not really the sentiment of the user with respect to the characters present in the comment.



Thanks for your attention!