Tracing the genealogy and change of TikTok audio memes

Team Members

Lucia Bainotti (facilitator), Sarah Burkhardt, Yan Cong, Jingyi Zhu, Jesper Lust, Kate Babin, Salma Esserghini, Iliass Ayaou, Amine Kerboute, Micky-Love Mocombe, Frédéric Lecat, Amina Mohamed, Simran Tamber, Devin Mitter, Sama Khosravi Ooryad, Jasmin Leech, Tommaso Elli, Kristen Zheng

Contents

Summary of Key Findings

- TikTok sounds are relevant as a feature that afford different kinds of behaviours, not only memetic ones.
- Among these behaviours, practices of '**sound surfing**' emerge, i.e. using the sound and trend with the aim to get visibility.
- Along this line, sounds usually start as a meme but become quickly used to boost the visibility of some unrelated videos and content.
- In the analysis, we identified three types of TikTok sounds, which reflects three ways in which TikTok audio memes can be generated: **integrated TikTok sounds** (e.g. the Stressed Out case study), **semi-native TikTok sounds** (e.g. the Mama said case study), and **native TikTok sounds**. Moreover, these sounds are used in different memetic ways and have different trajectories of development.
- The analysis provided insights not only into the genealogy of TikTok memes, but also pointed at the idea of the '**dilution of a meme**', i.e. when the trend reaches its peak in popularity, it starts getting diluted with non-related content that does not follow any specific template, and therefore the trend itself progressively ceases to exist.

1. Introduction

While there is currently an abundance of research in the field of new media into categorizing memes in formalistic and structural terms, there remains a paucity of empirical research into one of the defining features of the meme: it's "viral" spreading and "mutation" over time. The proposed project seeks to address this question through repurposing a single feature on a single platform that has, in recent years, become identified with such dynamics: TikTok.

TikTok is an increasingly popular platform, and part of its success is the innovative short video platform experience it offers, thanks to the personalized short videos offered by the algorithm in a stream of content that invites users to continuously scroll the For You page. Part of TikTok's popularity also resides in the creation of video largely based on musical snippets known as 'sounds', with users lip-syncing the lyrics and performing to the music, as well as joining so-called challenges. Tiktok affordances, indeed, encourage users to create, remix, imitate, and replicate content, thus prompting memetic practices and giving birth to imitation publics (Zulli & Zulli, 2020).

In this sense, the notion of internet meme (Shifman, 2012) is extended to the level of the platform affordances and includes a specific emphasis on audio content. TikTok's memetic logic, together with the possibility of creatively and easily reusing audio clips, sounds, and music it affords, makes it a particularly suitable platform to analyse issues of meme creation and evolution. In this context, this project aims to analyze the genealogy and evolution of a meme by looking at the 'sound' indexing feature on TikTok.

Secondly, the project aims to understand how TikTok memes emerge and develop over time, as well as how sounds are creatively or subversively used and appropriated. This is particularly important considering that users can engage with, and reinterpret, the meanings of sound and audio clips (Kumar, 2015).

Thirdly, this project is interested in developing and applying innovative ways of using visual analytics for social media research and particularly testing techniques to study audio-visual content. On TikTok, 'small stories' are vehiculated by means of short videos, which conflate audio and visual content and can convey different and even contrasting meanings. One first challenge in analysing TikTok is taking into account the multimodality of the platform, as well as the different levels of signification it offers. Another challenge related to the study of TikTok content lies in its expressive dimension, given that visual elements are functional to vehiculate performative and expressive messages and not only aesthetically pleasing and curated representations.

To address these purposes, the research focuses on two case studies, based on two TikTok sounds:
- the song '**Stressed out**' (by 21 Pilots), released in 2015 and then recently integrated as a TikTok sound in different versions - an 'official' sound, that we will refer to as 'Stressed Out Original' and a remixed sound, here called 'Stressed Out Remix'
- The song '**Mama said**' (by Lukas Graham), which again came out before TikTok existed, started to become popular on the platform after that the artist himself promoted it as a TikTok sound with its own trend.

2. Initial Data Sets

Data collection: following the sound(s):

- 's t r e s s e d o u t' (stressed out_remix) and 'stressed out' (stressed_out original)
- Mama said

Tool: Zeeschuimer Firefox extension (link).

Initial Data Sets:
- Stressed out_remix: 3,059 tiktoks
- Stressed out_original: 4,730 tiktoks
- Mama said: 3,926 tiktoks

3. Research Questions

How can we trace the genealogy of a TikTok meme by looking at TikTok sound indexing feature?

Do TikTok memes change over time and if so, in which ways are they interpreted, reappropriated, or subverted?

How can we analyze TikTok content by accounting for its multimodality as well as its expressive dimension?

4. Case study (1) - "Stressed Out" sound

4.1. Methodology

In this project, we aim at developing and testing research techniques useful to study audio-visual and memetic content, specifically on TikTok. We build on already established techniques for visual media research (Niederer, 2018; Rogers, 2021) to propose innovative methods of analysis in a Digital Method fashion.

4.1.1. Hashtag analysis

To make sense of TikTok audio meme, the first step of our analysis is to focus on the textual metadata which comes with TikTok content, and specifically, hashtags. We first collect the top 25 hashtags for both the original and remixed stressed out sound, and create a temporal bump chart to display the presence, absence and change in hashtags over time.

Secondly, we perform a co-hashtag analysis in order to account for the relationships among hashtags. We apply the modularity class algorithm to detect different clusters, which can point to the presence of subcultures.

4.2.2. Visual similarity analysis

To account for the visual components of TikTok videos, we start by performing a visual similarity analysis. Image similarity analysis consists of the visualisation and analysis of a large collection of images grouped by their visual attributes (such as colours) in an image wall.

More specifically, for each TikTok sound in our analysis, we plot the 'cover thumbnails' of each video in an image wall created with PixPlot. PixPlot is an open-source software developed by the Yale Digital Humanities Lab (DHLab), which algorithmically clusters similar images near one another by considering not only colours but also the type of content represented in each picture (e.g. faces, landscapes...). TikTok cover thumbnails consist of the 'preview' of each video, which is the first thing that users can see by

opening the platform. Usually, the cover represents not only a preview but the first frame of the video. Therefore, cover thumbnails provide insights into how each TikTok appears to viewers and how it starts.

The analysis of cover thumbnails in terms of visual similarity can be useful as a first step to detect the emergence of memetic practises and TikTok templates. However, as will be accounted more in-depth in Section 5.3.3., such an analysis presents some relevant limitations. First, it does not account for the dynamic nature of TikTok content, and, second, it only provides information about a small portion of TikTok videos, namely their cover.

4.1.3. Templatability analysis

In order to overcome these limitations, we integrate image similarity analysis with what we call 'templatability analysis'. Templatability analysis consists of the interpretation of video stripes, created by extracting frames by change of scene from a set of videos. Video stripes can be considered as a specific kind of composite image (Colombo, 2018), characterised by the juxtaposition of the different video frames next to each other. The templatability analysis consists of the visualisation of 250 videos for the 'stressed out' sounds (Fig. 1), which can be filtered by 'most followed account', 'most reproduced videos', and 'top-10-hashtag (sort plays)'.
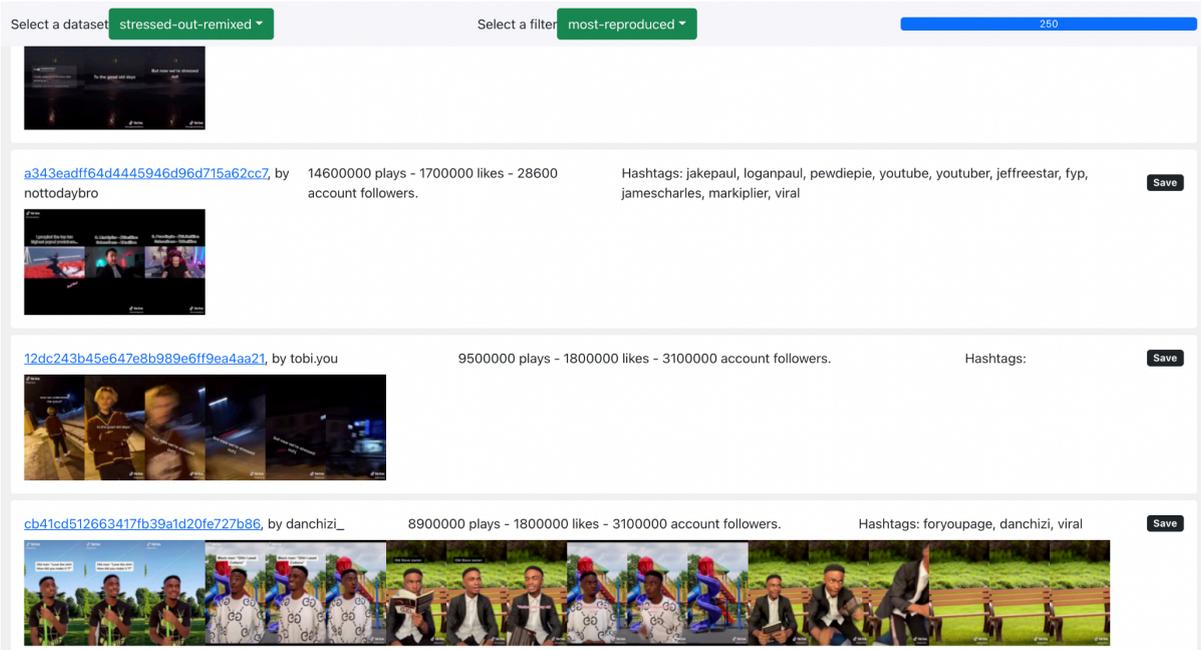

Figure 1. 'Stressed out - remix' image stripes, sorted per 'most-reproduced'.

Templatability analysis is useful to understand both the main themes emerging from the data, as well as the different templates that can be found in TikTok content sharing the same sound. Both this information is in turn relevant to highlight the changes, appropriations, and re-appropriations of TikTok memes.

4.1.4. Video stack analysis

Moreover, in order to examine visual patterns for both case studies more closely, so-called 'video stacks' can be used. Inspired by the existing tool Image stacks, a video stack comprises five TikTok videos superimposed over one another, each at 30% opacity. As a result of this, similar motions can be detected as well as other overlapping visual elements. In order to create more clarity, two video stacks exist for each collection of five videos: one where clips are continuously superimposed, and one where they enter one after the other, counting down from five.

The following categories are distinguished and used to create the video stacks:

'**Stressed Out**' and '**Stressed Out Remix**' case study, top five
- viewed videos
- most followed accounts' top viewed videos
- viewed 'duet' videos
- viewed videos from November 2020
- viewed videos from February 2021

'**Mama Said**' case study, top five (see also Section 5.2)
- viewed videos
- most followed accounts' top viewed videos
- viewed 'duet' videos
- viewed videos from June 2021
- viewed videos from September 2021

In order to create the video stacks, the top months of popularity are distinguished from the trend graph of each sound (see Figure 2, and Figure 26). Subsequently, the top five viewed videos from these months (September 2021 for 'Mama Said' and February 2021 for 'Stressed Out') are turned into video stacks, as well as the top five viewed videos from three months prior to peak popularity (June 2021 for 'Mama Said' and November 2020 for 'Stressed Out'). Here, the video stacks of the peak popularity months closely resemble the overall top videos, both in terms of actual content as well as the non-trend conforming nature. The pre-peak month top videos, however, are more visually similar:

transitions frequently match up perfectly, with many videos being more explicitly 'TikTok' in their incorporation of one or two individuals in front of a vertical video, displaying movements. As such, the trends can be seen reflected more clearly.

4.1.5. Sound analysis with audio recognition

For analysing the evolution of a song shared via TikToks, we developed a method to analyse its quality evolution with an audio recognition technique. To analyse how songs are re-appropriated, remixed, or kept unchanged, we took an approach similar to Shazam, an audio fingerprinting technique to compare audio files similarity (Cano et al., 2005).

Step 1: Audio data retrieval

From all TikTok data originating from the original song's id, we used the URLs which link to the TikTok page to download the mp4 files with a Python script. Here, in a loop, each page was visited with the Python requests package having the streaming option enabled. The video data was then retrieved in streamed chunks and saved as a mp4 file. Using Python, in a second step, we extracted the mp3 files from all mp4 files (Guven, 2021). Further, we made sure to have an mp3 version of the original song.

Step 2: Install Python library Dejavu

To fingerprint and compare all retrieved TikTok audio files to the original file, we used the audio fingerprinting Python implementation Dejavu (Drevo, 2021). Following the instructions on the GitHub repository which mainly contain installing Python based libraries and packages required to run the script, we then installed Dejavu, without Docker, by cloning the repository and running it locally with Python3. The installation and configuration of a local MySQL database is needed as well, which is going to store the audio fingerprints and data from the matching procedures.

Step 3: Recognise TikTok audio data in original song with Dejavu

Fingerprinting original song
In order to enable an audio comparison with the original song, Dejavu requires to first fingerprint the song we want to match against, whereas its fingerprints are then stored in the database. Hence, the first step was to 'fingerprint' the original song to match all other TikTok sounds against. Fingerprints are saved as hashes, which are uniquely indexed numbers that save structural features of a sound across a particular range in terms of frequency or time. Similar to a fingerprint, they uniquely encompass sound

information for a given part of the song. By comparing captured fingerprints across songs, highly similar fingerprints which both carry similar audio information are considered as matches. Given that the same song has similar pitches or silent parts, it can comparably well detect song similarity irrespective of noise, due to its structural similarity which overlaps at some point.

Recognizing TikTok audio in original song
As a second step, the matching procedure was run by 'recognizing' each TikTok audio file, which means its fingerprints are created and compared to the original song in the database. From this procedure, we retrieved the resulting similarity features per song in a csv file. The features that we retrieved are

- the amount of hashes (fingerprints) per TikTok sound input;
- the amount of hashes (fingerprints) in original song;
- the amount of matching hashes between input and original song;
- the input confidence (amount of matching hashes divided by the amount of hashes of TikTok sound).

We took the input confidence value (being a number between 0 and 1) as the significant original-song-similarity measure of a TikTok sound.

Fine-tuning
For the matching procedure, we did some fine-tuning of the algorithm in the settings.py file to enable that the matching is fine-grained and takes into account much local information across short spans. Because short audio samples are typical for TikToks, this should assure that short audio samples are matched with the long original song. Therefore, we changed

- the DEFAULT_FAN_VALUE from 15 to 5 to reduce 'the degree to which a fingerprint can be paired with its neighbours';
- the MAX_HASH_TIME_DELTA from 200 to 70 reduce 'the threshold on how close or far fingerprints can be in time in order to be paired as a fingerprint';
- the PEAK_NEIGHBORHOOD_SIZE from 20 to 1 to reduce 'the number of cells around an amplitude peak in the spectrogram in order to consider it a spectral peak'.

Assessing and interpreting results
Through our fine-tuning, the resulting similarity measures needed to be reinterpreted. Therefore, we implemented a testing and assessment strategy to detect which kind of audio files are located in which spectrum of similarity measure (input confidence).

Taking the original song as an input to be matched against itself, it turned out that the input confidence was 0.4. Grouping voice-dominant, noisy, remixed, completely different or mute sounds, and songs close to the original to test where they lay in the spectrum, we could find that from 0.4 towards 0, audio sounds become more 'diversified' in the sense of being remixed, noisy, with distinguishable beats or noise overlaying the sound. Whereas from 0.4 towards 1, audio sounds become more 'voice-dominant', with songs close to 1 matching files that contain only speech without any background sounds; and songs closer to 0.4 containing speech overlaid audio files. Therefore, we then decided upon 0.4 to be the 'threshold' value to match the original song, gradually going towards 'diversification' or 'voice-dominant' sounds.

Step 4: Visualisation of results

We then visualised our results in bar plots with Rawgraphs. For this, we uploaded a csv file which contained the similarity measure per TikTok sound, and furthermore all available information about the platform-based measures (e.g. likes, comments, followers, etc.) of the respective TikTok.

On the x-axis, we plotted the amount of TikToks, and on the y-axis, the similarity measure from 0 to 1. Therefore, a bar contains the number of TikToks which share the same similarity measure to the original song. Further, we combined this visualisation with the colouring of the bars to examine how platform-native measures (e.g. likes, comments, followers, etc.) correlate to the audio quality. Here, the total sum of the platform-native measure was used to colour the bar, showing which sound quality measure goes along with which 'popularity'.

4.2. Stressed out - Results

4.2.1. Tracing the development of the 'Stressed out' sound

Our first case study is the "Stressed Out" audio meme that went viral on TikTok in late 2020 through early 2021. The meme uses the 2015 song "Stressed Out" by the band Twenty One Pilots which was posted by the platform in 2019 and is affiliated with the band's verified TikTok account. Another version of the song, a remix made by user @crying.bandito.audios is associated with the trend as well and we included both sounds separately in our analysis.

TikTok creators use the song in their videos, in particular the chorus "wish we could turn back time / to the good old days / when our mamma sang us to sleep / but now we're stressed out" and often either lip-sync along or use text over the video to emphasize those lyrics.

The original song was considered to some as generation-defining when it was originally released in 2015, many millennials identified with the anxiety and malaise communicated through the lyrics and that same sad and nostalgic energy is one the main themes expressed in the videos being shared by TikTok creators who utilize this meme format.
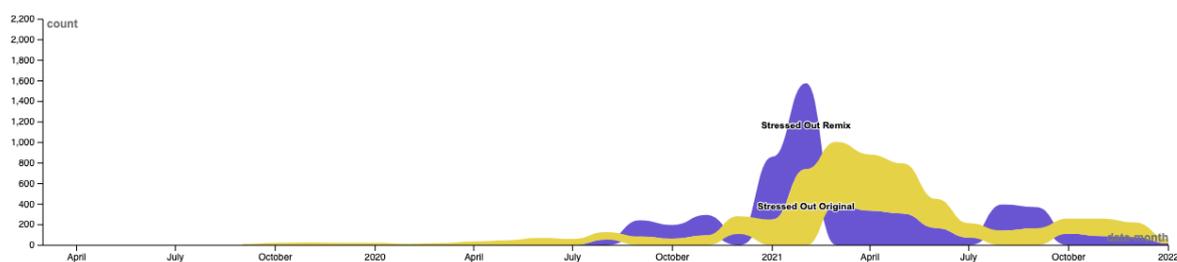
Figure 2. Stressed Out Original & Stressed Out Remix - number of posts per month.

The rise in popularity of this meme is visualized in the trend graph above (Fig. 2) and is much more significant with the remixed version of the audio. In the remixed version of the song, the vocals are pitched lower and reverb is added to create a more ambient sound which makes it complement the meme concept of nostalgic lyrics and aesthetic backgrounds better than the original song. As we will see, the original "Stressed Out" audio is more often used as background music to unrelated content whereas the remixed version becomes part of the narrative and memified.

4.2.2. Hashtag publics and subcultural appropriations

Collecting the top 25 hashtags for both the original sound and the remixed sound we created a temporal bump chart and a co-hashtag analysis to compare the emergent hashtags. In this way, it is possible to point out both the temporal changes of the audio meme emerging from the series of hashtags, as well as its sub-cultural uses and appropriations.

Firstly, most videos aim at virality by using tactics such as tagging #fyp, or other culturally agreed upon hashtags such as #xycba. This is an interesting finding as it indicates a trend towards the self-awareness of the posters, who tend to utilise the affordances of the platform to optimise exposure. To delve deeper in the subcultural use and appropriation of the meme, we run two versions of the chart, one with #fyp and other hashtags included, in order to investigate the prevalence of these hashtags, and a second eliminating them (Fig. 3).

By looking at the bump chart, we can see that many fandoms move in synchronisation with each other, i.e. a large amount of fan-created film franchises (#harrypotter) or gaming videos (#fortnite) being uploaded (Fig. 3). If we compare the bump chart to the co-hashtag network, it emerges that many users utilise these tags, which however do not overlap with each other, creating many separate clusters with a small number of unifying tags (some examples are shown in Fig. 4). Additionally, we identify the prevalence of regional hashtags in Spanish such as #tuhistoria and #aprendecontiktok (your history and learn on TikTok respectively). This indicates the rise of subcultures defined by not only interest but langues and also regions.

Figure 3. Stressed Out Original, Top 25 hashtags over time

Figure 4. Stressed Out Original, co-hashtag analysis, cluster close-ups.

We also noted a distinct difference in the evolution of trends in each version of the sound (original vs remixed). As shown in Fig. 5., tags such as #pain and #sad, evident in the remix, were not at all present in the top 25 of the original sound. Additionally, the co-hashtag analysis indicates that tags such as #sad served as central contact points with which many different subcultural groups mediated (Fig. 6). This indicates that in

some cases, there is a polarising use case of each sound. It is worthwhile to consider the possibility that Stressed Out Original and Stressed Out Remix operate differently, due to their origins, with the remix being a more native TikTok sound.

Figure 5. Stressed Out Remix, Top 25 hashtags over time

Figure 6. Stressed Out Remix, co-hashtag analysis, cluster close-ups.

The insights into the hashtag analysis highlight that the stressed out meme is being used by different subcultural groups, each of them represented by a set of specific and diversified hashtags (e.g. related to video games, movie fandom etc.) or kept together by the sharing of different feelings, such as sadness (in the case of the remixed sound).

It is important to notice, however, that the presence of these subcultural uses of the meme reflects what can be called a (subcultural) hashtag public (see, Arvidsson & Caliandro, 2016; Bruns & Highfield, 2016). Indeed, users that share a specific interest, an affective relationship to a topic, and possibly a similar subcultural background cluster around specific hashtags, which are devices for mediation, affect, and temporal aggregation rather than a sign of a strong and coherent community.

4.2.3. Image similarity analysis with PixPlot

Moving from hashtag analysis to visual similarity analysis, we can start addressing the visual component of TikTok video. In this case study, the display and arrangement of cover thumbnails in a cluster map with PixPlot was useful to start understanding the presence of different trends (for the advantages and disadvantages of using visual similarity analysis with PixPlot, see Section 5.3.3.). In this way, it is possible to grasp the changes in the meme and the different appropriations thereof.

Starting with the analysis of the original sound, we can see that the song 'Stressed out' is mostly used as background music for different kinds of content, in particular for thumbnails displaying activities such as painting and writing. A similar use of the sound can be found in images representing scenes of streets, driving and flying. Moreover, In line with the results of the hashtag analysis, we find plenty of subcultures and fandoms, and in particular, a huge cluster dedicated to Fortnite and other video games, which consists of gameplay, video edits/supercuts of film content etc. Moreover, the other two clusters consist of, respectively, representations of celebrities such as actors and K-pop idols, and cartoon anime.
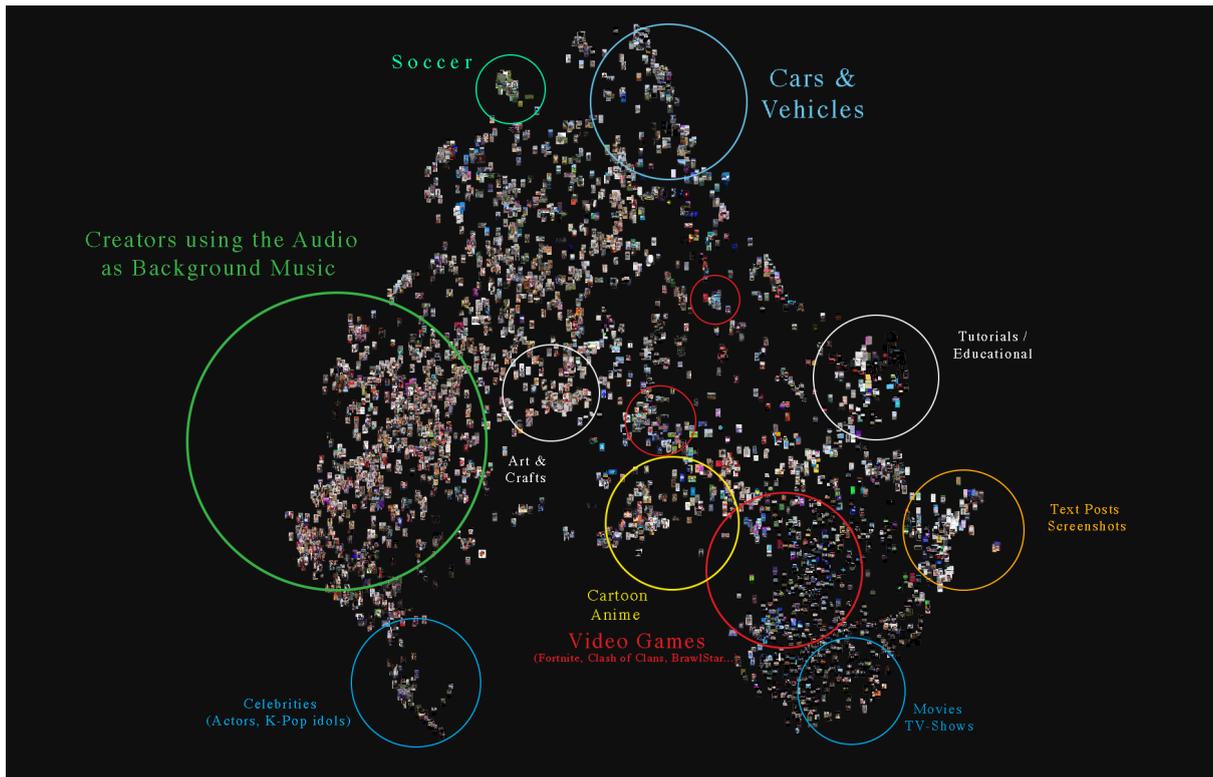
Figure 7. Stressed Out Original, PixPlot clusters, labelled

The Stressed out remix song appears to be used as background music too (e.g. in the case of the display of football videos and edits). However, this sound shows more trends/microtrends and similar video styles emerging from the thumbnails.

In particular, the 'nostalgia' trend, where users show images 'future generations will not understand' and the 'finally I understand this lyrics'-trend. Moreover, we can see a cluster dedicated to thumbnails suggesting the presence of aesthetic videos with beautiful pictures of the sky, the ocean, etc. accompanied by a meaningful quote. Such content goes hand in hand with the "slowed and reverbed" vibe of the song.
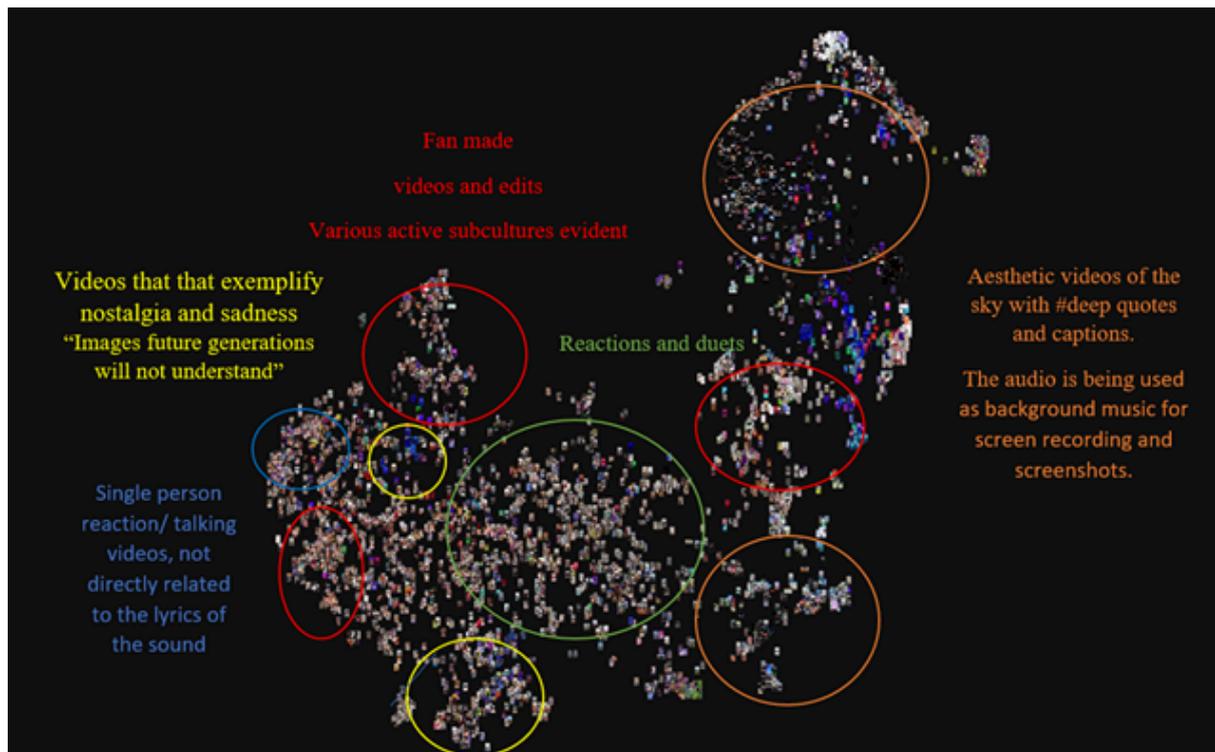
Figure 8. Stressed Out Remix, PixPlot clusters, labelled

The visual representations emerging from cover thumbnails present some similarities with the hashtag publics previously identified. Specifically, we can see a relationship between hashtags and visual representations in regard to different types of fandom, which is mostly present in the original sound. Moreover, in the case of the remixed sound, the representations about nostalgia couple with feelings such as sadness found in the hashtag analysis. These results further confirm that the 'Stressed out' sound is used differently in its two versions, and differently appropriated by specific categories of users, which find in the meme a way of expressing a variety of content, moods, and feelings.

4.2.4. Emerging templates

Considering the relevance of audio-visual content on TikTok, we rely on what we call 'templatability analysis' (see [Section 4.1.3.](#)) to consider TikTok videos as a whole by visualising their visual content in the form of video stripes. In this way, it is possible to understand what the emerging themes and templates related to the Stressed Out sound are. Templatability, that is, a set of norms and conventions in organising visual and audio content (Leaver et al., 2020), is a driving principle through which users' practices deploy and memes spread.

In this part of the analysis, we focus specifically on the remixed version of the sound, which, as previously seen, is the more 'memetic' one. We focus on the visualisation of the video stripes and analyse them qualitatively, in order to grasp the emergence of different themes and the presence of some recognisable templates. From the analysis, it was possible to identify the presence of three main themes, each one accompanied by a peculiar template.

- **Trend surfing**. As already mentioned, the sound is often used in order to increase users' visibility and create trending content. This practice, however, is not only aimed at acquiring a micro-celebrity status, but also at shedding light on contemporary issues. This is the case of a series of Tiktoks dedicated to sharing awareness about the political and economic situation in Latin American countries, such as Colombia, Chile, Mexico, and Peru. These videos are accompanied by hashtags such as '#soscolombia', and captions such as 'Latinoamérica resiste'. In these cases, the sound is reappropriated for activism purposes, to share information, and increase solidarity. This content shares a similar template (Fig. 9): it starts with the presence of the tiktoker, taking distance from other content creators and stressing the difficulties that Latin American countries are facing. What follows is a series of clips that symbolise these political issues. One peculiarity of this template, which also makes it recognizable, is the recurring representation of flags. The video then finishes by showing again the tiktokers, asking for support and solidarity.
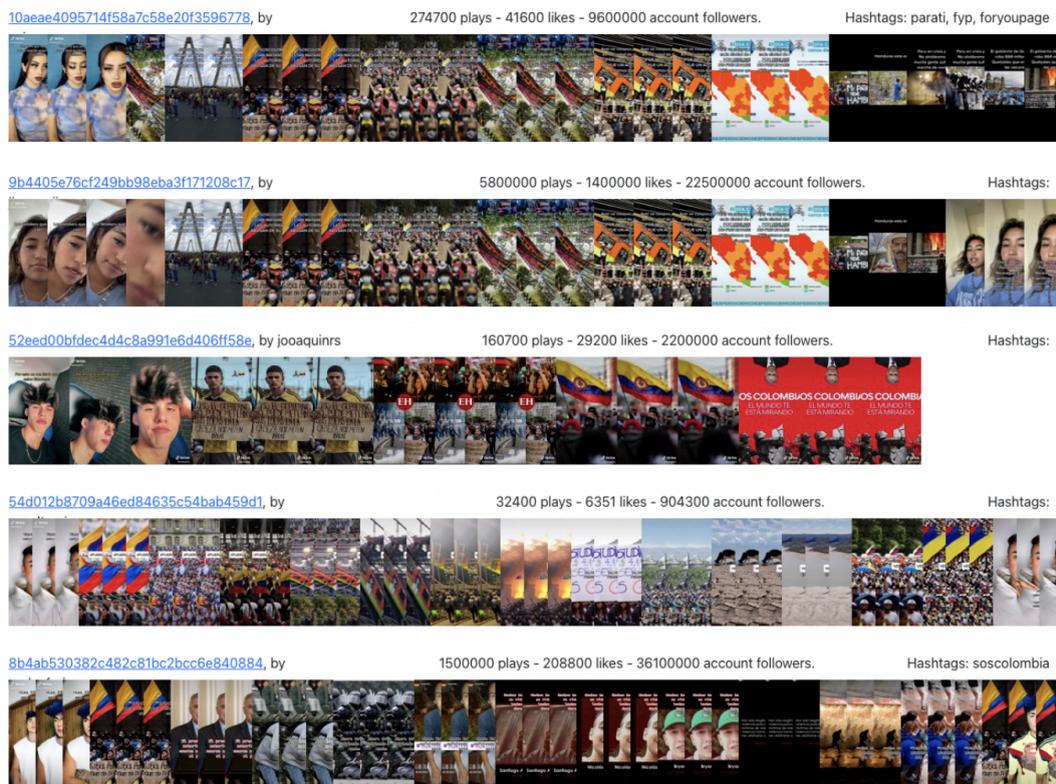
Figure 9. Image stripes - trend surfing

- **Nostalgia**. In line with the original meaning of the song and the results previously highlighted, the sound is used to convey a feeling of nostalgia. Firstly, the sound is used to share memories about the past, especially related to objects, cultural products, and feelings that the new generation won't understand. The feeling of nostalgia is characterised by a template that starts with the creators stating 'things that today's kind don't understand', or 'photos that the next generations won't get'. It follows a series of clips displaying such objects, cultural production (e.g. Toy Story is a recurring theme), images and photographs (Fig. 10a).

- Moreover, the results show that the sound can also be used to refer to the 'good old days' with dark humor. More specifically, some found some examples of the sound and the lyrics being used in a political context, either racist, sexist, homo/transphobic (Fig. 10b). From the data at our disposal, it is difficult to clearly point out the relevance of this kind of representation. Further research could investigate even more the relationship between nostalgia and radicalisation (see, e.g., Gandini, 2020)
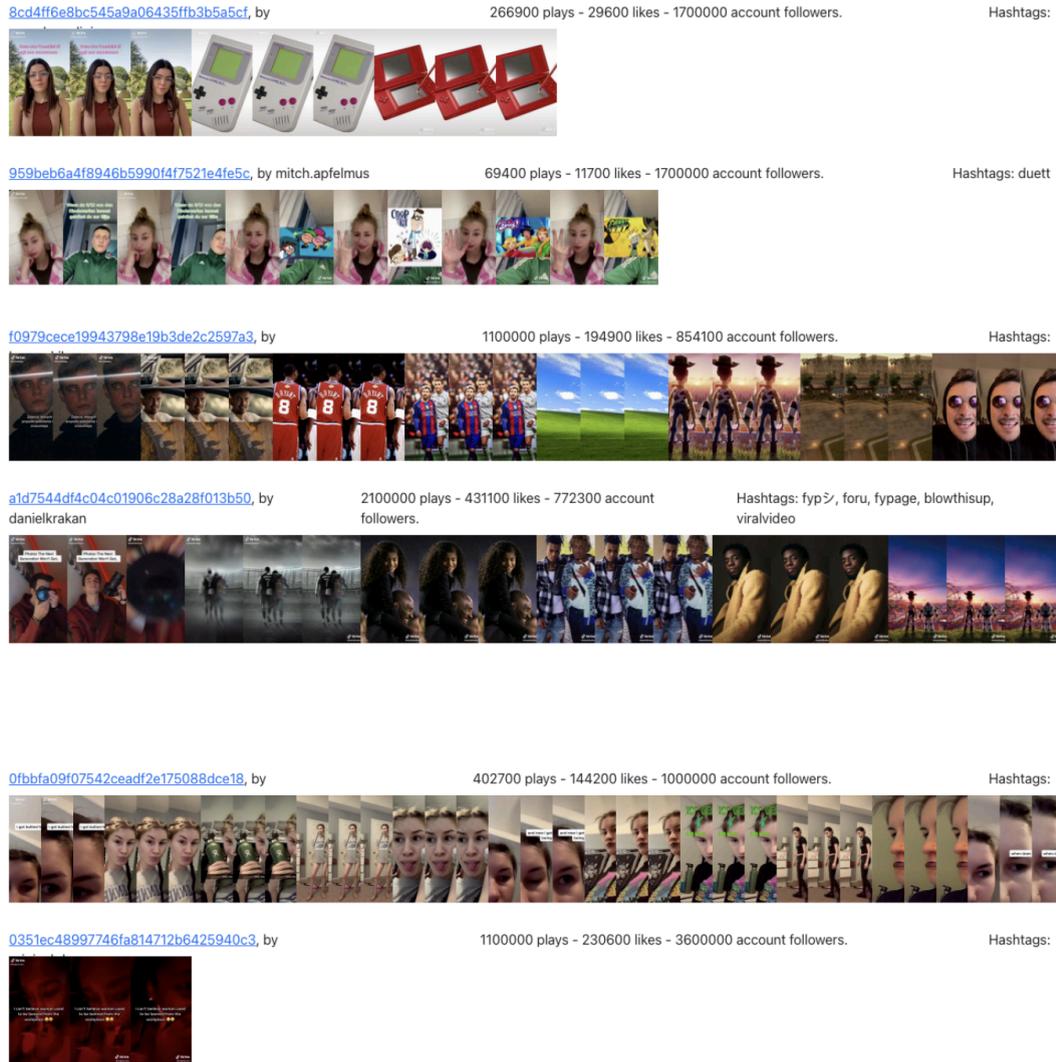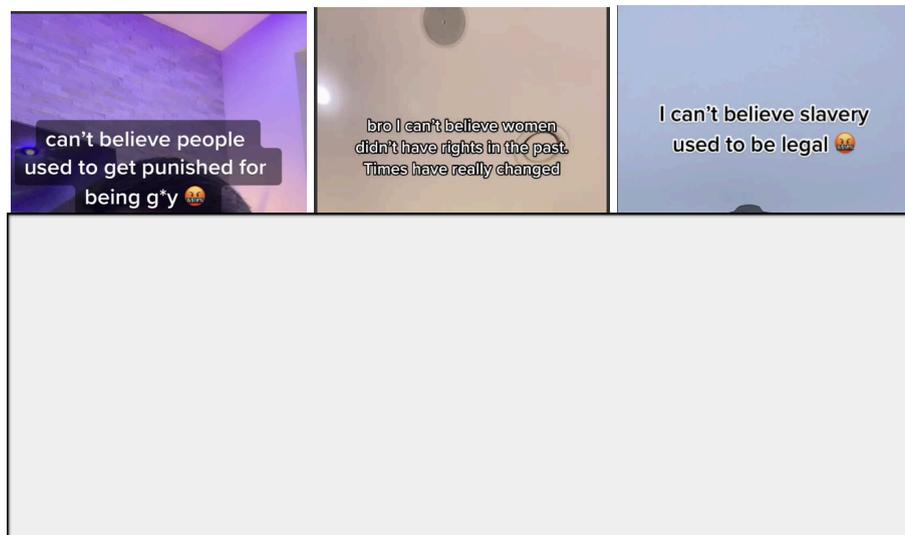
266900 plays - 29600 likes - 1700000 account followers.
Hashtags:

69400 plays - 11700 likes - 1700000 account followers.
Hashtags: duett

1100000 plays - 194900 likes - 854100 account followers.
Hashtags:

2100000 plays - 431100 likes - 772300 account
followers.
Hashtags: fypシ, foru, fypage, blowthisup,
viralvideo

402700 plays - 144200 likes - 1000000 account followers.
Hashtags:

1100000 plays - 230600 likes - 3600000 account followers.
Hashtags:

Figure 10a. Image stripes - Nostalgia

Figure 10b. Nostalgia and radicalisation, some examples.

- **Use of beat**. Lastly, the sound was used to show in a way that cuts and transitions of the videos are synchronized with sound beat. This way of using the trend is aimed at showing before and after representations (e.g. hairstyling, makeovers, etc.) and comparison (e.g. how the appearance of different actors change over time).
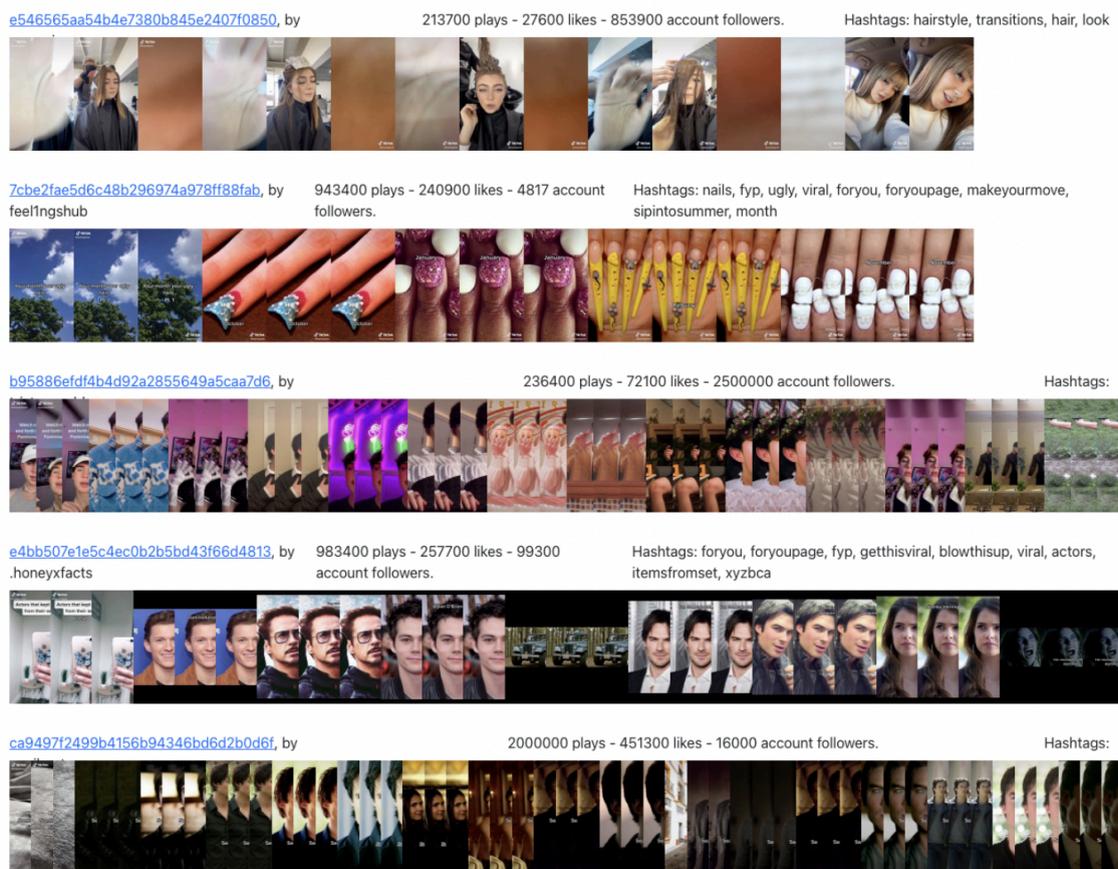

Figure 11. Image stripes - use of beats

4.2.5. Sound analysis with audio recognition

Having matched all Tiktok sounds with the original 'Stressed Out' song, we retrieved a similarity measure per Tiktok. Together with the quantitative platform-native measures 'followers' and 'likes', we have visualised how the sound quality–in terms of being diversified, voice-dominant, or close to the original–relates to platform-native popularity measures. For this, we have created two bar plots grouping TikToks with similar sound quality which are coloured by the amount of followers (Figure X), and the amount of likes (Figure XX).
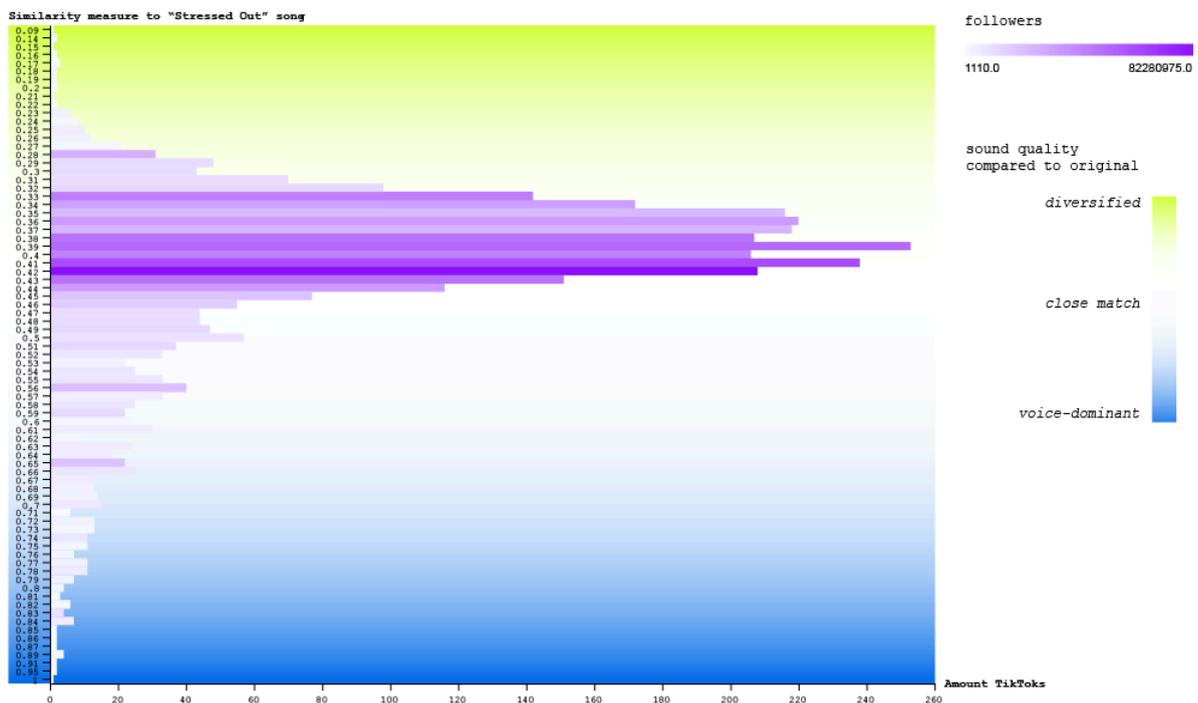

Figure 12. A bar chart showing the amount of TikToks per original-song-similarity measures between 0 and 1, ranging from diversified to close matching to voice-dominant sound quality. The colour intensity of each bar indicates the total sum of followers across TikToks per bar.

As a first finding, we can see in Fig. 12 that most of the 'Stressed Out' TikToks group around being a close match to the original song, given that the closest match possible has a similarity measure of 0.4 (see Section 4.2.5). Besides that, we can see that overall, more TikToks are voice-dominant instead of diversified (remixed or noisy). Moreover, we can see that the most followed bars group around a good sound quality which comes close to the original. Whilst it is obvious to expect that the more TikToks per bar, the higher the total sum of followers, we can nevertheless see that the bar with most followers is not the ones with most TikToks; instead, it is one where the sound quality is close to the original song.
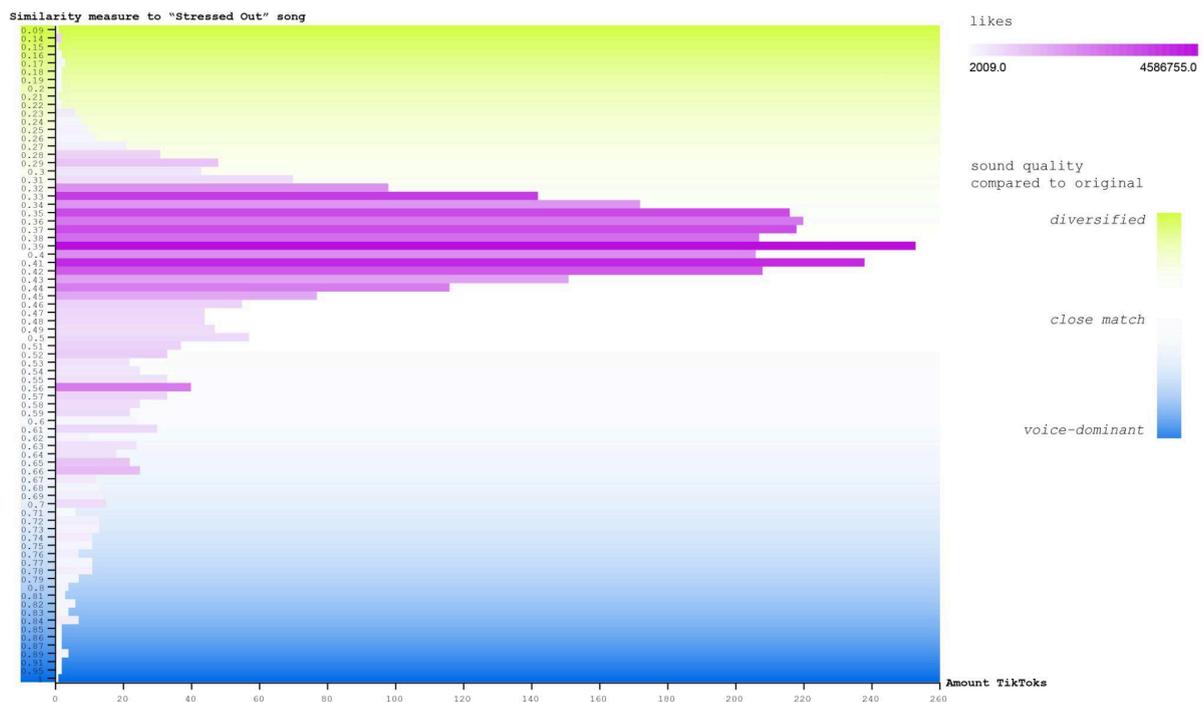
Figure 13. A bar chart showing the amount of TikToks per original-song-similarity measures between 0 and 1, ranging from diversified to close matching to voice-dominant sound quality. The colour intensity of each bar indicates the total sum of likes across TikToks per bar.

In comparison to the 'most followed' bars (Fig. 12), the 'most liked' bars group as well around closely matching the original song (Fig. 13); however, the amount of likes is more evenly distributed across centred bars irrespective of length (amount of TikToks). This shows that, a) the most followed 'Stressed Out' TikToks are not necessarily most liked, and that, b) the overall popularity in terms of 'likes' is more strongly tied towards matching the original song than the popularity in terms of 'followers', since bars with less TikToks have a comparable high amount of likes when being grouped around the centre. There is one exception for a bar located at a similarity measure of 0.56, showing that at least one TikTok which includes voice parts while still being relatively close to the original, got many likes. Given that the same bar has more followers as well comparable to its local neighbourhood, this could hint to one much-followed TikTok user having shared a much liked TikTok in which the user talks.

Comparing the findings of the most followed with the most liked TikToks grouped by sound quality, we can see that in both cases the most popular TikToks tend to match the original sound quality. This indicates that the songs shared around 'Stressed Out' are rather mainstream, as they do little remix or re-appropriate the original song, with one exception occurring for much liked more voice-dominant but still good quality TikTok(s). Moreover, the comparison shows that whilst both popularity measures are grouped around the original sound quality, much liked TikToks tend to be of more diversified

sound quality in comparison to the much followed ones. One hypothesis from this finding is that TikTok users that have shared 'Stressed Out' while having many followers make use of professional technical equipment and high quality sound; and that there are few users at the centre of attention. In comparison, likes are still mainly grouped around the original song quality, but with more fine-grained deviations across sound quality.

5. Case study (2) "Mama Said" sound

5.1. Data: scraping limitations and solutions provided

5.1.1. Initial data

In the first part of the analysis of the Mama Said case study, we relied on the data collected by means of the Zeeschuimer Firefox extension (see Section 2), which include a total amount of 3,926 tiktoks.

Considering that our study would focus on TikTok memes and that TikTok is a video-centric social network, video analysis was a necessary step, but we realised that the dataset provided contained expired video links. While we observed that the links were retrievable, this process could not be automated. Hence, the idea was to find a way to automate the process to get fresh links, and maybe more metadata to build up a qualitative *context layer*. For these reasons, the first data collection was coupled with a second one, which started with the exploration of different scraping techniques.

5.1.2. Exploring different scraping techniques

A new question emerged quickly: was scraping based on the "requests" and "Beautiful Soup" libraries going to suit our research path around the Mama said meme?

- *Requests* (https://docs.python-requests.org/en/latest/) is a HTTP library for the Python programming language. The goal of the project is to make HTTP requests simpler and more human-friendly.
- *Beautiful Soup* (https://beautiful-soup-4.readthedocs.io/en/latest/) is a Python package for parsing HTML and XML documents. It creates a parse tree for parsed pages that can be used to extract data from HTML, which in turn is useful for web scraping.

To answer this question, we needed to make a small test with a video determined above and to work on the filtering of the links of all the HTML pages.
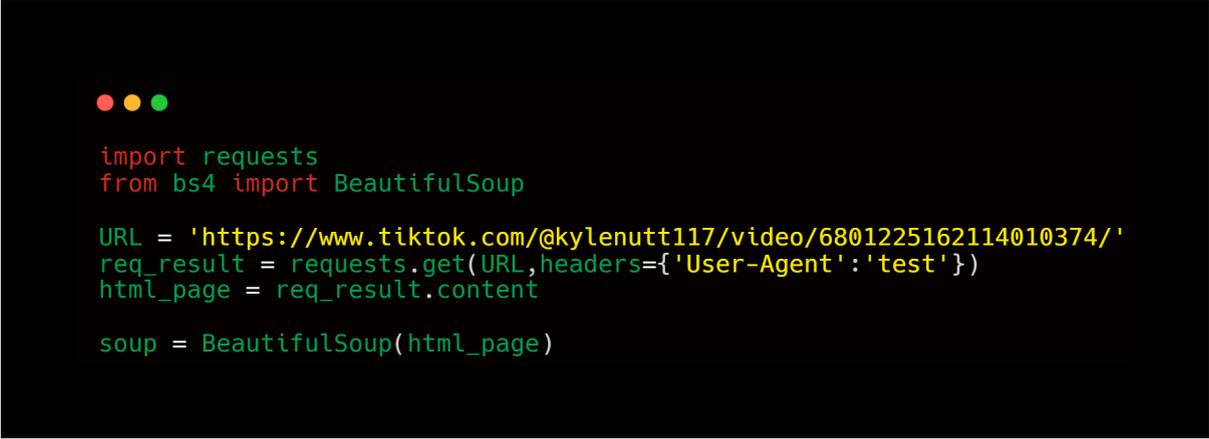
```python
import requests
from bs4 import BeautifulSoup

URL = 'https://www.tiktok.com/@kylenutt117/video/6801225162114010374/'
req_result = requests.get(URL,headers={'User-Agent':'test'})
html_page = req_result.content

soup = BeautifulSoup(html_page)
```

Figure 14: Tentatively re-scraping the data to obtain fresh metadata and videos, and a sound starting basis for exploration.

The result of this request was an HTML page with a set of tags, some of which are links to videos, as opposed to the link from the main video. Out of that, we understood that the TikTok video page was a dynamic page, and usually the video link is communicated by an API.

The solution was therefore to switch to a dynamic scraping method, and the "Selenium" library came to our mind. Indeed, Selenium Python bindings provide a simple API for writing functional/inbound tests using Selenium WebDriver. With the Selenium Python API (https://www.selenium.dev/selenium/docs/api/py/api.html), we could access all Selenium WebDriver features in an intuitive way.

While the method worked well, the first trials of scraping revealed a new issue: due to the link extraction and time required for video downloading, the links used expired meanwhile. We had therefore to scrape the links and download the videos at the same time, running the risk that the protection systems would ban our IP addresses, as an anti-BOT cyber-parade. We introduced an interval of one to two seconds between each request to slow the pace and perform an *ethical scraping* of the data, preventing a denial of service or an IP address ban.

```
from selenium import webdriver
from selenium.webdriver.chrome.options import Options
import urllib.request
Links=[]
driver = webdriver.Chrome(options=options, executable_path=DRIVER_PATH)
for i in data['tiktok_url']:
driver.get(i)
try:
temps = driver.page_source.split('tiktok-1h63bmc-DivBasicPlayerWrapper
e1yey0rl2')[1]
temps = re.search("(?P<url>https?://[^\s]+)", temps).group("url")
Links.append(1)
urllib.request.urlretrieve(temps, 'C:/Users/amine/Desktop/DMI/DMI Winter
School 2022 - TikTok memes/Data/Mama said/videos/'+i+'.mp4')
except:
Links.append(0)
urllib.request.urlretrieve(temps, 'videos/'+str(j)+'.mp4')
time.sleep(2)
```

Figure 15: Congratulations😬, we were able to extract and download all the videos of our meme song "Mama Said".

After performing an exploratory data analysis, we found out that some necessary data for further exploration was missing, namely :

- some data entries were not encoded properly,
- some metrics were not present (users metrics , video metrics),
- some data were duplicated.

So in order to have as many dimensions to analyse as possible, we decided to re-scrape the TikTok data from scratch. We decided to use a python package called "TikTokApi" and stored the data as a JSON-formatted file.

Opting to scrape the data using a programming language instead of a pre-made tool allowed us to have a much greater control in the scraping process.

```
1
2    import json
3    from TikTokApi import TikTokApi
4
5    api = TikTokApi.get_instance()
6
7    count = 1
8
9    # You can find this from a tiktok getting method in another way or find songs from the discoverMusic method.
10   sound_id = "222447171198074880"
11
12
13   tiktoks = api.by_sound(sound_id, count=count)
14
15   #print(tiktoks)
16
```

Figure 16: screenshot showing the us of the Tik Tok API library

5.2. Methodology

5.2.1. Testing TikTok research techniques

In the first part of the analysis, we replicated the techniques developed for the 'Stressed out' case study, in order to further test them and evaluate their efficiency in relation to the other case study. Replicating the same techniques can be relevant, especially if considering that, as anticipated in the introduction, the 'Mama Said' sound can be considered as a more native sound as compared to the 'Stressed Out' one. Indeed, the 'Mama Said' sound originated from outside the platform but was then largely publicised by the singer himself, firstly on Musically and then further more on TikTok.

More specifically, the analysis started with the following techniques:
- [Hashtag analysis](#)
- [Visual similarity analysis](#)
- [Video stack analysis](#)

After these first steps, we integrated the analysis with other approaches, which blend sociological investigation and computational research.

5.2.2. Blended and iterative approach: sociological investigation & computational approach

During the event, we adopted an iterative and expertise-based approach in order to answer the aforementioned research questions to the best of our knowledge. The following diagram illustrates the operating process we adopted throughout the data sprint.

Figure 17. Iterative workflow, visual representation.

We started by a short brainstorming session (typically at the start of the day) in order to come up collectively with the most pertinent ideas and approaches to answer the top questions of the day. We then deconstructed the ideas in order to elaborate a set of required steps to implement and evaluate the idea within a limited amount of time. Once the main steps elaborated and tasks divided, we assigned them to members of the team so that they match best with their skill-set and preferences.

At the end of the day we had an evaluation session where findings, limitations and roadblocks were highlighted, and decided what ideas to further expand upon on the following day.

Another important methodological choice we made was to blend a sociological outlook, based on observation, data collection and intuition, and computational and technical processes. This is especially illustrated by the elaboration of the audio memes categories after an NLP analysis of the hashtags (further detailed in section *5.2.3. Hashtags and clustering analysis* of the report) or the work completed around the semantic labelling (see e.g. section *5.2.4. Semantic labelling through computer vision: brainstorming and conception*).

5.2.3. Hashtags and clustering analysis

We decided to perform a textual analysis or Text Mining on one of the components of the data set, which was the body or "desc" data field, which refers to the description

written by the Tiktoker. It contains mainly hashtags and a small message that the TikToker wants to deliver to his followers.

As a quick reminder, textual analysis is a methodology that involves understanding the language and symbols in texts to obtain information about how people make sense of and communicate their lives and experiences on the platform.

Visual or written messages provide clues as to how communication can be understood. Often, messages are understood as being influenced by and reflecting larger social structures.

In our case, messages reflect and or may challenge the historical, cultural, political, and ethical contexts in which they exist. We had therefore to understand the larger social structures that influence the messages present in the text under study.

Before getting into our textual analysis, an identification of the most popular categories was necessary to help us categorise our videos.

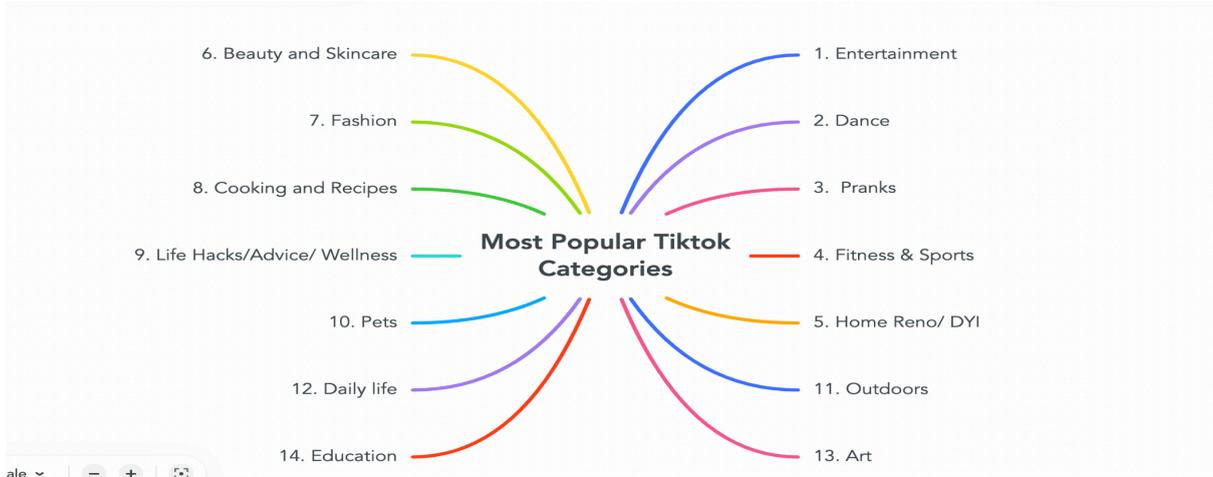In 2020, according to https://mediakix.com/blog/most-popular-tiktok-hashtags-categories-topics/, these categories were as follows :

Figure 18. most popular categories on Tik Tok. (source: https://mediakix.com/blog/most-popular-tiktok-hashtags-categories-topics/)

But we had to check these general categories on our particular research playground, the Mama said dataset, which would allow us to gain sociological insight at the same time.

The Tiktok API provided us with one of the most important components in the dataset: the hashtag. This variable mainly helped us in getting an overview of the most influential keywords related to the mama said challenge. We decided to carry out an analysis on hashtags in order to get to know the conversations taking place on the platform and the major present trends.

Hashtags are words or multi-word phrases that categorise content and track topics on different media platforms. Hashtags are preceded by the # symbol, such as #mamsaidchallenge in our case. People can search for posts with a specific hashtag. So, using hashtags helps the TikTok community find posts and topics that interest them. Users use hashtags so that their posts can be found out: adding a hashtag or two will help them to become visible to an audience.
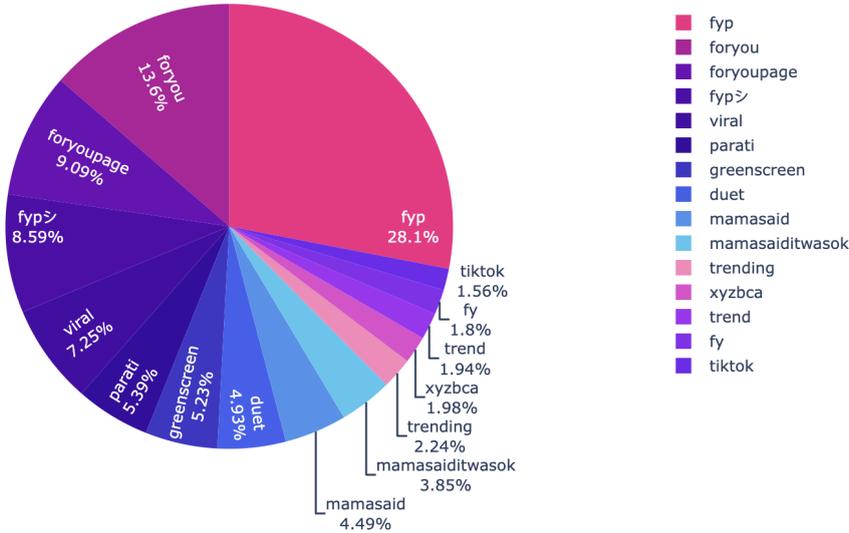

Figure 19 : word count of the Mama said posts by hashtag

As we could see from the word count illustrated in *Figure 19*, a large percentage of the hashtags in the dataset refers to #Foryoupage. On TikTok, the **For You** feed reflects unique preferences to each user. The system recommends content by ranking videos based on a combination of factors starting from interests that users express as new members of the community and adjusting for things that people indicate they are not interested in too, to finally form their personalised **For You** feed.

According to the outcome, we could say that the *mamasaidchallenge* took a considerable part in the feed of Tik Tok users which made it even more popular. The experiences mentioned in this trend actually range from childhood memories to explanations of life aspects. Which eventually led more users to try this challenge that made it go viral expeditiously.

This was an interesting intermediate result for us, that helped us to figure out categories and trends on Tiktok.

5.2.4. Semantic labelling through computer vision: brainstorming and conception

The main idea here was to find out a way to *summarise the semantics of a video* and came from the templatability concept whereby video stripes can be considered as a kind of composite image (Colombo, 2018).

While keeping in sync with our teammates working on the Stressed out meme, we wanted to take a step further in a slightly different direction: would the templates extracted tell us something about characters, objects and backgrounds that are staged by the TikTokers? How many sequences are there? What part is intentional, what part is not?

From there, we hoped we would be able to compare different videos and identify commonalities and differences in a computational way, considering the huge volume of material (4000+ videos): videos would speak *their own language*.

In the brainstorming session, we performed a first investigation of 10 sampled videos from a sociological perspective, and concluded that we could subsume the videos in three categories of growing complexity, namely:

- videos with a simple semantic: slow pace of the video (translating into a number of chained sequences limited to 2 or 3), each sequence being characterised by an homogeneous vocabulary over the duration of the sequence, as depicted by the colours (one lemme = one colour).

pole; unicycle; breakwater;
parallel_bars; seashore; bannister;
sandbar; bicycle-built-for-two; hook;
horizontal_bar

seashore; sandbar;
breakwater; wing; wreck;
promontory; lakeside;
container_ship; dock; liner

pole; unicycle; bow;
bicycle-built-for-two;
parallel_bars; sandbar;
seashore; crane;
aircraft_carrier; horizontal_bar

Figure 20 - Scene depiction: simple case

- **medium complexity videos:** the pace is quicker, but the semantic remains rather homogeneous throughout the video

Figure 21 - Scene depiction: medium complexity case

- **complex videos:** the pace is quick, and the scene depiction yields varying objects that do not match over time, and/or object extraction is not performing well enough.

Figure 22 - Scene detection: complex case

We then assessed that the extraction, whilst showing limitations, could yield additional valuable information, of a different nature, in comparison to the topic extraction from

TikTok postings, or from TikTok metadata including for example the soundtrack transcribed into text. Indeed, those latter pieces of information are intentional, whereas the semantic labels extracted account for a somewhat unintentional part of the video created by the TikToker: bodies, dressing up, elements of landscape, etc, that are also important cultural markers and, as such, could contribute to the question research around the expressive dimension or subcultural reappropriations of the memes.

As a next step, the process was designed as follows:
- We used the open-source AI Keras tensorflow model (see https://github.com/keras-team/keras and www.keras.io) implemented by Memespector (https://publicdatalab.org/2021/10/27/memespector-gui/) to perform the scene depiction. The google vision API was not used as we did not get access to it, but this could be a source of improvement (e.g. in our medium complexity case, Rihanna could possibly be recognised by the Google Vision API face detection).
- This allowed us to transform videos posted by TikTokers into collections of frames. After some initial elementary testing and validation, the extraction took a full night to be performed between day 3 and day 4.
- Each and every collection of frames was then injected into a dataframe. Memespector also returns probabilities coupled with objects found within each depiction. However we chose not to use them in a first exploratory approach in the interest of time. But this would probably bring interesting improvements in the label ranking algorithm (top 8) described hereafter (see section *5.3.6. Semantic labelling extracted from videos*).
- From our qualitative investigation we got the insight that the gathering of the collection of extracted labels obtained into a single frame would allow us to best summarise the information, because this is what we all did intuitively in the brainstorming session: look at how many times such word was occurring, if such or such object would be present and recognised by the Memespector AI throughout the video.
- Since the process showed limitations for complex videos, we decided that hashing and weighting would be explored on day 4, and if conclusive, added to the context layer to feed the final data visualisations and interpretations.

On day 4, we used 3 videos and tried out Top 3 to Top 10 most frequent occurrences in a spreadsheet (see *Figures 20, 21 and 22*) and concluded from our visual inspection that the Top 8 was probably the best compromise in terms of ratio *semantic information summarised* over *processing time and information load*.

With more time, this finding would deserve a deeper investigation, analysing scientific publications[1], and defining some metrics that would enable the measuring of information loss through a benchmark.

But for the time being, our frugal process was ready to be automated in Python.

5.2.5. Other Approaches brainstormed to categorise Tiktok videos' soundtracks

An approach we thought of while exploring the sound dimension of the Mama said meme was the analysis of the video soundtrack to discriminate among three categories:
- Whether the video soundtrack contains the song "Mama Said",
- Whether it contains the "Mama Said" tune, plus "other voice", or
- Whether it contains only dialogues or voices without any song.

Among the methods to be explored in this analysis, as depicted in *Figure 23*, we could opt for two paths.

The first path was to rely on a speech to text technique, followed by a text analysis:
- The videos containing the lyrics of the song "Mama said" with or without other phrases are categorised as memes.
- Videos that come up with empty output or only sentences distant from "Mama said" lyrics are considered as trend-surfing.

The second path was to make a signal matching, to get the same results as from the first method.

Both methods were found time (and processing) consuming, we had therefore to interrupt them in the framework of the data sprint.

---

[1] This topic is largely researched in the field of *video semantic segmentation* and uses for example Pytorch implementations to segment videos offline or in real time. A literature review would definitely be fruitful here, but the data sprint format did not offer us the necessary time span.
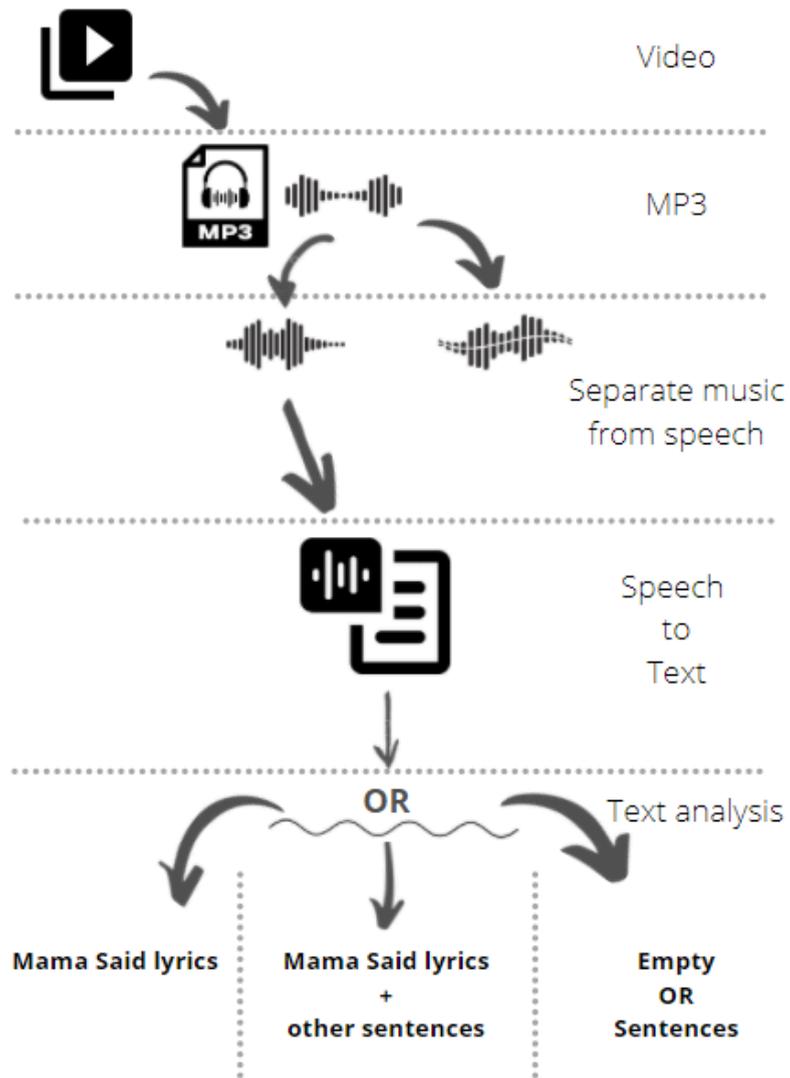
Figure 23. 'Process brainstormed to categorise Tik Tok videos as memes or "trend-surfing" postings, based on soundtrack analysis - the process was however not fully implemented

5.2.6. Integrating all data related to Mama said in a single context layer

On day 2, after completing the re-scraping of the Mama said dataset (see section <u>5.1.2. Exploring different scraping techniques)</u>, we then created a data pipeline in Python language in order to distil its content and select relevant data for our analysis.

The rescraping phase was time consuming but was performed in parallel and did not block any other active task, as we implemented a simplified version of one of the agile methods called Weighted Shortest Job First (WSJF).
And the benefit was definitely rewarding as our reflection moved on: a sound qualitative dataset was being built up and enriched with our new findings.

From day 3, as we obtained the first insights, we realised there would be a *great interest* to integrate data resulting from our different research tracks in the same "mega-frame":
- NLP-based topic categorisation,
- Tiktok metadata,
- semantic labelling,
- sound extraction (although this latter was not pursued).

The figure 24 below stages the "dreamed" columns of our mega-frame:
- on the left, first two columns: metrics,
- then in the middle, next two columns: video content (semantic labels, and Tiktok metadata),
- Then on the last four columns on the right, NLP-based categories: some geo-localised information to track the subcultural part (which would prove a dead-end), community or identity-oriented (LGBT, celebrities, etc), tiktok categories (Social challenges etc) and intention-oriented category (whether a video expresses an opinion, is personal, trend-surfing, an ad etc)
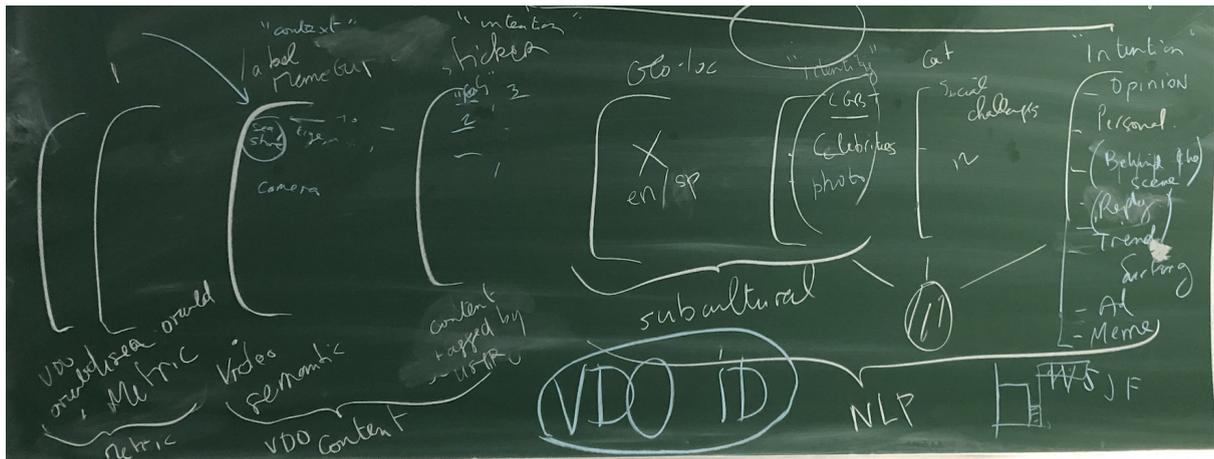
Figure 24. A photo of our board taken on the fly: at that moment of the brainstorming session, we start to bring together the pieces of the puzzle game… see also a simple drawing (at the bottom right corner) of the WSJF approach

At that moment, we thought that the different sources would bring different benefits, drawbacks and their cross-over might lead to new opportunities to make sense out of the posts:

- NLP-based topic categorisation, resulting from hashtags and users descriptions, bear what the user wanted to convey in each individual story,
- Semantic labelling tells us something about the context of the user, and is partly unintentional,
- Tiktok metadata bear also this information but link the different stories together over time (timestamps), since videos are often responses to other videos.

As a next step, we split the data according to our categories to obtain single spreadsheets containing all possible dimensions that may give us additional insights, and be used in parallel by the team members.

mama_sad_new_scrap_author.csv
mama_sad_new_scrap_authorStats.csv
mama_sad_new_scrap_challenges.csv
mama_sad_new_scrap_duetInfo.csv
mama_sad_new_scrap_effectStickers.csv
mama_sad_new_scrap_main.csv
mama_sad_new_scrap_music.csv
mama_sad_new_scrap_stats.csv
mama_sad_new_scrap_stickersOnItem.csv
mama_sad_new_scrap_textExtra.csv
mama_sad_new_scrap_video.csv
mama_sad_new_scrap_warnInfo.csv

Figure 25. Mama said integrated dataset with categories: how the dreamed mega-frame becomes reality, step after step

From day 4, our dataset matured on, and on day 5, final data amounted to 4379 lines and 30 columns, and would allow us to work out visualisations and interpretations without refactoring of the data: the effort spent initially on data re-scraping and curation was indeed rewarding during the final sprint.

Figure 25. Snapshot of the Mama said final integrated dataset

5.3. Results

5.3.1. Tracing the development of the Mama Said TikTok sound

"Mama Said" is a 2015 song from the Danish pop band Lukas Graham. TikTok is famous for using music as memes to illustrate situations in which people can relate to. The song was originally released in 2014, and gained popularity on the app Musical.ly in 2016. In this year, Graham himself played a role in popularising the song on the platform, by launching the 'official' dance related to the song, a series of thumbs up gestures that, when synched to the lyrics "mama said that it was okay/mama said that it was quite alright" creates a kind of literal, interpretive dance.

After this first phase, the 'mama said' meme peaked in popularity in September 2021, amassing over 1.1 million posts to date, but it previously saw boosts in engagement from February to March 2020 (Fig. 26).

Interestingly, in the beginning of the trend, the song was mostly used to explicitly speak about a mother/child situation but quickly it was used to illustrate any kind of situation that was involving judgement from society regarding a certain type of behaviour. From this point, the main usage of the meme was to show how people deviating from the societal norms (sexual deviance, not having kids, etc) could anyway feel validated through other people's experience and support. With more time passing, the song begins to serve new purposes, not necessarily connected to the initial trend. From a first qualitative overview of the trend and the content associated to it, we can see that in boosts of TikToks, many TikTokers used the meme to depict, and ostensibly justify, a variety of imagined violent episodes (e.g. one of the most popular videos depicts the user assaulting another person to protect his sister, as the lyrics "mama said it was okay" play in the background).
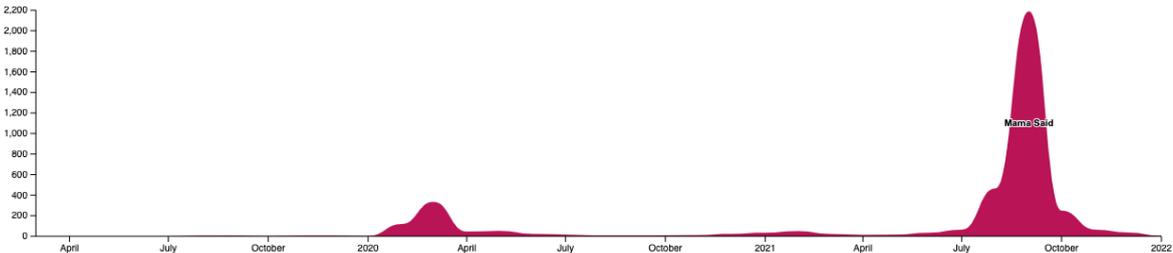
Figure 26. Mama said - Number of posts per month

5.3.2. Detecting hashtag publics

Similarly to the Stressed Out case study, one of the first steps of the analysis is to analyse the genealogy and change of a TikTok meme through hashtag analysis. To this end, we carry out a co-hashtag analysis and visualise it with Gephi.
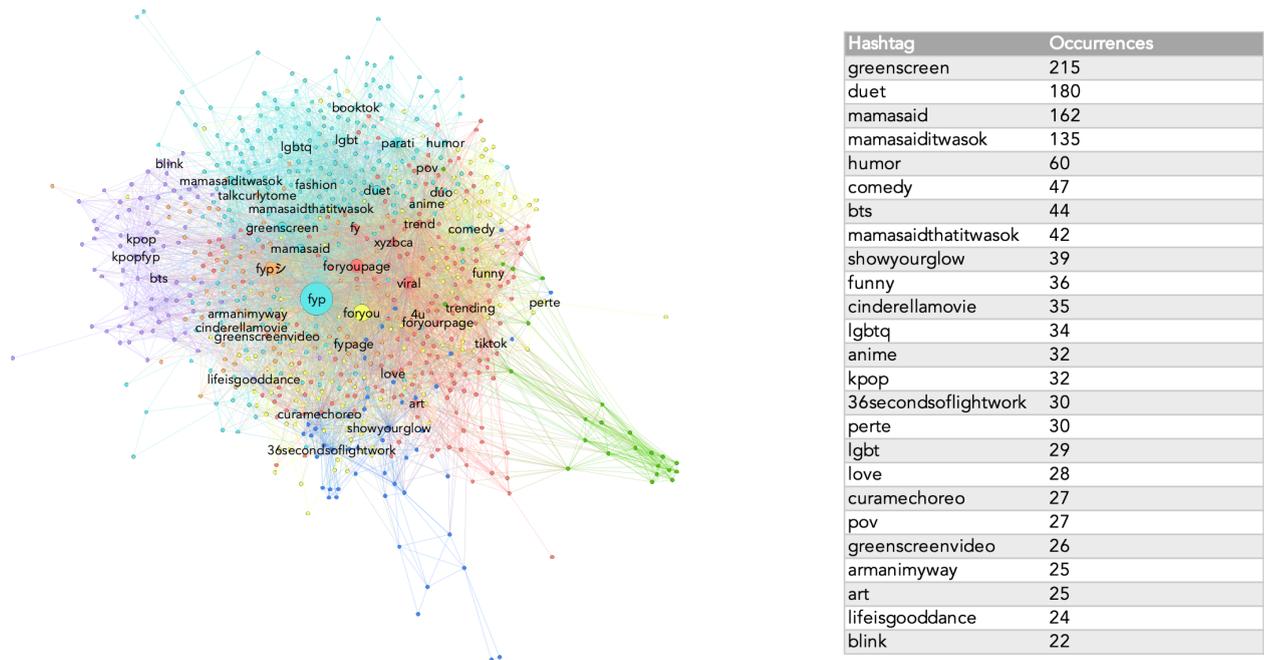

Hashtag	Occurrences
greenscreen	215
duet	180
mamasaid	162
mamasaiditwasok	135
humor	60
comedy	47
bts	44
mamasaidthatitwasok	42
showyourglow	39
funny	36
cinderellamovie	35
lgbtq	34
anime	32
kpop	32
36secondsoflightwork	30
perte	30
lgbt	29
love	28
curamechoreo	27
pov	27
greenscreenvideo	26
armanimyway	25
art	25
lifeisgooddance	24
blink	22

Figure 27. Mama said - Co-hashtag analysis and most-occurring hashtags.

When we look at the network, the most distinguishable clusters are the "lgbtq" cluster and the purple "kpop" cluster.

In the case of the 'lgbtq' cluster, we can notice nodes like #comingout, #pride, #nonbinary, #lesbian, #gay, #trans, & #bi, as well as hashtags like #loveyourself. This could indicate that the 'mama said' trend encourages more somber and supportive content than others on TikTok. When we put this in context with the meme itself, the TikTok environment can be read as largely being re-affirming and unifying.

The Kpop cluster consists of nodes that reference K-pop groups, artists, fandoms, and related media, with hashtags including #bts, #btsarmy, #blackpink, and #straykids. Similar to other celebrity hashtags found in this trend, many of these TikToks position the K-pop artists as 'mama,' implying that the K-pop artists have influenced their creators to willingly accept or appreciate something in their own lives. The presence of

these clusters points out the presence of two relevant hashtag publics with a similar subcultural background.

Moreover, hashtags related to advertising campaigns, stealth marketing, and product placement are widespread in the 'mama said' trend. As an example, the brand Armani is visible in our top hashtags, though others include #shiencares and #katespadenyonthemove.

Finally, "greenscreen" is the most frequently occurring hashtag once other attention-seeking labels are eliminated, which reflects the ways in which Mama said TikToks are edited, as well as a common characteristic of TikTok content more generally.

5.3.3. Visual similarity analysis: pros and cons of PixPlot for the analysis of TikTok videos

After this first overview of the main hashtag publics surrounding the mama said meme, we explored different possibilities for analysis TikTok visual content. Among those tools, we used Pixplot to try our first image analysis. We used Pixplot on two different datasets: the first one was the thumbnail picture dataset and the other one was the final frames dataset we analysed. We will focus here on the analysis of the first dataset only, which allows us to raise some interesting insights related to the research questions.
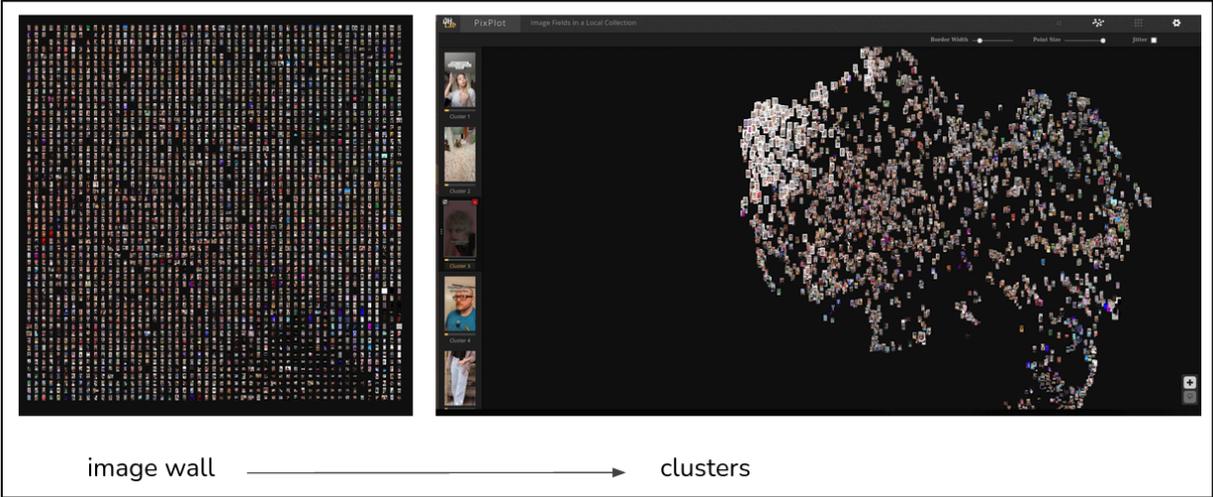

Figure 28. PixPlot: an example of image wall vs cluster visualisation.

In the above picture, we can see the result of the thumbnail PixPlot process. The result came with some indication about the clusters. 10 clusters were identified by the

algorithm but if some of them were pretty obvious (multiple people, women's face, outdoor scene...) most of them were difficult to interpret. Here is a couple examples of the clusters that were possible to analyse (Fig. 29):
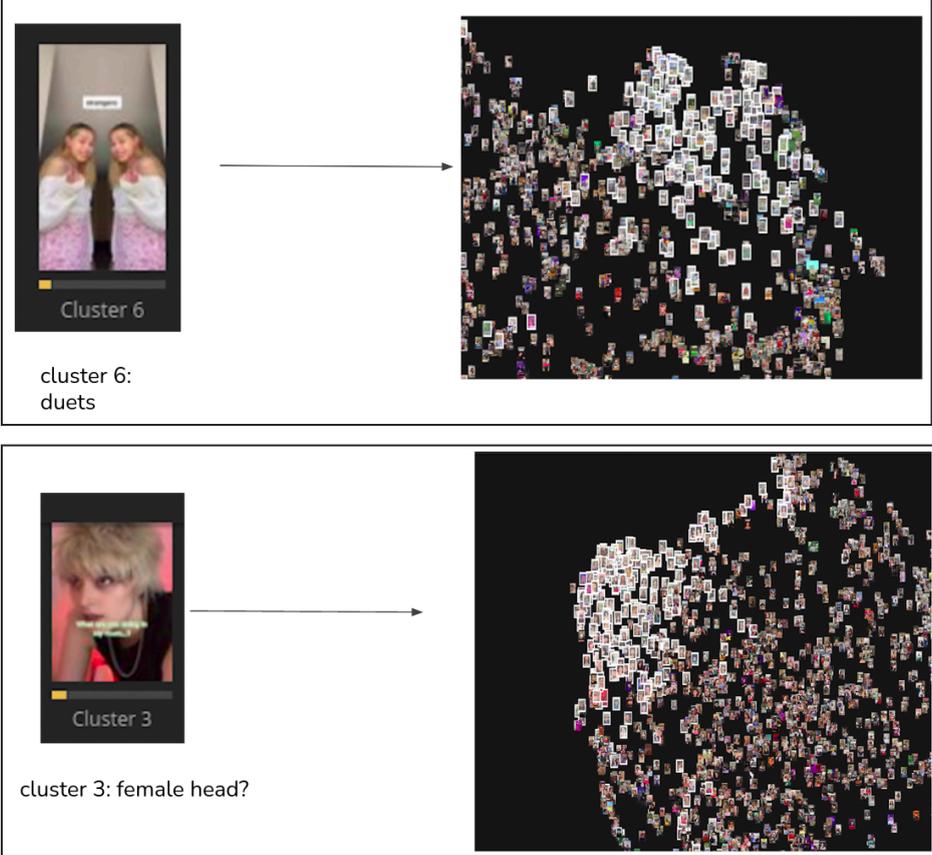

Figure 29. Mama said PixPlot clusters, some examples

Delving deeper into the analysis, we tried to label each of the clusters, in order to highlight the similarities and themes emerging from the image wall (Fig. 30).
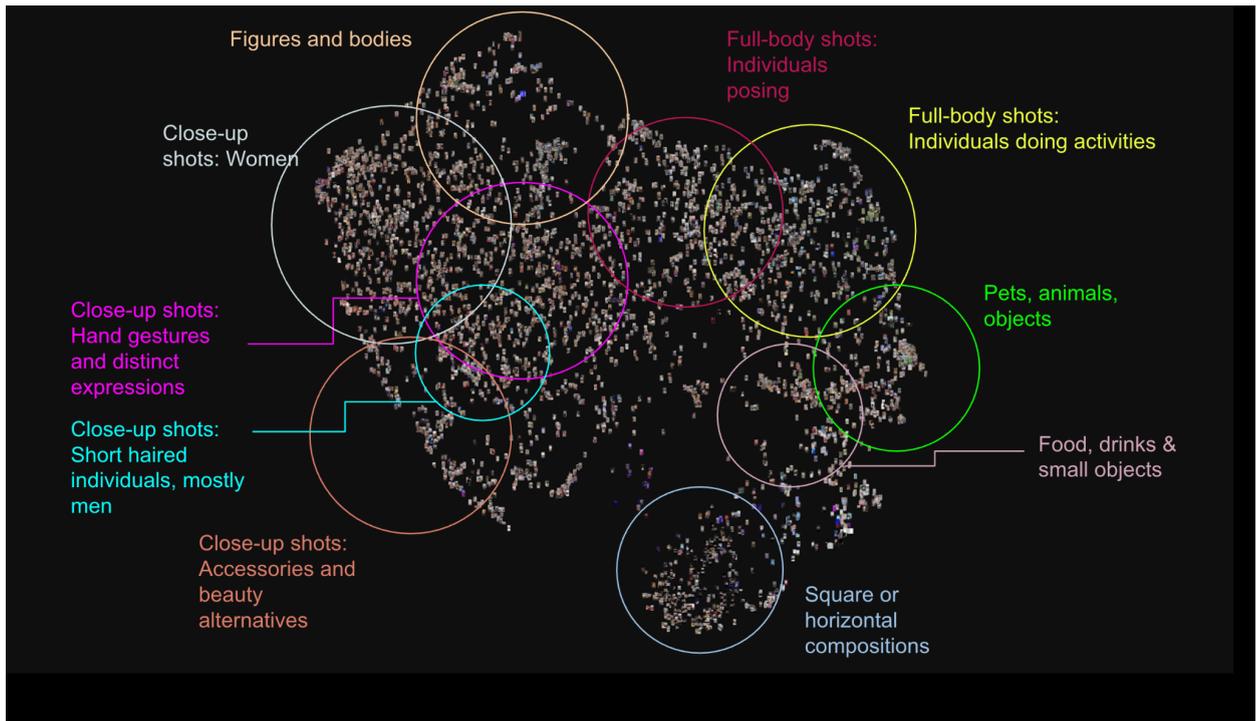
Figure 30. Mama said - PixPlot clusters, labelled.

From this visualisation we can see that the main clusters represent close up-shots, and specifically displays: women, hand gestures, and facial expressions, but also accessories and beauty products. Moreover, another relevant part of TikTok thumbnails consists of full-body shots, representing individuals posing or doing activities. By looking at TikTok thumbnails, the visual imaginary which characterises the mama said audio meme is dominated by selfies and self-portraits. If we zoom into each cluster, we can have a more fine-grained representation of some of the main themes addressed with the mama said audio meme, such as body positivity content and makeup tutorial (Fig. 31.). We can notice that these two visual clusters are in line with the hashtag publics previously identified in the hashtag analysis (see Section 5.3.2.).
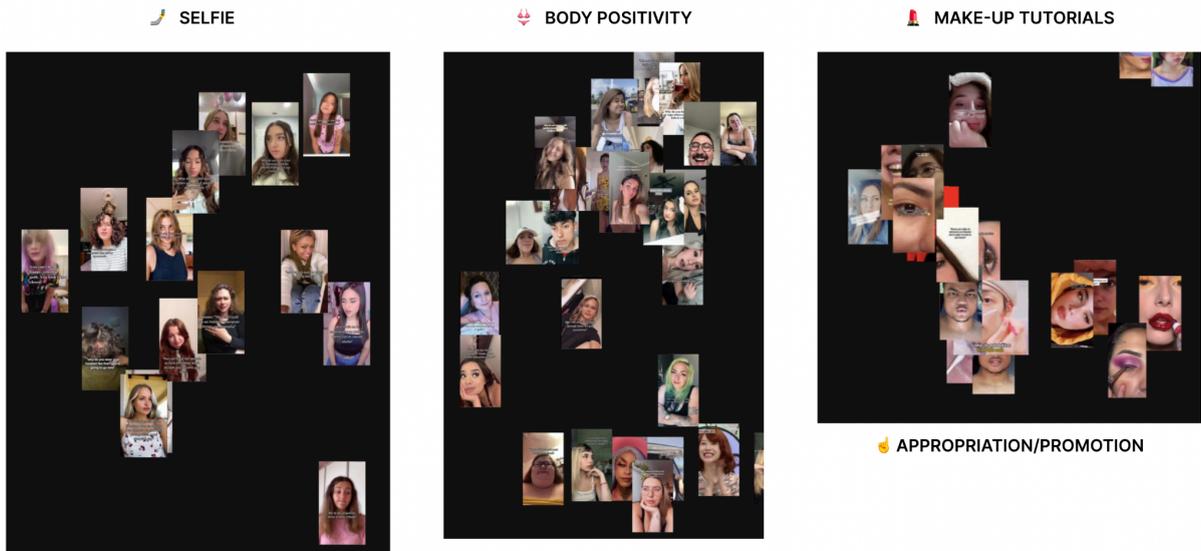
Figure 31. Mama said - PixPlot clusters, close-ups.

As we hinted before, visual similarity analysis of TikTok content can be helpful to grasp the visual trends present in video cover thumbnails. The results emerging from this case study are particularly useful to highlight the limits of this research technique, as well as the pros and cons of PixPlot. The 'mama said' audio meme consists of two main parts: a selfie introducing a problem or an issue, which is then followed by an array of role models who encourage and inspire creators to express their individuality, even if this means challenging existing stereotypes (for an example, see Fig. 32).
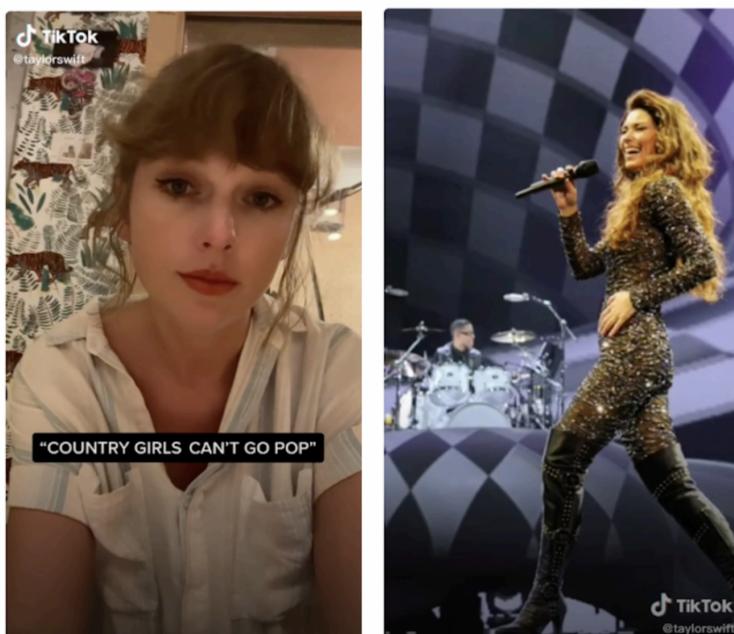

Figure 32. The two parts composing the mama said audio meme, an example. Source: @Taylorswift

Given its specificity, PixPlot only displays a static account of video content, and, more importantly, does not account for the entire video, which changes consistently from the first frame to the subsequent ones. In this way, image similarity analysis with PixPlot is problematic, because the interpretation emerging from the grouping of images in a metapicture overlooks some important elements of the object of study. On the other hand, PixPlot can be used to move from the aggregated image wall to a more qualitative and fine-grained analysis of the visual content. First, PixPlot allows to sort images according to different metrics (e.g., hashtags, etc...) and evaluate how the clusterisation changes. Secondly, for each thumbnail PixPlot provides a direct link to the TikTok website, a feature that allows for the manual exploration of each clip.

5.3.4. Video stack analysis: trend dilution

In order to go beyond these limits, in the 'mama said' and 'Stressed Out' case studies we try to account for both the video and dynamic content by means of video stacks.

The overall 'top' video stacks, from the most viewed videos and most followed accounts, revealed a rather unconnected selection of clips. Here, it can be seen that the sound clips primarily function as a, likely 'any', popular song to underscore what is happening in the video. Moreover, these videos can be interpreted as participating in 'trend surfing'; the using of a popular sound in order to garner more views, mostly regardless of what is happening in the video. Yet, certain overlapping motions can be distinguished, primarily on the beats of the sounds: these transitions happen at the exact same moments in several videos.

As noted previously, trends of both case studies are most accurately reflected three months prior to peak popularity. During peak popularity, a more unconnected selection of videos appears present: one where motions and other visual elements do not reflect one another. This reflection of the trend can be said to allow the trend to continue to grow, as can be seen from the popularity chart as well. Yet, once the sound hits peak popularity, its top videos no longer represent the trend: the sound enters 'trend surfing' territory, and subsequently the sound decreases in popularity as the trend mostly ceases to exist.

All the video stacks are available at this [link](#) (slides 18 and 25)

5.3.5. Networks of hashtags linked by videos posted on Tik Tok over time

One of our objectives was to understand the dynamics of a video on a social network such as Tiktok. Therefore we figured out that to understand the dynamic we would need two relations: a time variable and a variable to connect the videos between them.

The time variable was not difficult to obtain as it was present in a Unix Timestamp in our dataset, we just had to decode it to a date-time variable. The connection between the videos on the other hand was not so instinctive and we had to decide what would be the link between them. From that point, we decided to extract the hashtags from the videos and use the IDs to connect the hashtags. Simply put, in the network, each node represents a hashtag used in multiple videos. If two hashtags are connected, it means they are quoted by the same video.

We can see on the network that the most used hashtags are mostly what we will call "essential" hashtags. By "essential" we mean hashtags that are almost always cited in the videos. We can make a first hypothesis on why those hashtags are systematic: they help virality and visibility of the video. To verify this assumption, it could be interesting to make a comparison of the views on videos with and without those hashtags. Analysing the results would allow us to check if the use of those "essential" hashtags makes the video more visible just in the same way using a popular song will boost the visibility of the users.

The time analysis of the network allowed us to see two main results: first, we could see that the trend just keeps getting popular and that it doesn't seem to slow during the time period we studied. Secondly, we could clearly see the moment the song goes viral on Tiktok, as the network suddenly gets bigger. This behaviour can be linked to the usual virality of internet phenomena. Most of the time it is difficult to actually understand why a video or image goes viral.

5.3.6. Semantic labelling extracted from videos

Following our brainstorming and conception principles described in [5.2.4 Semantic labelling through computer vision: brainstorming and conception](), we moved on to the implementation step.

According to the output of the system of detection objects in the frames of the video is a multiple column spreadsheet (CSV format), but we were only interested in two types of

important columns, the name of the file and label_context columns (the objects detected in each frame).

The files follow a regular pattern name-wise: video_scene_frame_image.jpg.

Figure 33. Building up the semantic labels library.

Then we grouped the videos with all the existing labels in a single column. A simple algorithm based on frequency count of labels in decreasing order would determine the top 8 labels for each video.

Figure 34. Illustration of an excerpt of the library containing videos ids coupled with their semantic labels, and the most frequent semantic labels (top 8).

Then we extracted this dataset and made it ready for the creation of the networks, to form edges linking the nodes: we had our semantic labels that could be integrated into our dataset, associated with each video.

5.3.7 Natural language processing based approach to categorise TikTok videos: an intermediate result

We carried out a Natural language processing (NLP) Analysis using Python to analyze our data. This analysis refers to the branch of computer science and more specifically, the branch of artificial intelligence or AI concerned with giving computers the ability to understand the text and spoken words in much the same way human beings can.

A) Sentiment analysis

We started our analysis with a Sentiment analysis to identify and extracts subjective information from the description of the videos. The finality was to understand the social sentiment of the users based on their video's description or body.

Figure 35. Common words among most positive and negative posted bodies based on sentiment of users for the Mamasaid dataset.

The Word Cloud above represents the common words among most positive and negative post bodies based on sentiment of users. We could notice that the ForYouPage is particularly present on both negative and positive videos description. We would get to explain the reason shortly after our next analysis.

B) Topic Modelling : 12 videos categories

In this case study, our main objective was Topic Modelling or Topic Extraction, we wanted to know more about the categories that are most influencing on Tiktok and hopefully get a classification according to two variables "desc" and "video_id".

Essentially, we tested two main NLP techniques which are:

- LDA: Latent Dirichlet Allocation for topic modelling and was amazed at how powerful it can be and at the same time quick to run.
- LSA: Latent Semantic Analysis which involves creating structured data from a collection of unstructured texts.
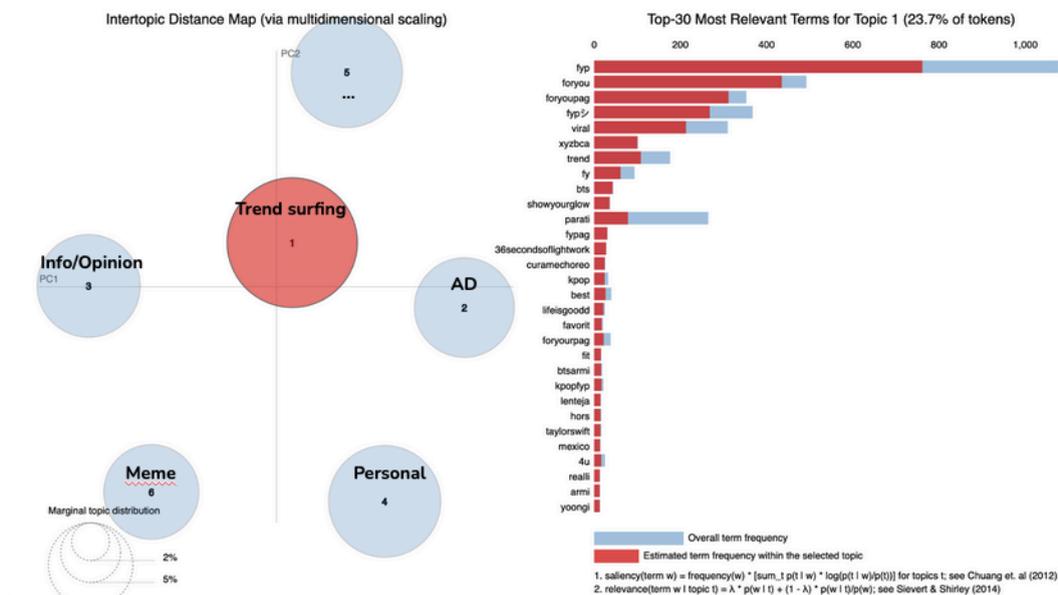
Figure 36. Topic extraction using LDA technique for the Mamasaid dataset, showing 6 clusters, mapping on guessed categories.
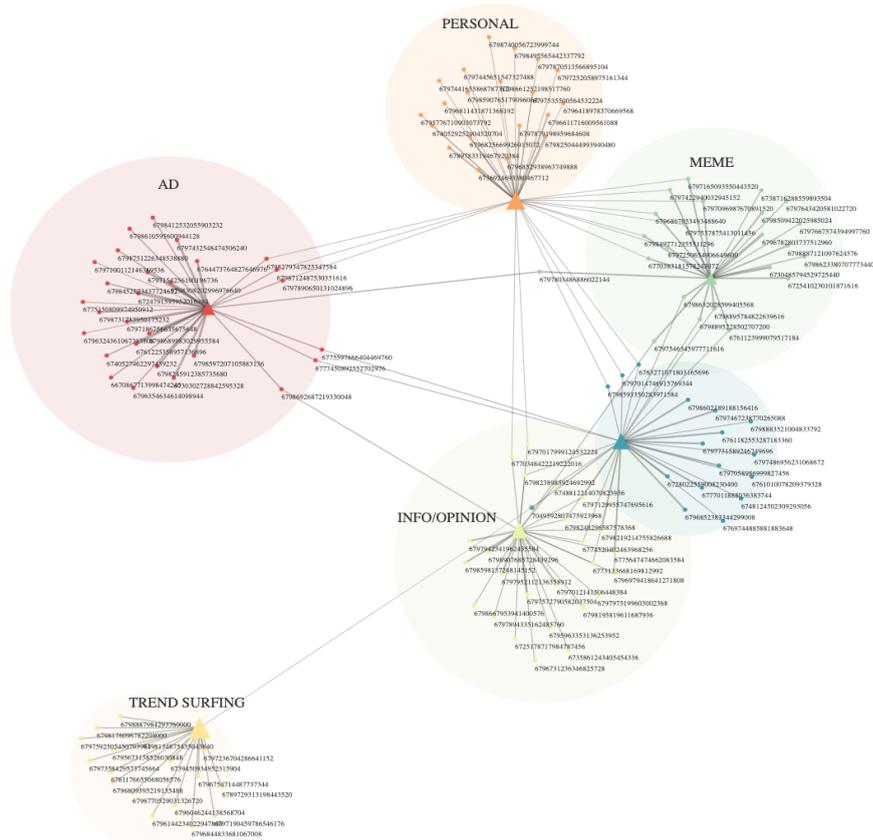
Figure 37. Hashtags co-occurrences for the Mamasaid dataset, showing 6 clusters, mapping on guessed categories.

According to the results of both techniques combined, we could not get a clear classification of the videos based on their categories, although some categories started to emerge. This made us conclude that Tik Tok data - or at least our particular dataset - or techniques applied (short documents, great variance in contexts, topics, a lot of information borne by videos and not text, since this is the essence of TikTok to communicate via videos) is not fully adapted for textual analysis since most of the content shared is random and the description or hashtags does not really carry a common sense behind.

We decided eventually to call out our own sociological expertise and background to elaborate a classification of the topics and see if day 5 would bring more light when crossing all data. We felt the lack of a true expert of TikTok to deepen our reflections.

In total, we obtained 12 videos categories. To name them, we use the "Most popular TikTok categories" presented above and also concepts drawn from theoretical readings (Shifman, 2011, 2012; Zulli & Zulli, 2020).

The 12 topics are: "Social Trend", "Feeling", "Fashion", "Daily Life/Hacks", "Activism", "Animals", "Social Challenges", Outdoors", "Person", "Food", "Arts" and "Sports".
The topic "Social Trend" is, for example, composed of viral hashtags like "#trending, #foryoupage, #foryou, #fyp". "Activism" contains words like "lgbtq, transgender, lesbian, pride". The "Art" topic is made up of "photography, photographer, artist, actor". The lack of expertise mentioned above was partly balanced by the fact that we came up with similar results as those found in section *5.3.2. Detecting hashtag publics.*

5.3.8. Visual results and interpretation: TikTok memes changing over time

We used our multidimensional dataset and fused it with the categories obtained (using a python script, see section *5.2.6 Integrating all data related to Mama said in a single context layer*) to create a more comprehensive basis for visualisation and interpretation.

To explore as many relationships in as many dimensions as possible, we opted to create different hypergraphs (a graph whose hyperedges connect two or more vertices).

We used graphistry (https://www.graphistry.com/), which is a cloud-based interactive graph and hypergraph visualization platform offering a python API client.

We constructed the hypergraph structure in python and sent it to graphistry via the aforementioned API. After creating multiple visual representations and exploring them, we observed some interesting patterns that are discussed below.

A) Networks of hashtags linked by videos posted on Tik Tok - Modularity, degree and dynamics

Hashtags (Nodes): 7 246
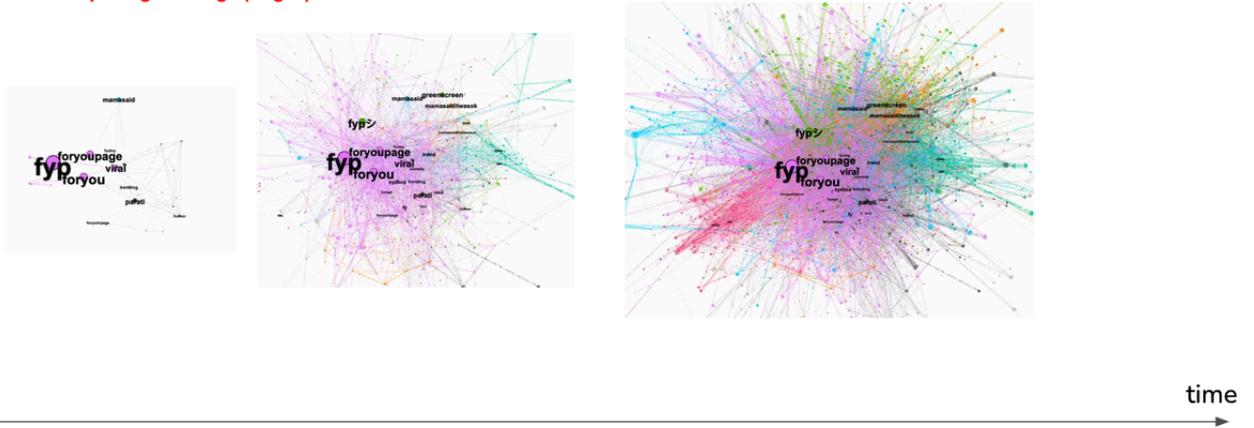Videos quoting hashtags (Edges): 38 691

time

Figure 38. Networks of hashtags linked by videos posted on Tik Tok over time. On the right part appears the peak of virality in October 2021.
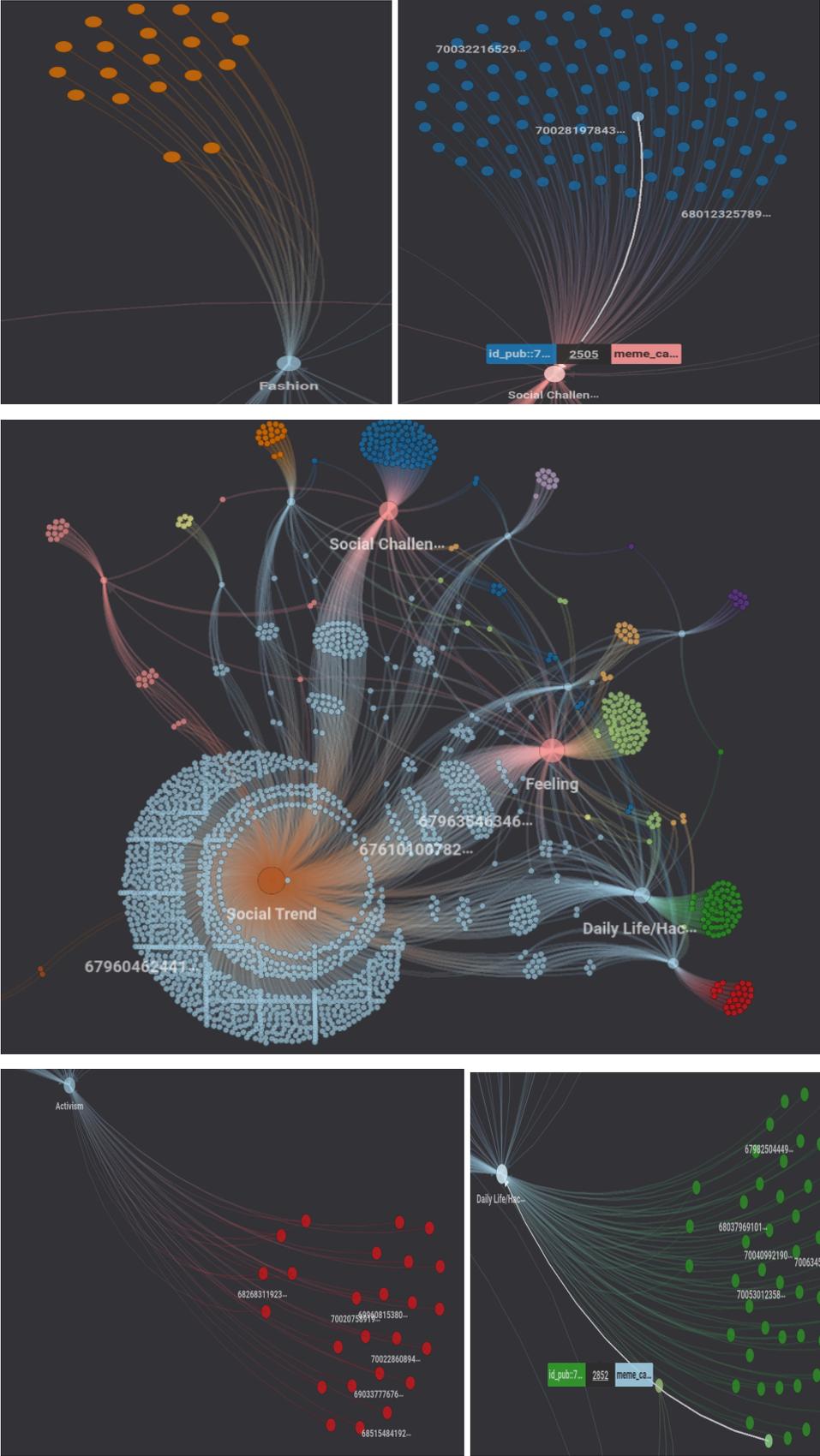
B) Videos by their categories

Figure 38. Hypergraphs linking videos and emerging clusters such as "Social Trend".

As mentioned before, we could clearly see that the "social trend" category is by far the most prominent, and as such, we can in part say that the audio meme is mainly used as a trend surfing tool.

We could also identify other important categories such as "fashion", "social challenges", "activism" and "feeling" related content.

This actually confirmed our categories inferred from literature and partly from NLP.

C) Use of hashtags and users representation
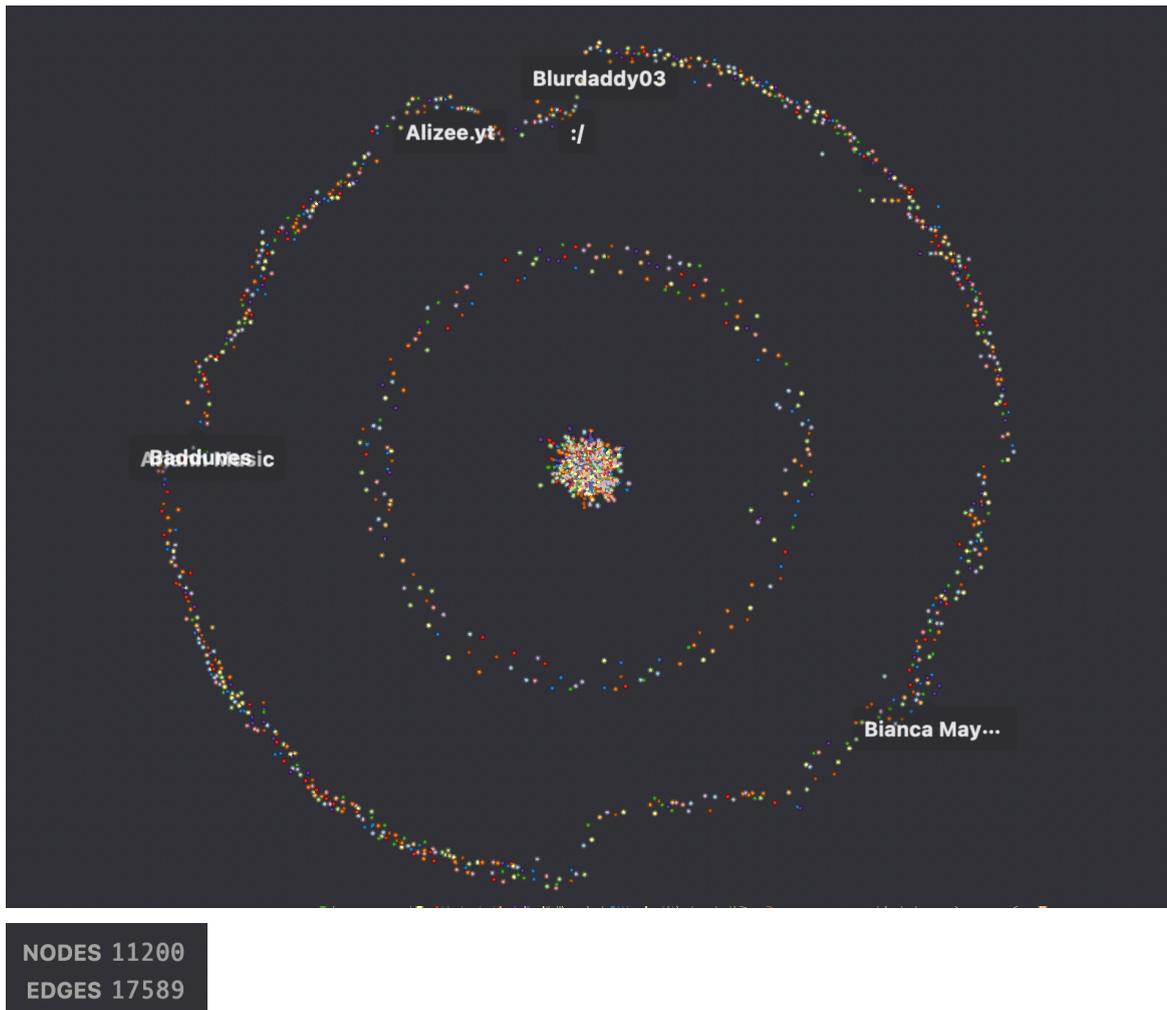
NODES 11200
EDGES 17589
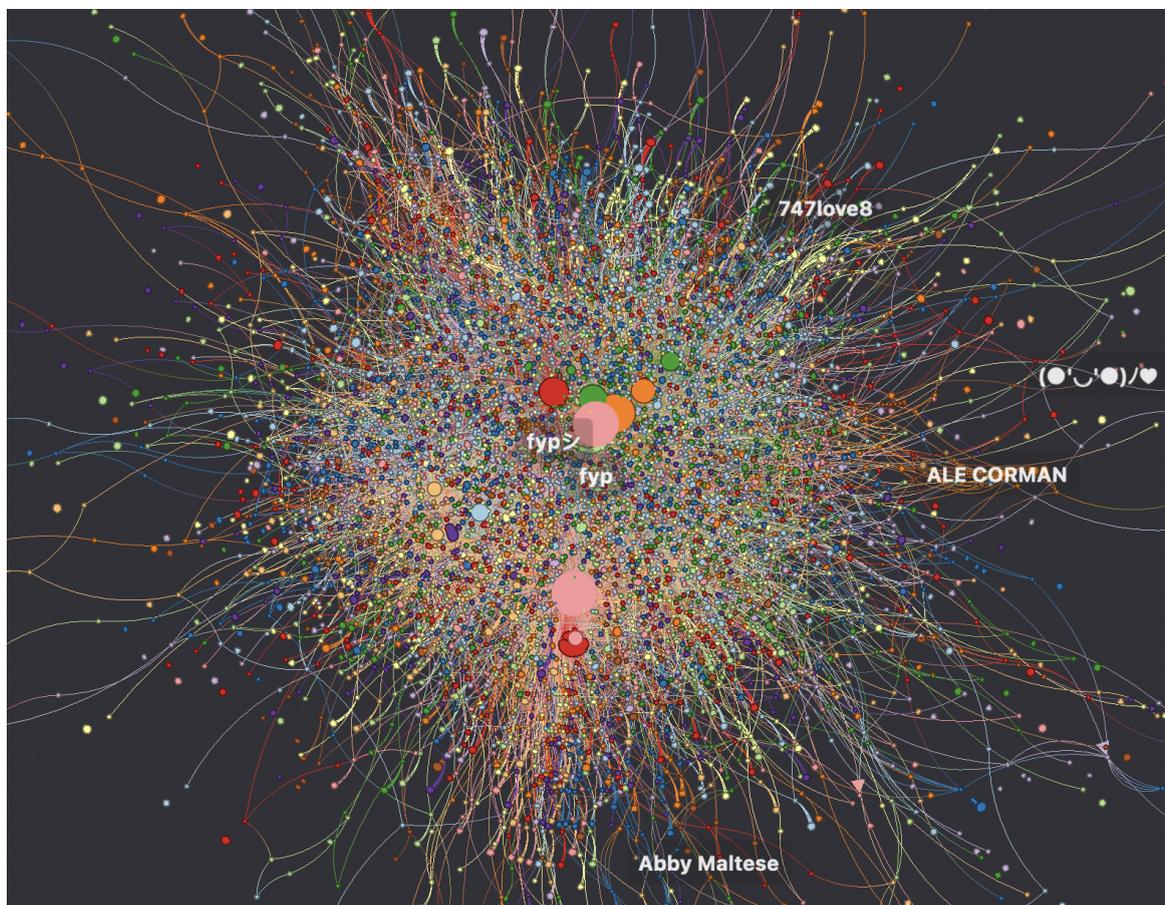
Figure 39. Hypergraphs depicting relations between TikTokers and hashtags used.

This graph represents how users are represented in relation to their hashtag use.

We could derive interesting results:

- The inner cluster, the highest in density, represents Tiktokers using and contributing to the most viral hashtags,
- the first belt-shaped cluster represents Tiktokers who utilise less frequently used hashtags,
- and the external belt-shaped cluster represents Tiktokers who did not use any hashtags in their "mama said" related videos, probably less skilled in Tiktok usage, or at least keeping away from trend surfing and virality.

This graph confirms that the meme use is primarily for trend surfing and allows us to see that most users of this meme (especially the more established ones) rely heavily on hashtags.

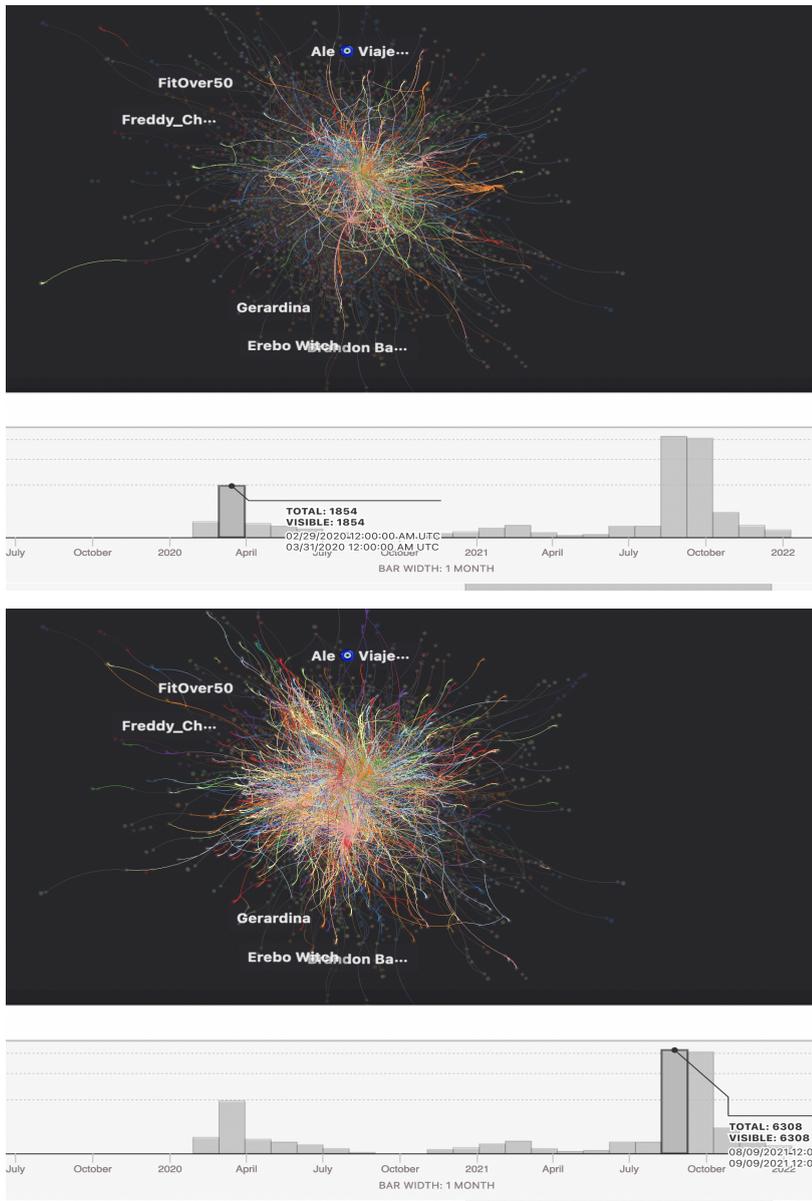D) Temporal exploration of the evolution of the meme

Figure 40. Hypergraphs depicting relations between TikTokers and hashtags used

Here we sliced our hypergraph using time as a dimension in order to explore the evolution of the use of the audio meme over time.

We could clearly identify two peaks concerning the meme's virality (around March 2020 and September/October 2021).

A first fast investigation - checking the peak time frame for possibly related social events that might be driving the peak - found no such event.

We can thus draw a primary conclusion that the virality of the meme is self-driven: a use by Tiktokers with high reach on the platform generates more use of the meme, but further investigation of these peaks would be necessary to establish a more definitive conclusion.

This idea of self-driven virality echoes also in the semantic set of labels derived from networks of co-occurrences and hypergraphs, from which 7 clusters emerge:

- Green = Human body through accessories,
- Purple = sport practices, outskirts
- Orange = cooking and utensils,
- Brown (middle) = Technology,
- Pink (top left) = Animals,
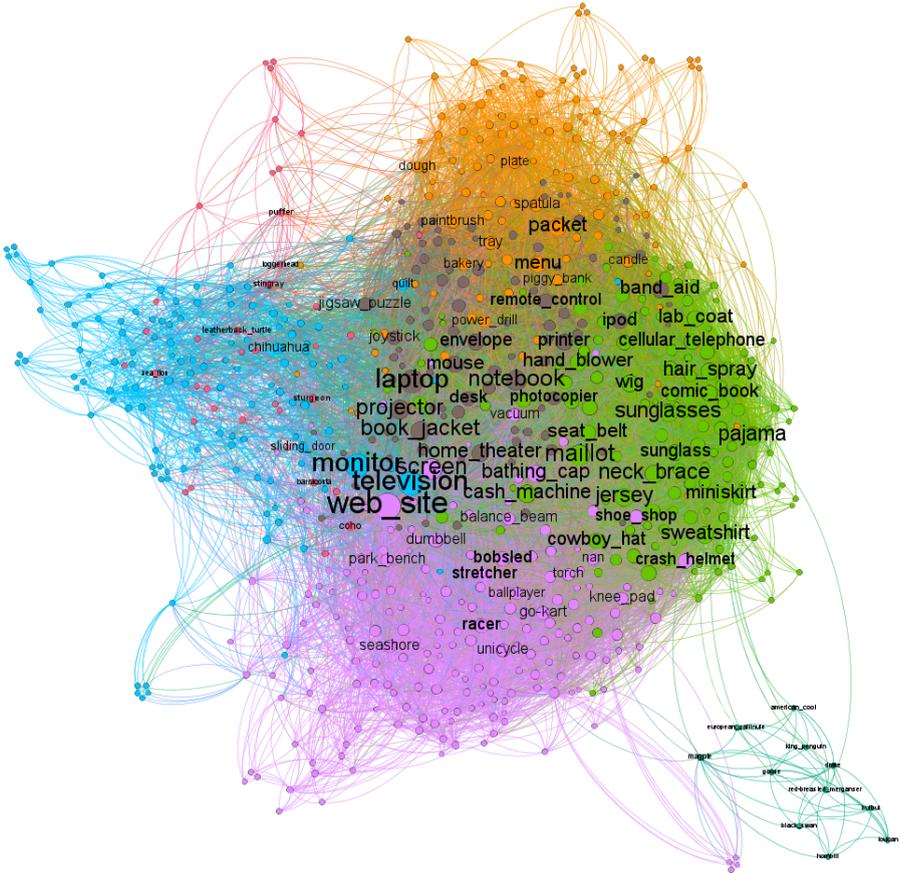- Blue = (left) Dogs and Cats,
- Dark Green (bottom right) = Birds

Figure 41. Labels co-occurrences depicting 7 clusters of semantic labelling extracted from videos
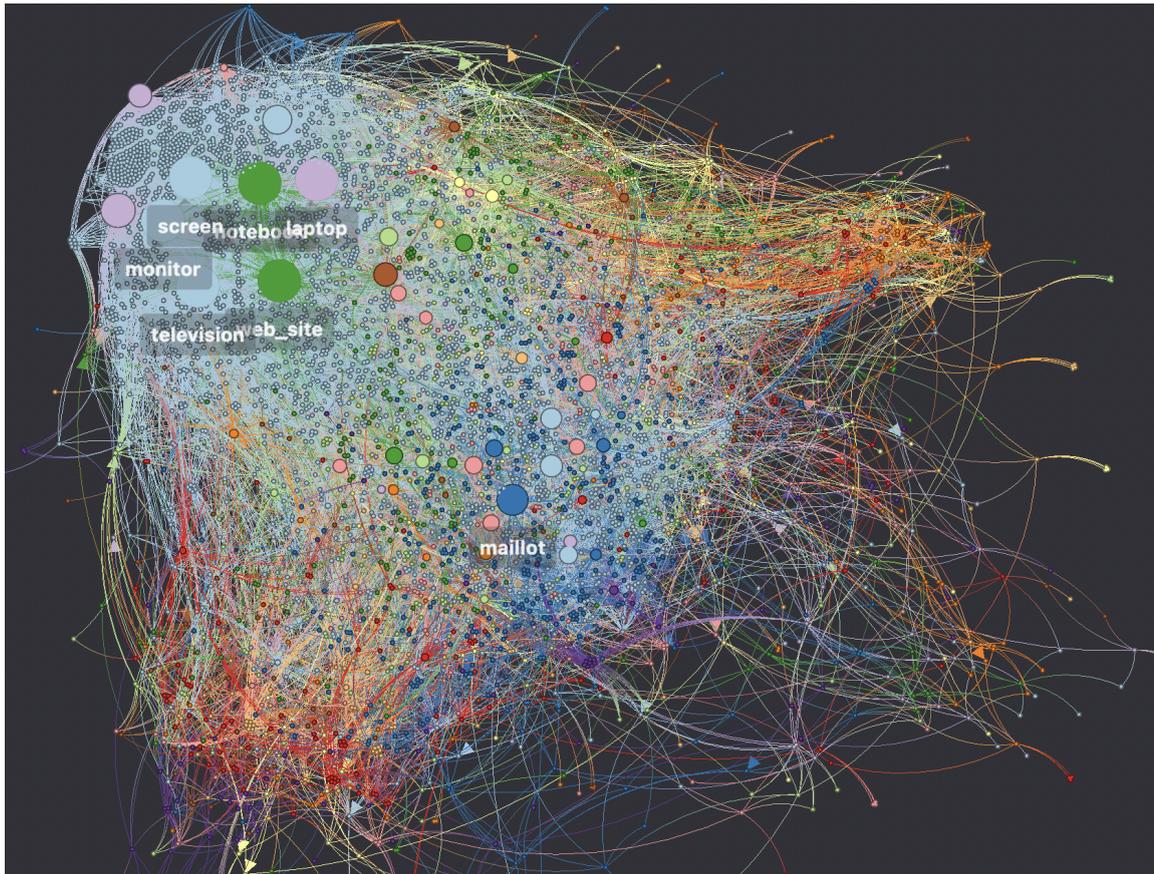
Figure 42. Hypergraphs depicting clusters of semantic labelling extracted from videos.

On a more philosophical touch, the theme of surpassing oneself, somehow authorised by a mythical Mama who "said it was all right", emerges from these categories; human bodies, accessories for staging oneself, or staging one's pets, becoming this hero who goes beyond one's own limitations and sees that performance reflected (be it in sport practices, cooking, arts, inside or in the outskirts) in the responses and gazes of others, calling, in turn, others to pass the marks, is indeed an interesting and attractive stepping stone. The calling Mama of the "Mama said" meme may have something to do with "a screen to project our fantasies, a support to back our self, a reassuring container of affects and emotions, but frightening by the regression and engulfment that would be possible" (Anzieu-Premmereur, 2011). In brief, all characteristics are required to become viral...

For further exploration of the graphs depicted and their various dimensions, we provide some links to all hypergraphs we created during the event.

- "Videoid_tags":
  https://hub.graphistry.com/graph/graph.html?dataset=95e8a232f593446d93702f088d5bc011

- "Meme_cats":
  https://hub.graphistry.com/graph/graph.html?dataset=c49aa1b3805f4a6dad0972964181f82c
- "user custom cats":
  https://hub.graphistry.com/graph/graph.html?dataset=c82c4ecfea2b49448fca9657d16c3826
- "user tags":
  https://hub.graphistry.com/graph/graph.html?dataset=c2a470b2e1474be8b69066070ffb5313
- "video objects memespector":
  https://hub.graphistry.com/graph/graph.html?dataset=5ef93c8097c84c109fe7f589a5652de0

6. Discussion & Conclusion

In this project, we aimed at understanding the genealogy and change of TikTok audio memes by analysing two case studies.

First, our purpose was to understand how TikTok memes are generated on TikTok, and thus study their genealogy. In the analysis, we identified three types of TikTok sounds, which reflects three ways in which TikTok audio memes can be generated: integrated TikTok sounds (e.g. the Stressed Out case study), semi-native TikTok sounds (e.g. the Mama said case study), and native TikTok sounds. The three different sounds are used in different memetic ways and have different trajectories of development, as emerged particularly from the comparison between the original and remix version of the Stressed Out sound, with the remix one being the most memetic.

Furthermore, the analysis provided insights not only into the genealogy of TikTok memes, but also pointed at the idea of the 'dilution of a meme', i.e. when the trend reaches its peak in popularity, it starts getting diluted with non-related content that does not follow any specific template, and therefore the trend itself progressively ceases to exist. By means of hashtag analysis and video stacks, we were able to see, in particular, that once the sound hits peak popularity, its top videos no longer represent

the trend and the sound becomes characterised by 'trend surfing' practices. With trend surfing, we mean the use of the sound, its trend, and popularity the aim of gaining visibility and creating trending content. Ultimately, the project provided useful insights into the functioning of TikTok sounds more broadly, being sound an affordance that enables different kinds of behaviours on the platform, and not only memetic ones.

In the analysis, we were also able to point out some of the ways in which the sounds object of study were used and appropriated by different users. The different usages of the meme clearly emerge from the templatability analysis, which shows how the trend is appropriate to vehiculate feelings of nostalgia (in line with the original meaning of the song), but also re-appropriated for activist purposes or as background music. These various appropriations are to be found in the hashtag analysis as well, where different hashtag publics emerge around both the case studies taken into consideration. However, given the specificities of the platform, more research is needed in order to understand to what extent these publics can be considered as cohesive communities, and what is the subcultural background they share.

The third research question that guided our project is about the ways in which it is possible to analyse TikTok audio-visual content. In this project, we experimented with a series of different strategies to account for the different layers of TikTok content (e.g, metadata, static visual content, dynamic visual content, audio content, among others) and, subsequently, combine them. In the text, we already acknowledge the limitations related to some techniques we tested, specifically the shortcomings of visual similarity analysis, and the complexities associated with the use of textual analysis algorithms applied to TikTok data. Other techniques, such as sound analysis, templatability analysis and video stacks, were particularly useful to extend visual media analysis to TikTok content. Therefore, it remains important to further expand TikTok research beyond the analysis of static images, in order to cover the totality of the video and its multimodality.

In terms of directions for future research, it would be interesting and a great contribution to the social network theory to analyze the relations between the users of the song (following/followers network). We could use Duncan Watts analysis of structured networks to understand what type of connection between people makes a video go viral or not. A limitation in this is however that very little is known about the TikTokers in the data. Moreover, investigating the influence of high-reach users specifically on the adoption and the evolution of TikTok memes is also an interesting perspective. Lastly, we deem it interesting to integrate visual analysis techniques with

the extraction of 'sticker text', either with OCR or by using metadata, if available, in order to further account for TikToks' content embedded in the video paratext.

7. References

Abidin, C., & Kaye, D. B. V. (2021). Audio memes, earworms, and templatability: The 'aural turn' of memes on TikTok. In C. Arkenbout, J. Wilson & D. de Zeeuw (Eds.), Critical meme reader: Global mutations of the viral image (pp. 58–68). Amsterdam: Institute of Network Cultures. ISBN: 9789492302779.

Anzieu-Premmereur, Christine. "Resumen", Revue française de psychanalyse. 2011, vol.75 n° 5. p. 1449-1488. http://www.cairn.info/revue-francaise-de-psychanalyse-2011-5-page-1449.htm .

Beauvisage, Thomas, Jean-Samuel Beuscart, Thomas Couronné, et Kevin Mellet. 2011. « Le succès sur Internet repose-t-il sur la contagion ? Une analyse des recherches sur la viralité ». *Tracés* (21):151-66. doi: 10.4000/traces.5194.

Cano, P., Batlle, E., Kalker, T., & Haitsma, J. (2005). A review of audio fingerprinting. *Journal of VLSI signal processing systems for signal, image and video technology*, *41*(3), 271-284.

Drevo, W. (2021, December 05). Audio Fingerprinting with Python and Numpy. Retrieved from https://willdrevo.com/fingerprinting-and-audio-recognition-with-python

Guven, B. (2021). Extracting Audio from Video using Python - Towards Data Science. Medium. Retrieved from https://towardsdatascience.com/extracting-audio-from-video-using-python-58856a940fd

Leaver, T., Highfield, T., & Abidin, C. (2020). Instagram: Visual social media cultures. Polity.

Mellet, Kevin. 2009. « Aux sources du marketing viral ». *Réseaux* n° 157-158(5):267. doi: 10.3917/res.157.0267.

Rogers, R. (2021). Visual media analysis for Instagram and other online platforms. Big Data & Society, 8(1), 1–23. https://doi.org/10.1177/20539517211022370.

Schellewald, A. (2021). Communicative forms on TikTok: Perspectives from digital ethnography. International Journal of Communication, 15, 1437–1457.

Shifman, L. (2012) An anatomy of a YouTube meme. New Media & Society, 14(2): 187–203.

Wasik, Bill. 2010. *And Then There's This: How Stories Live and Die in Viral Culture*. New York: Penguin.

Watts, Duncan J., et Peter Sheridan Dodds. 2007. « Influentials, Networks, and Public Opinion Formation ». *Journal of Consumer Research* 34(4):441-58. doi: 10.1086/518527.

Zulli, D. & Zulli, D.J. 2020. Extending the Internet meme: Conceptualizing technological mimesis and imitation publics on the TikTok platform. New Media & Society, https://doi.org/10.1177/1461444820983603.