Cyprus University of Technology

Faculty of Communication and Media Studies

Bachelor’s Thesis

Algorithmic culture and filter bubble: The case of YouTube’s recommendation system

George Irakleous

Limassol, May 2020

Copyrights

The approval of the thesis by the Department of Communication and Internet Studies does not imply necessarily the approval by the Department of the views of the writer.

Acknowledgements

I could not have done this thesis without the help of many people. First of all, I would like to express my gratitude toward Dr. Dimitra L. Milioni for her guidance and support toward the whole process of the thesis. Also, I would like to thank Dr. Costas Djouvas for assisting me in both the technological and theoretical parts in order to fully complete the research. Finally, I would like to thank all my friends and family for the patience they showed every time I was talking about the thesis and for their advice and support.

ABSTRACT

Day by day, algorithms of all kinds are becoming part of our daily routines and help us to improve our daily lives. The recommendation systems of many platforms we daily use, are using different algorithms in order to function properly. A strand of the relevant research is currently exploring the impact of algorithms on the identity of users and culture more generally, which leads to the notion of "algorithmic culture". Pariser (2011) first wrote about the phenomenon of the "filter bubble" and how they are being created, both by users and algorithms. Yet, it still remains controversial today if the phenomenon is real, and has split the academic community, as many tried to prove that such a phenomenon is not created by the algorithms. The YouTube platform has one of the most widely used recommendation systems and will be the focus of analysis for this thesis. The thesis examines the platform of YouTube for the existence of a commercial filter bubble in the case of music culture and ponders its impact on identity, or in this case, the music taste of users. Following the method of algorithm auditing, two fake accounts were created and loaded with two different types of music content in order to impersonate two different types of user with different music taste. By analyzing the recommended videos of the two accounts we show how different kinds of bubbles emerged through the recommendation of the platform, and how the platform of YouTube can faction as technology of the self.

Keywords: YouTube, Recommendations, Algorithms, Culture, Filter Bubble

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Introduction

It was the 13th to 14th of August 2019, when a group of LGBTQs YouTubers joined a class action lawsuit suing YouTube for discrimination, deceptive business practices and unlawful restraint of speech. As reported by *The Guardian*, producers and independent film makers joined the lawsuit, suing YouTube for restricting their content and effectively trying to push them off the platform (Kleeman, 2019, Nov 22). In the same article, *The Guardian* attempting to explain YouTube's alleged practices, writes: "Some think it is to appease advertisers wary of being associated with anything on YouTube that could be viewed as controversial". Taking this into consideration, can we really rely on algorithms and recommendation systems as a tool to help us find information, or should we have a more suspicious and critical attitude toward them?

Nowadays, we encounter the increase impact of datafication on our lives, which refers to "the quantification of aspects of life previously experienced in qualitative, non-numeric form, such as communication, relationships, health and fitness, transport and mobility, democratic participation, leisure and consumption" (Kennedy, 2018: 18) or "the transformation of part, if not most, of our lives into computable data" (Cheney-Lippold, 2017: 276-278 [e-book]).

This data – which are often "big" data – are collected and used by commercial platforms such as Facebook, Twitter, YouTube, Skype, free e-mail services etc. To use this data, platforms routinely employ search, filtering and sorting algorithms, leading, as a result, to the algorithmic mediation of users' everyday life, in aspects such as friendships, interests, information searches, expressions of tastes and many more.

All the above is the result of a pipeline of tools used for collecting and analyzing user-generated data. The sole purpose of these tools is the creation of models representing users' behaviors, that will then be used for feeding users with "interesting" content. This thesis focuses on the core element of this pipeline, the technology known as "Recommender Systems". Recommender systems, using various information collected from various sources, and relating this information with a user's behavior, suggests to a user content closer to his/her interests, called personalized content (Ricci et al., 2011). Yet, by doing so, there is a high risk of reproducing and suggesting to a user very similar content creating “echo chambers”. “Echo chambers”, as employed by (Dutton et al., 2017), suggest that users are exposed to the same kind of information with which they have already been exposed to and are familiar with. Pariser (2011), based on the idea of the echo chamber, wrote about the creation of a “filter bubble”, a phenomenon that is encouraged by recommendation algorithms and in some cases by users, so that they come across information that is similar to their own interests.

Recommendation algorithms map our preferences against other users’ preferences, at times suggesting new or forgotten bits of culture for us to encounter. Music, as an aspect of users’ culture, will play an important role in this study. This thesis will explore the existence of a filter bubble through the recommendation algorithms on the platform of YouTube in the case of mainstream and alternative music content. As argued by Cayari (2011), the platform of YouTube has become a powerful space that affords new ways to consume, create and share music, while Allen et al. (2017) consider YouTube one of the three most popular means of discovering music, along with recommendations from friends and family, and the mainstream AM and FM radios.

Taking into consideration the impact of algorithms on a daily basis, and the polarization that may occur by the filter bubble phenomenon, in different aspects of users’ lives generally and in shaping users’ music taste specifically, the research question that this thesis will attempt to answer is: In the case of music, does the recommendation system of the YouTube platform provide more mainstream content than alternative, creating as a result a commercial filter bubble?

YouTube is a platform that is famous for its user-generated videos. On the other hand, many commercial music companies are using the platform to share their music videos and songs with users all over the world. Separating music content into two different kinds will help us examine if the recommendation algorithms of the platform provide more content from those companies (mainstream), than from other, not so famous companies or users (alternative). Our hypothesis is that the algorithms, in order to increase the monetization of content and the platform’s profit, will provide more mainstream content and as a result the existence of a commercial filter bubble will be revealed.

In the next chapters we will discuss why it was considered necessary to study this problem, the theoretical background that the research focused on, a literature review of similar studies, the methodology that was used to collect and analyse the data of the research, the results from the analysis of the data and how those results are linked to the existing literature and previous research work. At the end, further research ideas are suggested.

1.1 Problem – Necessity of study

It is estimated that in each day that passes, 1 billion hours of videos are watched by users, on the platform of YouTube (https://www.youtube.com/about/press/). As one of the biggest video platforms, if not the biggest, YouTube has often been the object of research with contradictory results, especially on the investigation about the existence of a filter bubble. Moreover, YouTube as a video sharing platform, with different themes and genres of videos, has been examined mostly for the existence of ideological filter bubbles, such as the extreme-right and the extreme-left filter bubble.

Moreover, as internet users, we encounter everyday different kinds of algorithms. As we encounter social products with the mediation of algorithms, the result is the danger to be driven into more mainstream paths, limiting originality and creativity. Those algorithms are referred as cultural algorithms or algorithmic culture.

On the one hand, there are several studies than confirm the existence of filter bubbles both on YouTube and other platforms using recommendation algorithms. On the other hand, there are a few studies proving that something like a filter bubble does not exist. This fact increases the importance for further study of the platforms' recommendation system and so does the relative gap in the literature regarding the impact of the recommendation algorithms of YouTube and the creation of filter bubbles in the case of using the platform to find music. Furthermore, a possible filter bubble that is created by the platform's recommendation algorithms, as Allen et al. (2017) have argued, will reduce users' exposure to music that they are not familiar with. As a result of this practice there will be further polarization on users' tastes which can be considered a negative aspect, as it limits the possibilities for users to come across information, knowledge, ideologies and cultural content, different from what they are used to.

1.2 Theoretical background

In this chapter we will focus on and discuss the main theoretical concept which is at the heart of this thesis, which is the “filter bubble” concept, as well as other important and broader theories which inform this study and are connected to the concept under study.

1.2.1 Filter bubble

To begin with, the idea of filter bubbles was coined by Pariser (2011) in his book titled “*The filter bubble: What the Internet is hiding from you*”. As an idea is based on the theory of the echo chamber which, according to Dutton et al. (2017), might restrict access to a more diverse array of views and other political information than one might otherwise come in contact with, either online or offline. In the case of the online environment we have an algorithmic filtering, while on the case of the offline social world, we have a social filtering. When social filtering is transferred to the online activities is referred to as “confirmation bias”, which attracts people to information that confirms rather than challenges their preexisting views. Pariser (2011) referred to the filter bubble as the preferential attention to viewpoints similar to those already held by an individual, which is explicitly encouraged by social media companies. Namely, to increase metrics like engagement and advertise revenue, recommendation systems tend to connect users with information already similar to their current beliefs. Pariser (2011) identifies the emergence of filter bubbles in two possible ways. First, the filter bubble can appear through the suggestions of recommendation algorithms, and second, through users’ activities on the web and the choices of content s/he makes. Or, in most cases, as a combination of both, because if a user already has a diverse information or cultural “diet”, it is difficult for a filter bubble to emerge.

1.2.2 Algorithmic culture

The concerns over the filter bubble(s) do not exist in a vacuum. They are situated in a broader critical discussion of what is called "algorithmic governance", a term used to refer to "a more evidence-based and data-driven than traditional governance" (Just & Latzer, 2017: 245). A key tool in this process is automated algorithmic selection, which "governs a wide spectrum of individual action, and is heavily used for various societal functions", as algorithms co-govern or co-determine what can be found online, what is seen and found, is produced, is considered relevant, is anticipated and is consumed (ibid: 247). This transformed form of governance is based on proprietary big data that tends to strengthen selection criteria oriented on special interests concerned with profit maximization, thus weakening public interest goals and social responsibility in the construction of reality and eventually consolidating and creating new social inequalities (Just & Latzer, 2017). The filter bubble can appear as an effect of or can be used as a tool by the algorithmic governance in order to maximize profits, for example through advertisements, and to construct a reality.

1.2.3 Recommender systems

Recommender systems are a significant part of algorithmic governance, as they use (big) data generated by users in order to create a set of suggestions for them (Ricci et al., 2011). The most widely used form of recommender systems is known as Collaborating Filtering (Breese et al., 1998). Collaborating filtering, based on the assumption that users with similar "tastes" (interests, behaviors, etc.) in the past will have similar "tastes" in the future, clusters users according to their "tastes". Then, utilizing seen behaviors (i.e., clicking on a YouTube video) from members of a cluster, suggests similar behaviors to other users of the cluster.

The platform of YouTube heavily provides recommendation services and, as O'Callaghan et al. (2015) write, these recommendations are sets of videos that are presented to users, based on factors such as co-visitation and viewing history. Co-visitation is a factor based on machine learning algorithms, a type of algorithms that suggest to users videos that are based on what other users with similar characteristics tend to watch. The recommendation algorithms are an important factor in the creation of the filter bubble, both in the case of co-visitation and of a user's viewing history. Also, as argued by Karakayali et al. (2018), recommendation algorithms can function as "technologies of control" or as "technologies of the self".

1.2.4 Technologies of the self

Foucault, as discussed by Karakayali et al. (2018), contended that in modernity there is shift from how subjects are produced in knowledge-power networks to how human beings turn themselves into subjects. To confirm this idea, Karakayali et al. describe a cycle that evaluates the choices of the user, as can be seen in Figure 1. First, algorithms take into consideration the online and offline activities of a user and create a set of recommendations. Then, by the user's choices on these recommendations the algorithms take a recursive feedback of the data and create a new set of recommendations resulting in the "objectified" aspect of the user. But taking into consideration that users can change or modify their activities any time, the algorithms must begin from the point that calculates the users' online and offline activities; and so the circle begins again. Therefore, we need to consider how "technologies of the self" can impact users' behaviour and affect their point of view in subjects that are inseparable with cultural decisions.

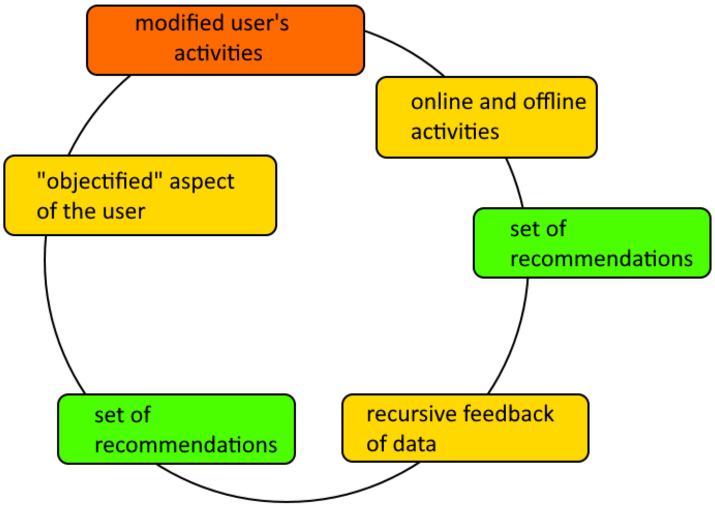

Figure 1: Depiction of the valuation cycle of users' choices (based on the discussion by Karakayali et al., 2018)

1.2.5 Algorithmic culture

Algorithmic governance in general and recommendation systems in particular also affect the realm of cultural production and consumption. Kroeber & Clyde Kluckhohn (1963, cited in Hallinan & Striphas, 2016: 3) wrote that the word "culture" is really hard to define and that there are more than 164 different definitions of it. Hallinan & Striphas (2016: 3), in their attempt to define it, refer to culture as "particular modes of fostering human refinement and their underlying frameworks of valuation and authority, patterns of social difference, commonality and interactions". They also introduce the idea of "algorithmic culture", as "the use of computational processes to sort, classify and hierarchize people, places, objects and ideas and also the habits of thought, conduct and expression that arise in relationship to those processes" (ibid). Based on the case of Netflix, they discuss how the "production of sophisticated recommendations produces greater customer satisfaction which produces more customer data which in turn produce more sophisticated recommendations, and so on, resulting – theoretically – in a closed commercial loop in which culture conforms to, more than it confronts, its users" (ibid: 122). Hallinan & Striphas raise the question "what happens when […] algorithms become important arbiters of culture". This issue is taken up by this thesis, which is empirically exploring this assumption in the case of music recommendations in YouTube.

1.2.6 The Frankfurt School

It is true that we live in a capitalis system, where over-consumption does not refer only to money and products, but the free market has a major impact on the cultural consumption as well. In the Frankfurt School, culture is represented "as the sum total of activities that possess the aura of intellectuality or spirituality, that is, the arts and sciences" (Piccone, 1978). As part of the arts and culture, music was a subject of critical scrutiny under the critical theory of the Frankfurt School. Writing about music, the question that the philosophers of the Frankfurt School posed was the relation between music and the public. "This question cannot be explored independently of the matter of the function and influence of music society" (Τσιτσέλα, 2015, author's translation). One philosopher that devoted a significant amount of time in his work to find an answer was Theodor W. Adorno. Some of his books that show the connection between consuming music and the impact on culture are “Philosophy of new music” (1949) and the “Essays on music” (2002), which is a book comprised by 27 essays written by Adorno. Taking into consideration that the philosophers of the Frankfurt School were considered Marxist intellectuals, we could make a connection between music consumption and the role of algorithms. Adorno always shared the view that society takes anything that tries to criticize its power and makes it part of the mainstream culture (in Τσιτσέλα, 2015), a process which is called “appropriation”. The question, then becomes, do recommendation algorithms play any part in this process?

1.3 Literature Review

The filter bubble phenomenon has become the subject of study in several research works. On the one hand, there are studies which reinforce the argument for the existence of the bubble in search engines and news recommendation (Flaxman et al. 2016; Dylko et al. 2017; Beam, 2014), while on the other hand there are several studies that refute the filter bubble effects, in cases like the search engines, music consumption and news consumption (Nikolov et al. 2015; Hosanagar et al. 2013; Dutton et al. 2017). In this part we will go deeper in three studies, two of which examine the existence of a filter bubble in the case of music on two different platforms with music recommendation algorithms, and one that examines the existence of ideological bubble in the platform of YouTube.

O'Callaghan et al. (2013), in their article “The extreme right filter bubble”, examined the recommendation system on the platform of YouTube asking whether it creates an extreme-right filter bubble. To do so, they first created a set of seed channels for extreme-right content from links propagated by extreme-right Twitter accounts and related channels, which were determined by using the results returned by YouTube’s related video services. To devise their method, they first generated an aggregated ranking of related channels for each seed channel. Next, they generated TF-IDF channel document vectors and then categorized the identified topics according to the set created. Lastly, they categorized the channels based on their topic weight and investigated whether an extreme-right filter bubble exists. Their main finding was that users who access an extreme-right video are highly likely to be recommended further extreme-right content, proving consequently the existence of an ideological filter bubble.

Allen et al. (2017), in their paper titled “The Effects of Music Recommendation Engines on the Filter Bubble Phenomenon”, set out to examine the contribution of the music recommendation algorithm of “Pandora” to the filter bubble phenomenon, in comparison to more mainstream radio services. For their research they recruited 18 participants. All participants were given one temporary email address and by the use of a google chrome extension, the investigators tracked the participants’ actions through the website of “Pandora” (to collect information about the music recommendation algorithm) and “last.fm” (to collect information about the mainstream radio). In the latter case, participants had to listen to 10 to 12 hours of music from a pre-determined set of radio stations. After this process, each participant was interviewed on his/her experience with both services. The result of their study was that, through the comparison of “Pandora” and AM/FM radio listening habits, a cultural filter bubble effect does occur with “Pandora” as it was less likely for the participants to challenge their music preferences on the service. In this case the filter bubble was induced both from users and the algorithms.

In their work titled “Recommendation Systems as Technologies of the Self: Algorithmic Control and the Formation of Music Taste”, Karakayali et al. (2018) aimed to show how music taste emerges as a significant aspect of the user’s self and how it becomes an object of care through his/her interactions with last.fm. The main object of their work was the role played by a recommender system in the care of music taste as an aspect of the self. To do that they delved into the experiences of the users of the music recommendation website “last.fm”. Their data sources were the comments of the users in several forums and ten in-depth interviews with users of the website. Their finding was that “last.fm” does not orient its users toward a definite “music taste” and its effects can be described more as disorientation for the users. From the results of this study, we conclude that, in the case of the music recommendation system “last.fm”, a filter bubble did not occur.

Judging from the findings of previous studies similar to this thesis, it is confirmed there is contradictory evidence regarding the existence (or not) of the phenomenon of filter bubble in the cultural/music field. The first study discussed shows that the recommendation system of YouTube indeed creates an ideological filter bubble in the case of extreme-right content. The other two studies examine two different music recommendation algorithms with diverse results.

2 Research Methodology

The method of this thesis is a relatively recent method called “algorithm audit” (Sandvig et al., 2014), which is adapted from the classic audit study, a field experiment used in social sciences usually to detect discrimination (ibid). More specifically, we will employ the subcategory of “sock puppet audit” (ibid), which entails the impersonation of users by creating false user accounts. Thus, the proposed study is based on an experimental design, and, following the processes of the algorithm audits, it used a mixed method with both qualitative and quantitative analysis. Qualitative analysis was used to categorize the music content, as described below, whereas the quantitative component refers to the statistical analysis of the collected data.

The first step of the research within the current thesis was the creation of the accounts that would be loaded with content. Since the YouTube platform belongs to the Google company, in order to login to the platform two different Google accounts were created. The first account was loaded with mainstream content in order to impersonate a user with mainstream or commercial music tastes and listening habits. The second account was loaded with alternative content following a similar rationale. The procedure to load the accounts with content entailed “watching” video content on YouTube for a period of one month (to ensure that a user profile was consolidated) and each account was loaded with one and half hour of content per day. The entire rationale of this approach was based on the assumption that YouTube personalizes its suggestions to users, according to their observed habits. The procedure began at the end of January 2020 and ended at the first days of March 2020.

Secondly, the categorization of the mainstream and alternative music content took place. As far as the mainstream content is concerned, the decision was made to identify mainstream content through YouTube channels, and more specifically the channel “Universal Music Group” (UMG), which is “the world’s leading music company”, according to their website (https://www.universalmusic.com/company/) and one of the three major players in the global music market; in fact, in 2019 it reached 30% of the total market share in the recorded music market and added more revenue than Warner Music and Sony Music combined (Music Industry Blog, 2020, March 5). Additionally, based on research inside the YouTube platform, one more successful music provider was identified, the VEVO channel, which is a multinational video hosting service owned by Universal Music Group (UMG) and Sony Music Entertainment (SME), along with other music companies (Wikipedia, Vevo). Both channels host some of the most mainstream and popular music videos of the YouTube platform. In the case of the alternative music content, Wikipedia was used as a source of categorization. A Wikipedia page (https://en.wikipedia.org/wiki/Alternative_rock, retrieved 05/05/2020) categorizes the alternative music content in two big genres, alternative rock and alternative metal. Each of these genres had a list of sub-genres. Based on this typology, a song was deemed alternative if it belonged to one of the genres or subgenres of those lists. An additional criterion was the popularity of the song or band. Namely, as specific songs or bands of alternative music become a huge success also among mainstream audiences, we worked with a narrow definition of alternative music, excluding songs or bands that are hugely popular within the YouTube platform from loading the account and thus "training" the algorithm. For example, the song "Chop suey" from the alternative metal band "System of a down", was not used to load the alternative user, as the video today (10/05/2020) has almost 1 billion views. Also, a list was created with all the alternative artists that were found during the process, independently from their popularity, in order to be used for the categorization of the content that would be collected.

As Davidson et al. (2010) write, recommendations on YouTube are featured in two primary locations: the "Homepage" and the "Browse" page. Based on the researcher's engagement with the platform, in the Browse page there are recommendations that are related to the platform; for example, a visit to the "Browse" page towards the end of the year (visited in 2019), the first recommended video is the "YouTube Rewind video", a video that the company of YouTube creates at the end of each year summing up the platform's activities. As can be seen in Figure 2, in a random visit conducted by the researcher, the recommendations of the homepage are more relevant and personalized to the user compared to the "Browse" page.

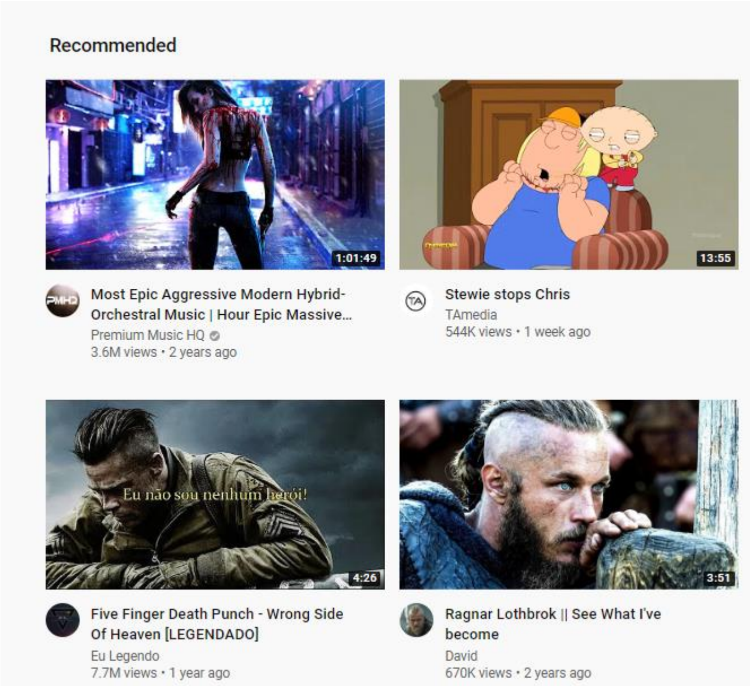

Figure 2: Recommended videos from Homepage

Based on these observations, the decision was made to begin the process of data collection following a path from the "Homepage" to recommended videos on the right side of the screen, as seen in Figure 3. From the researcher's experience and experimentation with the platform, the videos that appeared on the right side of the page are personalized and relevant to the user's preferences and previous viewing history.

Figure 3: Recommendation in YouTube

After the long process of loading the accounts with content, we logged in with each account, and from the "Homepage" we selected the first recommended song-related video, avoiding any other content irrelevant to this study, e.g. sports videos, video games videos, etc. Next, from the landing page, i.e., the YouTube page of the selected music video, the first recommended video was selected, i.e., the first video that appeared on the right side of the page, where a list of recommended video appears.

At this point, with the use of Python language, a program was created to automatically extract useful information from the recommended videos, which was the id of the videos, the title, the channel that uploaded the video, the views, the seconds since the publication of the video, the "Recommended for you" tag (if it existed) and the order in which the video appeared. For the programme to work, we had to save the html code of the YouTube page and then run the programme to save the information mentioned above in a csv file. Papadamou et al. (2019) found that through a 10-click rate, a toddler can come across disturbing video recommendations. Based on the same idea, we followed a 10-click path, which means that we clicked on the first video from the "Homepage", the html of the page was saved and then clicked on the first recommended video. This procedure was repeated 10 times, and for 10 days straight for each account. As a result, a database of 2000 videos, along with their information, was created. The next step was the qualitative analysis of the videos and their categorization into mainstream and alternative (or not song-related videos). Following Davidson et al. (2010: 294), we did not use metadata from the videos for this purpose, such as tags and description of the video, because "video metadata can be non-existent, incomplete, outdated, or simply incorrect". Therefore, the categorization of the videos was completed manually by the researcher, drawing on the typology of the artists on Wikipedia, through the list that was created for the alternative artists and the two YouTube channels. Also when an artist was neither on the channels nor on the list, the researcher manually searched information about the artist's genre. Finally, the analysis of the collected information was completed by statistical analysis and the results will be presented in the next chapter of the thesis.

3 Results/ Findings

3.1 Artists appearance

Through the analysis that follows, the main purpose was to answer the research question of the thesis: In the case of music, does the recommendation system of the YouTube platform provide more mainstream content than alternative, creating as a result a commercial filter bubble? In order to analyse our data, we performed a statistical analysis on the 2000 collected videos using SPSS. In order to investigate the existence of a filter bubble, we first explored the appearance of mainstream artists in the alternative music "fan" account's recommended videos and the appearance of alternative artists in the mainstream music "fan" account's recommended videos. If the recommendations corresponded to the accounts' registered tastes, we could speak about the existence of a filter bubble. Moreover, if mainstream content prevailed in both accounts, we could speak of commercial filter bubble, in the sense of YouTube promoting popular musical content irrespective of the users' preferences. The findings confirm the former assumption. As seen in Table 1, the vast majority of the recommended songs recommended to the mainstream music "fan" were mainstream music songs (90%, n=841); similarly, 96% (n=899) recommended to the alternative music "fan" fall into the alternative music genre. This difference is statistically significant [$x^2(1, n=1869)=1395, p<.001$). Yet, quite unexpectedly, the number of alternative videos that appeared as recommended to the mainstream music fan (10%) was higher than the number of mainstream videos that appeared as recommended to the alternative music fan (4%). The reason we consider the results of this analysis unexpected emerges from the assumption that more mainstream content would be recommended to both fans, in order to propel the profit of the platform, which as a result would be a clear creation of a commercial filter bubble.

		User	
		Mainstream	Alternative
		Count	Count
The artist is categorized as mainstream or alternative	mainstream	841	33
	alternative	96	899

Pearson Chi-Square Tests

		User
The artist is categorized as mainstream or alternative	Chi-square	1395.030
	df	1
	Sig.	.000*

Table 1**:** Recommended mainstream to alternative and alternative to mainstream

3.2 Recommended videos: views and popularity-age

Because of this unexpected result, the analysis continued to examine the views of the recommended videos on both of the above cases. In the case of alternative content (Table 2), the mainstream music fan was recommended content with significantly more views (M = 215137822.9, SD = 369248343), compared to the alternative music fan (M = 5431534.4, SD = 13327210) (*p*<.001). The same goes for the mainstream content (Table 3): again, the mainstream music fan was recommended content with significantly more views (M = 345739592, SD = 577670779), compared to the alternative music fan (M = 2162700, SD = 6453661.5) (*p*<.001).

user	N	Mean	Std. Deviation	Std. Error Mean
mainstream "fan"	96	215137822.92	369248343.047	37686251.201
alternative "fan"	899	5431534.37	13327210.238	444487.347

Table 2**:** Views of alternative content

user	N	Mean	Std. Deviation	Std. Error Mean
mainstream "fan"	842	345739592.11	577670778.887	19907849.727
alternative "fan"	33	2162700.03	6453661.518	1123438.269

Table 3**:** Views of mainstream content

But those numbers are depended and can be affected by the time the video was uploaded on the platform of YouTube. If, for example, two videos had the same number of views but were uploaded three or four years apart, we consider that those two numbers of views are not entirely comparable. For this reason, a new variable was created named "popularity - age". This variable was created by dividing the number of the views with the seconds that had passed since the upload of the video. Then, the analysis was computed again but this time using the new variable of popularity-age of the videos instead of the views. In the case of the alternative music fan (see Table 4), the minimum value of popularity of the mainstream songs was 0.0003638020 and the maximum 0.5921270420 (M = 0.04, SD = 0.10), while in the case of the mainstream music fan (see Table 5) the minimum value of popularity of the alternative songs was 0.0093903710 and the maximum value was 15.69593706 (M = 1.17, SD = 2.73). In the alternative music fan, the mainstream videos that were suggested had a lower value of popularity even compared to the value of popularity of the alternative suggested videos in the mainstream music fan. This can be considered as one step forward to the creation of a filter bubble but not of a commercial one, as the assumption, but a different filter bubble in each fan, according to the popularity-age of it.

N	Valid	33
	Missing	0
Mean		.0435538087
Std. Error of Mean		.0182232008
Median		.0181307060
Mode		.001068721[a]
Std. Deviation		.1046843184
Variance		.011
Range		.5917632400
Minimum		.0003638020
Maximum		.5921270420

Table 5: Popularity-age of mainstream songs in alternative music fan

N	Valid	96
	Missing	0
Mean		1.178638430
Std. Error of Mean		.2788885613
Median		.4234102385
Mode		.7943302160
Std. Deviation		2.732538681
Variance		7.467
Range		15.68654669
Minimum		.0093903710
Maximum		15.69593706

Table 4: Popularity-age of alternative songs in mainstream music fan

3.3 The "Recommended for you" videos

The researcher noticed that some of the recommended videos fell into a distinct category, identified by YouTube by the tag "Recommended for you", as seen in Figure 4. A recommendation system consists of a variety of algorithms. Could that mean that the tag is used to signify information more intensely personalized for a certain user? To answer this question, the first step was to analyze and compare the

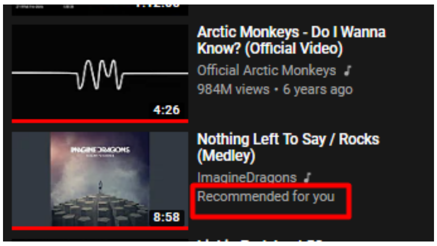

Figure 4: "Recommended for you" tag

frequency in which the tag appeared in the recommended videos of each fan. By running the analysis in song-related data (see Table 6), the results showed that in the alternative music fan 8% of the videos that were recommended had the tag, while the mainstream music fan had 18%. The analysis was significant at the 0.05 level with a significance of $p < .000$. This difference is statistically significant [x^2(1, n=1870)=39941, p<.001).

		User	
		Mainstream	Alternative
		Count	Count
“Recommended for you” tag	Without the tag	769	856
	With the tag	169	76

Pearson Chi-Square Tests

		User
“Recommended for you” tag	Chi-square	39.941
	df	1
	Sig.	.000*

Table 6: “Recommended for you” tag frequency

Yet, the frequency of appearance alone cannot help us understand how those tagged videos are related to the entire set of recommendations for each user. The next step was to compare the popularity-age of those videos to the popularity-age of the whole set of recommendations. In the alternative music fan (see Table 7), the minimum value of popularity-age was 0.0000145660 and the maximum was 1.166508029 (M = 0.10, SD = 0.24). In the mainstream music fan (see Table 8), the minimum value of popularity-age was 0.0020529470 and the maximum was 2186.223529 (M = 21.42, SD = 167.91). Then the two means were compared by T-Test to examine if their difference is statistically significant. As seen in Table 9 this difference, although large, is not statistically important (p = .270).

N	Valid	76
	Missing	0
Mean		.1028482925
Std. Error of Mean		.0275728220
Std. Deviation		.2403742891
Minimum		.0000145660
Maximum		1.166508029

Table 7**:** Popularity-age of “Recommended for you” in alternative music fan

N	Valid	169
	Missing	0
Mean		21.42739483
Std. Error of Mean		12.91650182
Std. Deviation		167.9145236
Minimum		.0020529470
Maximum		2186.223529

Table 8**:** Popularity-age of “Recommended for you” in mainstream music fan

user	N	Mean	Std. Deviation	Std. Error Mean
mainstream “fan”	169	21.42739483	167.9145236	12.91650182
alternative “fan”	76	.1028482925	.2403742891	.0275728220

Table 9: T-Test popularity of the “Recommended for you” songs

But are those values mirrored in the entire set of the recommended videos or are they different? So, the next step was to compare those values of popularity with the entire recommended datasets of each user (including the tagged videos). In the mainstream music fan (Table 10), there was a mean of 6.41 (SD = 71.74). In the alternative music fan (Table 11), there was a difference in all the values of popularity of the videos. The minimum value was 0.0000008266 and the maximum value was 5.719632382 (M =0.07, SD = 0.25).

N	Valid	938
	Missing	0
Mean		6.417845476
Std. Error of Mean		2.342394337
Std. Deviation		71.74000931
Minimum		.0020529470
Maximum		2186.223529

Table 11: Popularity-age of mainstream music fan (whole set)

N	Valid	932
	Missing	0
Mean		.0797000041
Std. Error of Mean		.0084807815
Std. Deviation		.2589070219
Minimum		.0000008266
Maximum		5.719632382

Table 10: Popularity-age of alternative music fan (whole set)

Then, once again, means were compared with the use of T-Test. The results in Table 12 showed that the difference in the means is statistically significant at the 0.01 level (p = .007). From the results, what we conclude is that the recommendation algorithms of YouTube, tends to use the tag in videos with higher values of popularity-age, and that can be seen by the 14.1 difference in the means. In the alternative music fan, the algorithm tends to use the tag in videos that were not in the edge of the popularity-age value but close enough to the mean of this value, the difference is only

0.03, which could also be translated that the algorithm uses the tag in videos with higher value of popularity-age. That is shown by the fact that the 76 videos with the tag have a mean of 0.102 while the 932 videos have a mean of 0.079.

user	N	Mean	Std. Deviation	Std. Error Mean
mainstream “fan”	938	6.417845476	71.74000931	2.342394337
alternative “fan”	932	.0797000041	.25890702219	.0084807815

Table 12: T-Test popularity of all songs

3.4 Viewed vs Recommended videos

In order to examine the filter bubble assumption, it was considered necessary to compare the videos that were recommended with the videos that were used to load both users with content and construct the users’ profile in terms of musical taste. Unfortunately, during the process of loading, the researcher did not manage to record the information about the seconds since publication of the videos, so it was not possible to compute the value of popularity for those songs; this is the reason why we will compare only the views of the videos. The purpose of the analysis was twofold: first, to examine whether more popular content is promoted, through the recommendations, to the mainstream content fan; second, to examine whether the recommended content was more popular compared to the content the two “users” consumed in the first place (namely, the songs used to load the accounts).

For the alternative music fan (Table 13), the minimum number of views of a video (in the dataset used to load the accounts) was 41 and the maximum was 500798 (M = 44284.84, SD = 71371.81). For the mainstream music fan (Table 14), the minimum number of views (again in the dataset used to load the account) was 272211355 and the maximum was 4725409580 (M = 784206748.9, SD = 640457467.5). The recommended videos of the mainstream music fan (Table 15), had a minimum number of views of 138200 and a maximum number of views of 4683051115 (M = 332413196, SD = 561143147.7). The recommended videos of the alternative music fan (Table 16) had a minimum number of views of 1 and a maximum number of views of 122051751 (M = 5315792.38, SD = 13157346.18). As noted above, we cannot provide a more accurate analysis of this data as the seconds since publication of the video that were used for the loading of the accounts was not obtained. Nevertheless, in the mainstream music fan, there was a difference in the means of the videos that were used to load the account and the recommended videos, but in both cases the number of views was more than 100000000. In the alternative music fan the difference in the means was also huge with 44284.84 in the loaded videos and 5315792.38 of the recommended videos.

N	Valid	938
	Missing	0
Mean		332413196.0
Std. Error of Mean		18321973.25
Std. Deviation		561143147.7
Minimum		138200
Maximum		4683051115

Table 14: Views of loaded videos, Alternative music fun

N	Valid	470
	Missing	43
Mean		784206748.9
Std. Deviation		640457467.6
Minimum		272211355
Maximum		4725409580

Table 13: Views of loaded videos, Mainstream music fun

N	Valid	513
	Missing	0
Mean		44284.84
Std. Deviation		71371.816
Minimum		41
Maximum		500798

Table 15: Views of recommended videos, Mainstream music fun

N	Valid	932
	Missing	0
Mean		5315792.38
Std. Error of Mean		430983.204
Std. Deviation		13157346.18
Minimum		1
Maximum		122051751

Table 16: Views of recommended videos, alternative music fun

A T-Test analysis was also conducted, in order to examine if the difference in the means was statistically significant. The first analysis (see Table 17) compared the views of the videos that were used to load the alternative music fan with content and the views from the videos that the user was recommended. Their difference is statistically important at the .00 level with $p < .001$. Regarding the mainstream music fan (Table 18), the difference was once again statistically important in the .00 level with $p < .001$, however the mean number of views was smaller in the recommendations dataset compared to the viewed videos dataset. What we conclude from this analysis is that in the case of the alternative music fan, the algorithm suggested videos with more views on average, compared to the views of the videos that were used to load the user's account. This means that the user was suggested more popular content, in terms of views, albeit still within the alternative music genre. Thus, it did not make it easier for the user to encounter undiscovered or inconspicuous content. In the mainstream music fan, the algorithm suggested videos with fewer views compared to the views of the videos that were used to load the user's account, on average. Yet, the mean number of views was still significantly high (over 300 million views).

user	N	Mean	Std. Deviation	Std. Error Mean
Viewed videos	513	44284.84	71371.816	3151.143
Recommended videos	1000	5146791.91	13212337.92	.0084807815

Table 17: T-Test alternative user: loaded vs recommended videos

user	N	Mean	Std. Deviation	Std. Error Mean
Viewed videos	470	784206748.9	640457467.6	29542100.06
Recommended videos	1000	312938149.6	548767037.4	17353537.43

Table 18: T-Test mainstream user: loaded vs recommended videos

Another analysis that was conducted is related to diversity of the artists or bands, namely the possibility that the recommended videos suggested more songs of the same artists or opened up the user to more diverse content. A new variable, coded for the sameness/uniqueness of the artists or music groups was created. The findings show, quite surprisingly, that regarding the alternative music fan, only 5.6% of the recommended videos were from artists that were used during the process of loading the account with content. This suggests that recommendations steer the user towards discovering new (for her/him) content. On the other hand, regarding the mainstream music fan, 70.5% of the recommended videos were from artists that were used during the process of loading the account with content. Here, this diversity is significantly diminished.

3.5 The non-songs

The results, so far, show a clear profiling of the users, in both accounts. But to go even deeper we conducted a qualitative analysis of the recommended videos that were not music-related, to understand what the algorithms tried to "serve" the users. In the alternative music fan, 68 videos (7%) could not be classified as mainstream or alternative, so they were classified as non-music content. Of those 68 videos, five of them had as main subject the coronavirus, and four were connected to the genre/taste of the account (i.e. a deathcore song about the virus –and deathcore was classified as a subgenre of alternative metal). Those videos were considered in the analysis as non-songs because they were not created or recorded by an official studio or artist, but they seemed to be made for fun. All the other videos were close to the culture of alternative rock and metal, e.g. top 10 lists, interviews with artists, tutorials on how to scream, headbang challenges etc. Fifteen of those videos (22%) were marked with the tag "Recommended for you". In the mainstream music fan, there were 62 videos (6%) that were classified as non-music content. From those videos there were again some lists, like top 10 songs of a specific artist or top 100 songs from the 2000s etc., while other videos included content that was gossip about the life of mainstream artists, like Selena Gomez and Justin Bieber. Moreover, there were videos from Disney's productions, like movies and series, and videos from talent shows like "The Voice". Furthermore, 16 of those 62 videos (26%) had the "Recommended for you" tag.

3.6 Non-songs & "Recommended for you"

On both users, non-music content that had the "Recommended for you" tag, appeared more often than in the other videos of the dataset. So, we ran an analysis to compare the non-music recommended videos with both mainstream and alternative videos (Table 15). The results showed that the song videos had 245 videos (13%) with the tag and 1625 videos (87%) without the tag, while in the non-music videos, 31 of the videos (24%) had the tag and 99 of the videos (76%) had not. This difference is significant at the .05 level [$x^2(1, n=2000)=11796, p<.001$]. The results show that the "Recommended for you" tag appears more often in non-music videos. When we compared the non-music content of the two accounts (Table 16), it was the only result that was not statistically significant [$x^2(1, n=130)=251, p=.616$]. The platform of YouTube does not provide only music videos but a variety of topics and kinds of videos. The question here is: why would the algorithm suggest videos that are not songs to a user that has been watching only songs on the platform? What is more, why are videos not only recommended, by appearing at the recommended list, but also added a tag, which can make the user feel more confident that the content of this video is relevant to his/her preferences? Those questions cannot be answered from the current analysis, but they could be used to trigger further research on the platform of YouTube, with real users, so their behaviour could be examined regarding the recommended videos.

		Song or non-song	
		Song	Non-song
		Count	Count
"Recommended for you" tag	Without the tag	1625	99
	With the tag	245	31

Pearson Chi-Square Tests

		User
"Recommended for you" tag	Chi-square	11.796
	df	1
	Sig.	.001*

Table 19: Songs and non-songs with and without the tag

		User	
		Mainstream	Alternative
		Count	Count
"Recommended for you" tag	Without the tag	46	53
	With the tag	16	15

Pearson Chi-Square Tests

		User
"Recommended for you" tag	Chi-square	.251
	df	1
	Sig.	.616

Table 20: Non-songs with and without the tag

4 Discussion and conclusion

The analysis that preceded focused on answering the research question of the present thesis. But from the analysis of the data that was collected, more aspects of the recommendation system of YouTube were examined, for example, the use of YouTube's recommendation system as a cultural algorithm.

First, as it is evident from the results, a filter bubble was created according to the content that was used to load the accounts. In the mainstream music fan, the majority of the videos that were recommended were classified as mainstream music, whereas in the alternative account most recommended videos were classified as alternative music. Thus, in this case we encounter a manifestation of the filter bubble phenomenon, based on the cultural content recommended by the platform. As a result, a user that starts to use the platform having a specific taste in music and over time seeks to use the platform in order to widen her/his horizons in music, s/he is more likely to encounter content very similar to his/her preexisting taste in music.

Second, in the case where alternative music videos were recommended in the mainstream music fan, the popularity of those videos was close to the popularity of the mainstream videos. This result leads to the conclusion that the views as well as "freshness" of the videos are compatible with the entire set of recommendation, irrespectively of the genre of the music. This means that even when alternative music content makes it into the recommendations of YouTube to a user with mainstream tastes, this content is highly likely to be already popular, in terms of views, as well as "fresh" (recent). This renders unlikely that a user with mainstream music tastes encounters diverse "undiscovered" cultural content, as there seems to be at work a process of making more salient particular songs that are already popular, recent or both, even within the alternative music genres.

The same goes for the alternative music fan: whenever a mainstream music video appeared as recommended, the popularity of the video was similar to the popularity of the alternative songs what were recommended. Compared to the mainstream music fan, the popularity of the mainstream recommended videos in the alternative music fan, was lower. The paradox of the analysis was that even the recommended videos to the alternative music fan, that were categorized as mainstream, had a popularity similar to the alternative recommended videos. In this case, a filter bubble, this time according to the popularity of the videos, was created too.

The big difference between the two accounts lies in the videos that were not music-related content. In the entire dataset, these videos appeared more often tagged as "Recommended for you", compared to the music videos. The analysis of those videos, both qualitative and quantitative, leads us to the examination of a commercial filter bubble. In the alternative account, the non-music videos were really close to the genre of the content that was used to load the account, namely content that belonged to the alternative rock and metal culture. On the other hand, in the mainstream account, the recommended videos were highly commercial content originating from popular culture products such as talent shows, videos about the lives of celebrities or parts from Disney's productions. The main difference here is that in the case of the alternative account the non-music recommended videos can be seen as additional information that reaffirms the user's belonging to this kind of content and culture. On the other hand, in the mainstream account, the recommended videos try to "pull" the user into other kind of activities, outside music consumption, that are connected to mainstream culture, like the productions of Disney.

Because of the diversity of content that is hosted in the platform of YouTube, we examined Pariser's (2011) idea of the filter bubble from different aspects and in all cases it was verified that the recommendation were similar to the viewpoints that were already held. To come to this result, it was necessary to examine and compare the frequencies of the videos that were used to load the accounts with content to the recommended videos of each user. The results showed that in the alternative music fan, only 5.6% of the 1000 videos that were recommended was from the same artists that were used during the loading process, which means that a filter bubble was not created as far as the artists or bands are concerned. On the other hand, in the mainstream music fan, 70.5% of the 1000 recommended videos were from artists that were used during the process of loading. That high percentage of repetition of the same artists leads to the conclusion that much of the content that was used to load the account was repeated, in terms of individual artists or bands. Here, we observe the creation of a filter bubble, not only in terms of the genre, but of the artists too.

In light of the results of this study, we now turn to consider anew the concept of the “algorithmic culture”, introduced by Hallinan & Striphas (2017). We can argue that the alternative music fan seems to have fallen victim of the “algorithmic governance”. From the start, the recommendation algorithm suggested videos with more views and popularity than the songs that were used in the procedure of loading the account. This means that simulating the tastes and preferences of a user who prefers “fringe” bands and artists, in the recommendations we encountered more popular and well-known music videos, albeit still in the alternative music genre. Thus, artists with an already low popularity in the YouTube platform are unlikely to see their videos in the recommended lists of users (even users who actively select this kind of music). Because neither the exact logic of the recommendation algorithm of YouTube nor the actual motives behind it are known, we cannot explain why this is so. A plausible cause points to profit maximization as the videos by artists from more official discography companies, or from artists who already have a social influence and impact, are more likely to spend money in order to promote their videos from YouTube to other social and mass media in their attempt to increase their popularity. It works as a win-win situation for both sides. The platform of YouTube gets more clicks, and the artists/companies gain more fame.

But what if we look at the recommendation algorithm of YouTube as a “technology of the self”, as discussed by Karakayali et al. (2018)? In their work they argued about how the algorithms make a circle in order to recommend more personalized content and engage the user to their content. The circle begins with the algorithm tracking both the offline and online activities of the user. To develop this discussion, we focused completely on the online activities of the users, because there were not any offline activities, as the accounts were fake and were only used for this specific research purpose. Thus, it appears that the algorithm used only one user action, the history from the videos that the user had already viewed. The first set of recommended videos appeared in the “Homepage”, where different videos appeared. The fact that the user chose one of these recommended videos was the recursive feedback of the data, which drove the user to the second set of recommended videos, the ones that appeared in the right side of the webpage. From the choices the user made on the second set of recommendation, the algorithm created an “objectified” aspect of the user but was always ready to modify the user’s selections, which in this case occurred the next day, where the cycle began once again. As a result of this, the recommendation algorithm of YouTube can be considered as a technology of the self. This result can be clearly seen if we analyze the separate days of the recommended videos. For example, the second day in the mainstream account, the first suggested song was from an alternative artist with high popularity. From the 100 recommended videos that we collected, only 17 of them were from mainstream artists. This means that the algorithm may have modified the user's activity and did not disregard previous activities. More important is the impact of the technologies of the self in the creation of a possible identity, or in the case of this study, the creation of music taste of the users in order for the algorithm to engage the users with content that serves the interests of the platform, namely, steering the user toward specific videos to satisfy the goal of maximizing profits within new media capitalism.

The filter bubble phenomenon has triggered the interest of the academic community to examine its existence in different platforms, from different point of views and different concepts. For example, Allen et al. (2017) examined real users and their behaviour, in the platforms of "Pandora" and "last.fm" and the creation of a filter bubble in the case of music. The results were similar to this thesis. A filter bubble emerged to surface in the platforms, but in some cases the bubble was not created by the algorithm but by the users' activities, who were prone to create their own bubble, a musical "safe zone" for them to consume their music taste, over and over again. Similar studies took place in the YouTube platform examining the existence of an ideological filter bubble this time. O'Callaghan et al. (2013, 2015) published about the creation of an ideological filter bubble in the extreme right content and the creation of a pipeline. The results showed that in a case a user consumes ideological right content is more likely to be recommended extreme right content. Cross-checking those results with the results of this thesis, some similarities were observed. For example, the mainstream music fan was recommended videos with gossipy news about artists, or videos from mainstream productions, like Disney. The user begun to consume mainstream music content and then the recommendation systems began to suggest videos that are part of the mainstream culture but not only from the music aspect. By replacing the alternative videos with the LGBTQ+ YouTubers, it could be mean that truly, the recommendation algorithms might underestimate their work in order to suggest more mainstream YouTubers who could increase the profits of the business.

5 Limitations and suggestions for future research

As every research, the thesis had some limitations, as it was not possible to control all factors. First and most important limitation of the research was that fake accounts were used, which tried to mimic the behavior of a real user. So, the results cannot reflect on the behavior of real users of the platforms, who are more likely to have a variety of preferences in music and not restrict their choices to one particular genre, which in turn would affect the construction of their profiles and hence their recommendations. However, it was decided to proceed with the fake accounts and indeed construct them as more or less extreme types of users in order to be able to simplify the number of different variables that play some role in algorithmic filtering. One more limitation was that we excluded a big part of the music culture, like different genres of music, such as reggae or opera; on the other hand, however, such a limitation was necessary due to time and resources. Furthermore, the categorization of the music into two rough categories (mainstream and alternative) leaves out multiple shades of music genres in between. It is important to mention that the loading of the accounts was made manually by the researcher and needed a minimum of three hours per day. Moreover, more data from the videos could have been gathered, especially in the process where the two accounts were loaded with content. It would be important to gather information of those videos that were not collected from the start, such as the channel and the seconds that have passed since the publication of the song when it was first used for the purpose of the research.

From the experience I gained through the entire process, I would suggest that this kind of research takes place in an environment more specialized in music, like “Spotify” or “last.fm”. It would also be interesting to conduct a similar algorithm audit specifically of a service offered by the platform of YouTube, namely the YouTube Music. The examination for different kinds of filter bubbles would also be interesting, such as the creation of a filter bubble in the case of YouTubers. In their “About” page, the creators of YouTube mention that “Our mission is to give everyone a voice and show them the world”. It may be high time that researchers took a more active role in showing how commercial recommendation algorithms, like YouTube’s, work and they do not show us the world as it is, but only from the platform’s profitable point of view.

REFERENCES

Abdollahpouri, H., Burke, R., & Mansoury, M. (2020). Unfair Exposure of Artists in Music Recommendation. arXiv preprint arXiv:2003.11634.

Adorno, T. W. (1991). On the fetish character in music and the regression of listening. The culture industry: Selected essays on mass culture, 26-52.

Andrejevic, M. (2009). Exploiting YouTube: Contradictions of user-generated labor. The YouTube Reader, 413.

Allen, D. P., Wheeler-Mackta, H. J., & Campo, J. R. (2017). The Effects of Music Recommendation Engines on the Filter Bubble Phenomenon. Retrieved from https://digitalcommons.wpi.edu/iqp-all/3086.

Cayari, C. (2011). The YouTube Effect: How YouTube Has Provided New Ways to Consume, Create, and Share Music. International Journal of Education & the Arts, 12(6), n6.

Cheney-Lippold, J. (2017). We Are Data: Algorithms and The Making of Our Digital Selves. New York: New York University Press (Kindle Edition).

Chitra, U., & Musco, C. (2019). Understanding Filter Bubbles and Polarization in Social Networks. arXiv preprint arXiv:1906.08772.

Chodos, A. T. (2019). Solving and Dissolving Musical Affection: A Critical Study of Spotify and Automated Music Recommendation in the 21st Century (Doctoral dissertation, UC San Diego).

Davidson, J., Liebald, B., Liu, J., Nandy, P., Van Vleet, T., Gargi, U., ... & Sampath, D. (2010). The YouTube video recommendation system. In Proceedings of the fourth ACM conference on Recommender systems (pp. 293-296). ACM.

Dutton, W. H., Reisdorf, B., Dubois, E., & Blank, G. (2017). Social Shaping of the Politics of Internet Search and Networking: Moving Beyond Filter Bubbles, Echo Chambers, and Fake News.

Francesco Ricci and Lior Rokach and Bracha Shapira (2011). Introduction to Recommender Systems Handbook, Recommender Systems Handbook, Springer, pp. 1-35

Gillespie, T. (2014). The relevance of algorithms. Media technologies: Essays on communication, materiality, and society, 167, 167.

Hallinan, B., & Striphas, T. (2016). Recommended for you: The Netflix Prize and the production of algorithmic culture. New media & society, 18(1), 117-137.

John S. Breese; David Heckerman & Carl Kadie (1998). Empirical analysis of predictive algorithms for collaborative filtering. In Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence (UAI'98).

Just, N. & Latzer, M. (2017). Governance by algorithms: Reality construction by algorithmic selection on the Internet. Media, Culture & Society, 39(2), 238-258.

Karakayali, N., Kostem, B., & Galip, I. (2018). Recommendation systems as technologies of the self: Algorithmic control and the formation of music taste. Theory, Culture & Society, 35(2), 3-24.

Kennedy, H. (2018). Living with Data: Aligning Data Studies and Data Activism Through a Focus on Everyday Experiences of Datafication. Krisis: Journal for Contemporary Philosophy, 1, 18-30.

Kruse, H. (1993). Subcultural identity in alternative music culture. Popular music, 12(1), 33-41.

O'Callaghan, D., Greene, D., Conway, M., Carthy, J., & Cunningham, P. (2015). Down the (white) rabbit hole: The extreme right and online recommender systems. Social Science Computer Review, 33(4), 459-478.

O'Callaghan, D., Greene, D., Conway, M., Carthy, J., & Cunningham, P. (2013). The extreme right filter bubble. arXiv preprint arXiv:1308.6149.

Pariser, E. (2011). The filter bubble: What the Internet is hiding from you. Penguin UK.

Papadamou, K., Papasavva, A., Zannettou, S., Blackburn, J., Kourtellis, N., Leontiadis, I., ... & Sirivianos, M. (2019). Disturbed YouTube for Kids: Characterizing and Detecting Inappropriate Videos Targeting Young Children. arXiv preprint arXiv:1901.07046.

Piccone, P. (1978). *The essential Frankfurt school reader* (p. 270). A. Arato, & E. Gebhardt (Eds.). New York: Urizen Books.

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). Auditing algorithms: Research methods for detecting discrimination on internet platforms. Data and discrimination: converting critical concerns into productive inquiry, 22.

Van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. Surveillance & Society, 12(2), 197-208.

Τσίτσελα, Έ. (2015). Η μαζική κουλτούρα από τη Σχολή της Φρανκφούρτης στον μεταμοντερνισμό: η περίπτωση της δημοφιλούς μουσικής. Retrieved from: http://dspace.uowm.gr/xmlui/handle/123456789/189

Internet resources

https://musicindustryblog.wordpress.com/tag/record-label-market-shares-2019/
Retrieved: 03/05/2020

https://en.wikipedia.org/wiki/Vevo Retrieved: 03/05/2020

https://en.wikipedia.org/wiki/Alternative_rock Retrieved: 27/01/2020

https://www.youtube.com/about/ Retrieved: 07/05/2020

https://www.theguardian.com/technology/2019/nov/22/youtube-lgbt-content-lawsuit-discrimination-algorithm Retrieved 10/05/2020

https://www.youtube.com/about/press/ Retrieved 10/05/2020

APPENDIX I

Python code

```
from bs4 import BeautifulSoup

import pandas as pd

import datetime

import dateparser


user = 'giorgos' # The name of the user

file = 'Alt Day 10 Path 10' # The file where the search results were stored - without the .html extension


# Open and process the html file

with open(file+'.html', encoding='utf8') as fp:

    soup = BeautifulSoup(fp)


# Extract and process all the results returned

data = []


# Works for: ytd-compact-autoplay-renderer  ytd-compact-video-renderer

# Does not work for: ytd-compact-radio-renderer ytd-compact-playlist-renderer

counter = 1

for i, video in enumerate(soup.select('ytd-compact-autoplay-renderer')):

  link = video.select_one('a.yt-simple-endpoint.style-scope.ytd-compact-video-renderer')

  video_id = link.get('href').split('?')[1].split('&')[0][2:]

  title = link.select_one('h3 span#video-title').string.strip()
```

```
  title_meta = link.select_one('h3 span#video-title').get('aria-label').strip()
  meta = link.select_one('div.ytd-video-meta-block');
  channel = meta.select_one('yt-formatted-string.style-scope.ytd-channel-name').string;
  try:
    second_meta = meta.select_one('div.ytd-video-meta-block span.ytd-video-meta-
block').string.strip()
  except:
    second_meta = ''

  tmp = title_meta[len(title)+4:].strip().split(' ')
  views=tmp[len(tmp)-2].replace(",", "")
  tmp = tmp[:-2]
  for i in range(len(tmp)):
    try:
      int(tmp[i])
      published = " ".join(tmp[i:])
      break
    except:
      pass
  seconds_since_publication = round((datetime.datetime.now() -
dateparser.parse(published)).total_seconds())
  recommended = 1 if second_meta=='Recommended for you' else 0
  data.append([user, video_id, title, channel, views, seconds_since_publication,
recommended, counter])
  counter = counter+1
```

```
for i, video in enumerate(soup.select('ytd-compact-video-renderer')):
  link = video.select_one('a.yt-simple-endpoint.style-scope.ytd-compact-video-renderer')
  video_id = link.get('href').split('?')[1].split('&')[0][2:]
  title = link.select_one('h3 span#video-title').string.strip()
  title_meta = link.select_one('h3 span#video-title').get('aria-label').strip()
  meta = link.select_one('div.ytd-video-meta-block');
  channel = meta.select_one('yt-formatted-string.style-scope.ytd-channel-name').string;
  second_meta = meta.select_one('div.ytd-video-meta-block span.ytd-video-meta-block').string.strip()
  tmp = title_meta[len(title)+4:].strip().split(' ')
  views=tmp[len(tmp)-2].replace(",", "")
  tmp = tmp[:-2]
  for i in range(len(tmp)):
    try:
      int(tmp[i])
      published = " ".join(tmp[i:])
      break
    except:
      pass
  seconds_since_publication = round((datetime.datetime.now() - dateparser.parse(published)).total_seconds())
  recommended = 1 if second_meta=='Recommended for you' else 0
  data.append([user, video_id, title, channel, views, seconds_since_publication, recommended, counter])
  counter = counter+1
```

```
df = pd.DataFrame(data, columns=['user', 'video_id', 'title', 'channel', 'views',
'seconds_since_publication', 'recommended', 'order'])
df.to_csv(file+'.csv', index=False)
# print(df)
```