Utrecht University | Faculty of Humanities | New Media & Digital Culture

Curation or Control:

Analyzing the Socioeconomic Context of Recommendation Engines for Cultural Content

Master's Thesis

Written by Hester Julia Voddé
Student number: 9788573

Date: 27-12-2021

Daily supervisor: Tim de Winkel MA
Supervisor: Dr. Mirko Schaefer
Second reader: Dr. Imar de Vries

Curation or Control:

Analyzing the Socioeconomic Context of Recommendation Engines for Cultural Content

Master's Thesis

Written by Hester Julia Voddé

Student number: 9788573

Date: 27-12-2021

Citation Style: MLA 8

Daily supervisor: Tim de Winkel MA

Supervisor: Dr. Mirko Schaefer

Second reader: Dr. Imar de Vries

Utrecht University

Faculty of Humanities

MA New Media & Digital Culture

Acknowledgements

My deepest gratitude to:

Tim de Winkel for being a competent and inspiring supervisor;

Nicky Heijmen for designing the beautiful cover page for this thesis and being the highlight of this entire Master's together with Julia Alexeenko;

Rosa-Menno, Puneet, Marlijn, Maaike, Karin, Niek, Rob, Mila, Susanne, Richard and Jules for your kindness and everlasting support. I highly appreciate you all - you know this.

So it goes. Thank you for reading.

Table of Contents

Introduction

Recommendation engines: A brief introduction

A considerable part of our daily lives has by now moved to the online sphere: the news, the weather, train departure times and entire films or albums have become available to many people in just a few clicks or swipes. In order to be able to navigate through this increasing amount of content due to further datafication and digitization, various algorithms have become more and more important in helping us select the most relevant information. According to Tarleton Gillespie, leading academic in the field of science and technology studies and principle researcher at Microsoft Research, these algorithms have become "a crucial feature of our participation in public life" (Gillespie, "Relevance of Algorithms" 2). The specific algorithms that are developed to select and suggest personalized content are called recommendation engines. "These recommendation engines not only algorithmically anticipate what 'people like you' desire, they nudge users to explore options and opportunities that might never have crossed their minds" (Schrage ix).

More recently however, these recommendation engines have become the topic of critical debate - both in public and academic spheres. They are no longer only seen as a useful way of navigating through vast amounts of content, but also as tools that could cause concrete societal problems. The conscious design of these recommendation engines could lead to its users being presented with less diverse content, since users are mostly recommended content that they are already interested in. In this way, users would only see things that correspond to their interests and worldview, without being confronted with other information from time to time. Recommendation engines would therefore be the cause of various potential societal dangers. Examples of this are the slightly debated filter bubble, as coined by internet activist Eli Pariser, and echo chamber, as introduced and popularized by legal scholar Cass Sunstein. Concepts stemming from psychology and medical studies, such as confirmation bias (Peter Wason) and popularity bias (David Sackett), are now often applied to the context of new media to study the consequences of algorithmic recommendations that could lead to the increased spreading of fake news, dis- and misinformation and even to radicalization. These information spheres - consisting of news, information and communication - on which these imaginaries of the possible dangers are often projected could have a direct societal impact. The potential negative consequences of these recommendation engines should therefore be carefully studied.

Interestingly enough, the information sphere is not the only context in which recommendation engines are implemented on a large scale. Recommendation engines have become indispensable when it comes to services and platforms that provide artistic and cultural content as well. Important players in this field, such as Spotify (music) and Netflix (films and TV series), greatly rely on their recommendation engines. Apart from the difference in context in which these recommendation engines for both information and cultural content are used, their general workings are rather similar as they are built with the same goal in mind: making sure that users stay on or make use of a certain platform or website as long as possible by suggesting content that they might find relevant or interesting (Knijnenburg et al. 11).

At the end of April 2020, a group of researchers at Spotify published their research *Algorithmic Effects on the Diversity of Consumption on Spotify*. Whereas many online platforms profit from algorithmically recommending their users more of the same content and therefore decreasing content diversity, this publication claims that Spotify makes more profit when the diversity of the music that users listen to is increased. In a section of the paper where business metrics are combined with musical diversity, it is stated that "generalist users [those who have a more diverse musical diet, HV] are up to 25 percentage points less likely to leave the platform and up to 35 percentage points more likely to become a paid subscriber. Thus, our key user metrics are very strongly associated with *diverse* listening patterns" (Anderson et. al 2156). This puts Spotify in a rather unique position as compared to those platforms that recommend their users more of the same content in order to keep them on the platform.

Even though recommendation engines are unmissable when it comes to recommending either news or cultural content, there seems to be a different discourse for both spheres. The use of recommendation engines for selecting and recommending news and information and the potential problems that come with it are elaborately discussed and studied in public and academic spheres, but when it comes to the possible negative consequences of recommendation engines for cultural content both the public and academic debate seem to be largely absent. Even given the fact that the problems that may arise in the information sphere might be of a different size than those in the sphere of cultural content, the remarkable difference in expectations is still interesting. Is the final impact of these recommendation engines for cultural content that much smaller, are journalists more vocal about their personal experiences in the field of news and information or is it a matter of waiting for the same sort of polarization in the area of cultural content? Or are there socioeconomic causes that might help in explaining or interpreting this difference in expectations?

Academic research about recommendation engines highlights the importance of their socioeconomic context, as well as the algorithms that they consist of. According to Gillespie, algorithms are "inert, meaningless machines until paired with databases upon which to function" and in order for algorithms to function, certain choices need to be made (*Relevance of Algorithms* 4). Usually, these choices are in line with the bigger strategy of a platform: "Any knowledge system emerges amidst the economic and political aims of information provision, and will be shaped by the aims and strategies of those powerful institutions looking to capitalize on it" (*Relevance of*

Algorithms 11). Even when algorithms or recommendation engines are black-boxed and unavailable to the public, their workings are influenced by and visible in the socioeconomic context in which they are designed and implemented. I shall come back to this in more detail in my theoretical framework.

Do recommendation engines for cultural content depart from the same socioeconomic principles, but do we see them as less dangerous? Are the business models, ownership status and governance of recommendation engines for cultural content that much different? And if so, would this mean that these recommendation engines could provide a template for a better system of recommendations? These questions have led to the design of the following research question that I shall answer in this thesis:

> *What do the socioeconomic elements of recommendation engines for cultural content look like, and what do they mean in regards to the differences in academic and public debate about recommendation engines for cultural content and recommendation engines for news and information?*

Introduction of theoretical framework and methodology

Through the framework of José van Dijck's analysis of platforms as distinct microsystems I shall study recommendation engines in their appropriate context. In doing so, my thesis shall take on the shape of a literature review. In her book *The Culture of Connectivity: A Critical History of Social Media*, Van Dijck states that platforms should be analyzed "as distinct microsystems that together form an ecosystem of connective media". Van Dijck proposes a model to disassemble platforms into their techno-cultural components and socioeconomic elements. In order to answer my research question, I shall focus on the latter. Practically, this entails that I shall focus on Spotify as a case study and analyze the platform's business models, ownership status and governance and ultimately connect those to its recommendation engines. Following Van Dijck, my methodology shall be in the field of political economy as described by sociologist Manuel Castells. The sub-questions that I have designed to guide my analysis are as follows:

1. *What are the socioeconomic elements of Spotify in terms of ownership status, governance and business models?*
2. *How do these findings fall within the larger debate surrounding the socioeconomic context of recommendation engines for information and news?*

I shall come back to these questions in more detail in the methodologies section.

Academic relevance

With this thesis I aim to stress the importance of analyzing technological developments in the appropriate context. While different types of recommendation engines are often developed with the same general goal in mind, their exact workings and real-life consequences largely depend on the socioeconomic context in which they are implemented. By taking Gillespie's approach and unpacking the 'warm human and institutional choices' (*Relevance of Algorithms* 4) that lie behind algorithms, I will be able to connect both technological and institutional choices. This enables me to see technological developments in their grander socioeconomic context. By analyzing a different field than is the norm in the general debate, I aim to make more clear the importance of a general socioeconomic context in technological developments.

As stated before, a focus on the potential negative consequences of recommendation engines for cultural content seems to be largely absent from both public and academic debate. This absence seems rather strange, as the cultural and creative industries are an elaborate field. As journalists have extensively written about their experiences with negative consequences of recommendation engines in their field, I aim to shed light on another field by analyzing the socioeconomic elements of recommendation engines for cultural content. The cultural and creative industries tend to be an often overlooked but equally important field that should be added to this debate. Adding this creative sphere will help in gaining a better understanding of the extent to which analyzing socioeconomic elements in general is a useful way of giving meaning to a rapidly growing academic and public debate. This thesis is an explorative research into the alleged influence of varying socioeconomic elements on this debate surrounding the potential negative consequences of recommendation engines by analyzing a platform from a different industry.

Furthermore, with this thesis I aim to make more clear the extent to which specific types of content influence the debate surrounding recommendation engines. Diving into the specifics of content or analyzing platforms in light of overarching institutional agencies requires different theoretical and methodological approaches. By analyzing platforms that specialize in both news and information and music, I will be able to recommend whether it is productive to focus more on specific types of content or on the bigger socioeconomic context of platforms in general for future research.

Finally, the overarching aim of this thesis is an exploration of socioeconomic elements of recommendation engines for cultural content in general as a blueprint for sustainable, futureproof recommendation engines that will be able to endure more intense future criticism and debate. I would like to argue that being more aware of the socioeconomic elements surrounding specific recommendation engines that have not been studied extensively so far increases the understanding of the future direction of the general tone of the debate as well.

Theoretical Framework

Introducing: Recommendation engines

Recommendations are as old as time. The word "recommendation" stems directly from the Medieval Latin word *recommendatio*, meaning something as "act of representing in a favorable manner". The current definition of the word is not far removed from its original meaning: "Recommendation: A suggestion or proposal as to the best course of action, especially one put forward by an authoritative body" (Oxford Languages). Interestingly enough however, even though the meaning of the concept itself has remained, the definition of this "authoritative body" providing the recommendations has broadened over time. While recommendations used to be put forward from person to person, relatively new developments have paved the way for algorithmically automated recommendations. These new authoritative bodies are recommendation engines (or recommender systems): "Recommender Systems are software tools and techniques providing suggestions for items to be of use to a user" (Ricci et al. 1). Their function is to predict what a certain user is interested in because of different machine learning algorithms that these engines are built on. The first primitive versions of recommendation engines are mentioned in research papers related to both cognition science and information retrieval.[1]

As recommendation engines are currently used in many fields and industries for varying purposes, there are many types that make predictions by means of (combinations of) different algorithmic models. At the base of these models are two directions of recommendation engines that have evolved over time: collaborative filtering and content-based filtering (Apáthy). The first model, collaborative filtering, predicts "a person's affinity for items or information by connecting that person's recorded interests with the recorded interests of a community of people and sharing ratings between like-minded persons" (Herlocker et al. 241). The system is thus built on similarities between people with similar tastes: "A fundamental operation in memory-based CF is the search for nearest neighbor (NN) users (UB CF) or items (IB CF), whose ratings are then combined to generate predictions" (Nanopoulos et al. 293). According to Larry Hardesty, editor of the Amazon science blog, collaborative filtering is "the most common way to do product recommendation online" (Hardesty). Amazon has implemented collaborative filtering models as (a part of) their recommendation system since 2003. Other services such as YouTube and Netflix have reported using the same model (Smith

[1] It can be argued that Grundy, "a system that plays librarian", was the first version of a recommendation engine (Adomavicius 745). Grundy was introduced in the 1979 research paper *User Modelling via Stereotypes* by computer scientist Elaine Rich. Grundy is a system that "builds models of its users, with the aid of stereotypes, and then exploits those models to guide it in its task, suggesting novels that people may find interesting" (Rich 329). Even though Grundy was a rather primitive system, it paved the way for the various elaborate recommendation engines that we rely on nowadays.

13). The second model, content-based filtering, is built to match knowledge of various characteristics of the object to be recommended with the users' preferences for these characteristics. "Traditional content-based filtering methods usually utilize text extraction and classification techniques for building user profiles as well as for representations of contents, i.e. item profiles" (Maidel 303). American media service Netflix uses content-based filtering (amongst other models) to classify characteristics of their series and movies to be able to recommend them to the appropriate users (Sheikh 116). This is another way to approach providing suggestions that might be interesting for a user. When studying these different approaches however, it is important to clearly make a distinction between both recommendation engines themselves and the algorithms that the systems (partially) consist of. These algorithms are often black-boxed and not available to the public because of commercial interests.

Mapping the debate: Recommendation engines for news and information

Recommendation engines are widely used in the information sphere. By algorithmically curating relevant content for each individual user of a certain website or platform, they turn the consumption of news into a personal experience. These recommendation engines for news and information are impactful and generally well-known: they are the ones that first come to mind when thinking about algorithmically curated content. Despite their popularity however, as of recently these recommendation engines have been the topic of serious debate in both public and academic spheres. In this section of my thesis I shall provide an in-depth overview of this critical academic debate.

When it comes to the potential negative consequences of recommendation engines for news and information that could impact both individual users and society at large, many publications focus on the political economy and socioeconomic context of the platforms that use these recommendation engines. Taking this broader socioeconomic context into consideration when analyzing recommendation engines is highly important, as it is almost impossible to approach them as a separate phenomenon. Recommendation engines are not just multiple lines of code that recommend content to users; they are carefully designed systems that are subject to the bigger context of the platform on which they are implemented. Recommendation engines use machine learning algorithms that work in relation to the databases that they are connected to. In order for these algorithms to properly function, certain choices need to be made. Examples of this are the way in which these databases are ordered - what is being included and excluded - and how much knowledge about its users is taken into account when recommending new content. Furthermore, choices must be made in regards to the meaning of what kind of content is "relevant". According to Gillespie, "relevant" is a fluid and loaded judgement that is open to interpretation:

> As there is no independent metric for what *actually* are the most relevant search results for any given query, engineers must decide what results look "right" and tweak their algorithm to attain that result, or make changes based on evidence from their users, treating quick clicks and no follow-up searches as an approximation, not of relevance exactly, but of satisfaction (*Relevance of Algorithms* 10).

All of the above choices and decisions are supposed to be in line with the general strategy of a platform.[2] This general strategy is not always instantly clear, as most platforms in the field of news and information are big commercial platforms with complicated ownership structures. Still, even though algorithms are often black-boxed, their workings are continually influenced by and therefore visible in the socioeconomic context in which they are designed and implemented. The order in which search results on Google News are presented is an example of this. Even though the long list of equally presented search results might evoke the idea of neutrality, it is yet another example of algorithmically curated content that is in line with a bigger socioeconomic context:

> As in the case of Google News, filter settings limit the sources of news but within these settings the algorithm selects content that fits with user interests and is popular among other users and generates more interaction between users elsewhere in the social network. More interaction generates more traffic, more user attention and more advertising revenue for the social media network. (…) News publishers can partially monitor but have no control over these social media propagation mechanisms (Aguiar et al. 22).

The more “relevant” certain content is to a user, the bigger the chance that this content appears earlier due to algorithmic processes. It is important to note that this “relevance” to a user might have a different meaning for both the user and the platform itself, on which content that generates more traffic is also seen as relevant.

Now that it is established that both algorithms and recommendation engines are inseparable from and continuously influenced by their socioeconomic context, it is important to understand that this is also the case when it comes to the potential negative consequences of these systems. In reality however, more often than not only the algorithm itself seems to be blamed for this even though the ownership status, governance and business models of platforms are also involved. The

---

[2] As I have cited before: “Any knowledge system emerges amidst the economic and political aims of information provision, and will be shaped by the aims and strategies of those powerful institutions looking to capitalize on it” (Gillespie, “Relevance of Algorithms” 11). The specific meaning of ‘relevance’ that is at the base of algorithms and recommendation engines is intertwined with the business strategy of a platform or company.

shift from newspapers to online publications is the result of a significant interplay between these three elements, among others, and is therefore a fitting case. When it comes to the distribution of news, news outlets used to be in charge of both the production and distribution of their content. Nowadays, they have weaker control over the latter (Aguiar et al. 19). The process of curation and distribution now largely depends on black-boxed algorithmic processes, turning the consumption of news into a specific service: that of a personalized experience. The fact that these algorithmic processes are black-boxed however does not mean that this content is not regulated. More often than not, this content itself is what drives the business models of many online platforms.

When it comes to Facebook, the fabric of social interaction is monetized by turning all interaction that takes place on the platform into data, metadata, attention, affect and engagement (Schwarz 119). This data is then captured and sold to advertisers.[3] In order to be able to sell this content to advertisers, platforms need to make sure that their content is properly governed. Most of this governing is also done by means of algorithms: "Algorithms play an increasingly important role in governance, especially online, where social interactions are fully captured as digital data, thus rendered simultaneously economically exploitable, knowable and governable by algorithms" (Schwarz 127). Users must accept and comply to the platform's Terms of Service. Facebook's Terms of Service consists of five different levels: Facebook Principles, Statements of Rights and Responsibilities, Privacy Policy, Data Use Policy, and Platform Policy. Furthermore, Facebook's Terms of Service change frequently and users are never notified of these changes (Van Dijck 59). As these Terms of Service are rather opaque but quite strictly governed by Facebook, they are a common cause of complaints for users. As more processes rely on a platform's black-boxed algorithms, the process of curating and distributing content becomes more unclear and is therefore more influenced by socioeconomic models of these platforms rather than journalists who used to be largely responsible for this. Furthermore, articles do no longer only appear in a set format such as a newspaper: because of the fluid and dynamic structure of the internet, news and information are shared amongst multiple platforms. This makes it even more difficult for journalists and news outlets to keep track of the afterlife of their articles.

It is not too difficult to imagine how these developments might lead to concrete societal problems. Looking at a famous example, we see how all these aspects interplay. As of October 2021, Facebook whistleblower Frances Haugen leaked tens of thousands of internal Facebook documents to the press in an attempt to shed light on the company's alleged damage to society. In the following court cases, which had as its unofficial mantra "Facebook puts profit before its people", Haugen

[3] A grand total of 98.3 percent of Facebook's 2017 revenues came from targeted advertising that was made possible through this monetized data (Schwarz 121).

claimed that the platform's strategy for further expanding and increasing their profit would make it a more unsafe and harmful place. According to her, this mostly has to do with Facebook's current algorithmic engagement-based ranking models that "determine which content they believe is most relevant to users' interests. After taking into account a post's likes, shares and comments, as well as a user's past interactions with similar content, the algorithms powering someone's Twitter feed or Facebook's news feed will place posts in front of that person" (Mac). This algorithmic emphasis on the users' interests shapes a highly personalized approach to curation. This personalization of news and information plays a central role in public and academic debates, with various articles being published in regards to the potential dangers of filter bubbles and echo chambers (Bozdag 2013; Pariser 2011; Gillespie 2014). Even though the effects of the filter bubble have not been verified by media scholars (Lewis et al. 72; Muis et al. 24), the phenomenon itself has gained enough attention to be an important idea in the public discourse. "Regardless of the actuality of the effect, the *idea* of the Filter Bubble has gained a lot of traction in popular opinion (…)" (Knijnenburg et al. 12). A certain form of polarization however is still difficult to avoid: "A platform that delivers news through a filter that is heavily weighted towards personally focused algorithmic values may bake this potential for polarizing personalization directly into its design" (DeVito 767).

Furthermore, Haugen pointed out that Facebook's engagement-based ranking could still be harmful for its users. When asked about the dangers of engagement-based ranking in court, she stated that: "Facebook knows that content that elicits an extreme reaction from you is more likely to get a click, a comment, a reshare" (Mac). According to her, these clicks, comments and reshares are not even beneficial to Facebook's users: Facebook is aware that these interactions work as an encouragement for other users to produce even more content. Instead of presenting content in chronological order, Facebook uses engagement-based ranking to present their users with content. This has proven to be more effective in getting users to spend more time on the platform - even if this time is spent reading and sharing more extreme content. As Facebook is an advertisement-based platform, more active users simply means more profit. This also means that when users indicate that their recommendations are not "correct", "relevant" or "accurate" since they are not satisfied with the recommended content that they see, it is important to keep the actual goal of these recommendation engines in mind. In the end, these recommendation engines are developed with the purpose of increasing interactions and therefore advertisement revenues - not necessarily with the purpose of giving their users the content that appeals to them. Still, my hypothesis is that this specific situation might be rather different when it comes to recommendation engines for cultural content. I shall therefore scrutinize the socioeconomic elements of music streaming service Spotify and uncover to which extent they create a different situation than is the case for recommendation engines for news and information.

Recommendation engines and cultural content: Introducing Spotify

In December 2018, eminent media scholar Lev Manovich created a taxonomy of the various uses of artificial intelligence in the cultural industries, a concept that he called "cultural AI". According to him, cultural AI is currently used for selecting content from larger collections, targeting content, assistance in creation/editing of new content and fully autonomous creation (Manovich 5).[4] According to Manovich, the use and influence of artificial intelligence in the cultural spheres "continues to grow" (5). Although most of the academic debate is focused on research in regards to recommendation engines for news and information, it is therefore also important to take a closer look at the ways in which recommendation engines are used in the spheres of cultural content.

Swedish music streaming platform Spotify is a perfect example of the increasing importance of AI in the cultural industries - especially when it comes to "selecting content from larger collections" (Manovich 5). Spotify was founded in 2006 and launched in 2008 by Daniel Ek and Martin Lorentzon and is currently "the world's most popular audio streaming subscription service" with over 365 million monthly active users and 165 million subscribers as of 2021. Currently, there are more than 70 million tracks available on the music streaming service (Spotify Newsroom). Although Spotify is known as a service that has various ways of recommending new music, their focus on recommendations still is a relatively recent development. Whereas Spotify functioned as an elaborate music library in which the user could actively search for music up until 2011, general artistic changes in the industry as a result of increasing digitization inspired the service to implement new ways of recommending songs to users. From this moment on, streaming services would often start competing with their best music recommendation features. Whereas Spotify used to be praised for being a clean and simple music service, it was now criticized for not actively recommending the right music to its users (Eriksson et al. 59).[5] Even though this focus on personal recommendations was not originally engrained in the company's DNA, Spotify now seems to take pride in their extensive knowledge about their users. Statements such as "We know our fans. Like, really know them." often appear on the company's website for advertisers.

Even though Spotify hardly ever releases official information regarding their recommendation engines, many users and enthusiasts have tried to unpack these black-boxed algorithms. According to a popular article published by Clark Boyd on Towards Data Science, Spotify is said to work with three types of recommendation models: collaborative filtering, Natural Language Processing, and audio models that are used on raw audio (Boyd). In the case of Spotify, collaborative

[4] Although my research focuses only on the first category - the selection of content from larger collections - this taxonomy does show the increasingly important influence of artificial intelligence on cultural field at large.

[5] Examples of features that have their roots in this time period are the weekly "Discover Weekly" playlist, six different "Daily Mix" playlists that change every day and general recommendations based on artists and albums that a user listens to.

filtering models - which have already been discussed[6] - are used to develop individual “taste profiles” for each user that consist of the things they listen to, how they listen, when they listen etc. The collaborative filtering model compares these taste profiles and creates groups of similar users. If a certain user has never listened to an artist that other users from a similar taste profile group often listen to, the collaborative filtering model therefore suggests this artist. The second model, Natural Language Processing (NLP), does not focus on the users themselves but rather on the music and musical context.[7] Although this model is originally used for text rather than music, it is still a model that has proven to be quite helpful in Spotify’s case when it comes to providing more accurate recommendations. The model takes into account various textual clues such as song- and playlist titles that can be found on Spotify itself, and scrawls the internet looking for more contextual information such as song lyrics, news regarding artists and album reviews. This textual information provides important context about factors such as mood and genre. Finally, Spotify uses audio models on raw audio in order to get a better understanding of musical elements such as key, tempo and loudness. The combination of these three models allows Spotify to have a deep understanding of its users and their behavior, context surrounding songs and actual musical details of these songs themselves. Although Spotify has never officially confirmed working with these three models, they have been explicit about the consequences of their recommendation engines.

As mentioned earlier, a group of researchers from Spotify published their paper *The Algorithmic Effects on the Diversity of Consumption on Spotify* in April 2020. Interestingly enough, it is claimed in this paper that Spotify’s “key user metrics are very strongly associated with *diverse* listening patterns” (Anderson et. al 2156). This appears to be a rather unique position for Spotify to be in, as many platforms tend to profit from recommending their users more of the same content in order to engage them and keep them on their platform as long as possible - as mentioned in the case of engagement-based ranking on Facebook. This unique position of Spotify and the general absence of the platform in the debate surrounding recommendation engines is rather interesting. Clearly, this situation cannot be explained by only analyzing these recommendation engines themselves as they are used by various platforms and are heavily intertwined with the bigger context of these platforms. Instead, when trying to get a better understanding of this situation both the algorithms and recommendation engines themselves, as well as their bigger socioeconomic context, should be analyzed.

[6] Collaborative filtering is a widely used type of recommendation engine that predicts “a person's affinity for items or information by connecting that person's recorded interests with the recorded interests of a community of people and sharing ratings between like-minded persons” (Herlocker et al. 241).
[7] Natural Language Processing is a “theoretically motivated range of computational techniques for analyzing and representing naturally occurring texts at one or more levels of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks or applications” (Liddy 2).

Opening up the black box: Studying algorithms, recommendation engines and platforms

Even though there are various types of recommendation engines that are used in many fields, their exact workings are often black-boxed. These recommendation engines consist of algorithms that are a result of specific choices depending on the greater context in which they are implemented. In order to get a better grasp on the workings of these algorithms and the recommendation engines at large, it is therefore important to understand the different factors at play in this broader context. However, provided that these recommendation engines themselves and the algorithms that they consist of are black-boxed and only its broader context is visible, how does one go about analyzing them?

When it comes to algorithms, a more obvious way to study them would be by dissecting and analyzing them on the level of code to discover exactly how they are built to function. However, their often black-boxed nature does make this a rather unrealistic approach (Bucher, "Algorithmic Power" 1176). Furthermore, algorithms are often complex and constantly changing systems: "Compounding this lack of access, the algorithms in question also change rapidly and without notice and are sometimes not fully understood even by their creators" (DeVito 754). Nonetheless, these difficulties do not make it any less relevant or important to analyze algorithms according to Tarleton Gillespie. In his 2013 essay *The Relevance of Algorithms*, Gillespie discusses "the need for an interrogation of algorithms as a key feature of our information system" (3). Interestingly enough, Gillespie proposes an analysis of algorithms "that does not conceive of algorithms as abstract, technical achievements, but suggests how to unpack the warm human and institutional choices that lie behind them" (4). From this perspective, studying algorithms does not involve opening up the black box but actually looking at the real life implications that they have. When analyzing algorithms in this light however, a relevant danger is that of an assumed algorithmic neutrality. Algorithms are often considered to be unbiased by the public, and have a "technologically inflected promise of mechanical neutrality" (Gillespie 16). This is an often-heard misconception as algorithms are not a standalone tool. They are drenched in value judgements, biases and assumptions: "These implicit value judgements start long before a simple line of code is written" (DeVito 756). They should therefore be studied in the context in which they are developed and applied. In this case, that means in the context of platforms.

In his article *The politics of ´Platforms´*, Tarleton Gillespie aims to define the term "platform": a concept that is ambiguous in its meaning. In the oldest sense of the word, "platform" is used in the context of architecture.[8] In a more recent computational sense, platforms are defined as "an infrastructure that supports the design and use of particular applications, be it the computer hardware, operating systems, gaming devices, mobile devices or digital disc formats" (*Politics of*

[8] In this case, the definition of the word is as follows: "A raised level surface on which people or things can stand, usually a discrete structure intended for a particular activity or operation" (Gillespie, "Politics of Platforms" 349).

'Platforms' 349).[9] As my analysis shall focus on platforms in a computational context, I shall depart from this definition. Van Dijck further expands this definition by putting more emphasis on how platforms function as "a *mediator* rather than an intermediary: [a platform] shapes the performance of social acts instead of merely facilitating them" (Van Dijck 29). This process of a platform as a mediator is as follows:

> Technologically speaking, platforms are the providers of software, (sometimes) hardware, and services that help code social activities into a computational architecture; they process (meta)data through algorithms and formatted protocols before presenting their interpreted logic in the form of user-friendly interfaces with default settings that reflect the platform owner's strategic choices (29).

As can be seen in this definition, platforms are complex and multi-facetted systems that are difficult to unpack and analyze. In order to be able to analyze them, they need to be taken apart on various levels.

In her work *The Culture of Connectivity: A Critical History of Social Media,* José van Dijck introduces a multi-layered model that dissects individual platforms - also referred to as microsystems - as techno-cultural constructs and socioeconomic structures. When studying the techno-cultural components of a platform, there needs to be a focus on the technology itself, the content that appears on a platform, and on its users. Focusing on the socioeconomic structure of a platform, the elements that need to be analyzed are the platform's ownership status, governance and business models (28).[10] The model that Van Dijck introduces, considers the importance of analyzing the bigger context of a specific technological phenomenon. Van Dijck's analytical model makes use of two methodological approaches that each play a role in disassembling platforms: actor-network theory, as developed by Bruno Latour, Michel Callon and John Laws, and political economy as theorized by Manuel Castells. Although there is much more to say about the use of both methodologies, I shall first focus on the role they will play in my own analysis.

[9] More recently, the term is also used in a more figurative way: "We now refer to the issues a political candidate or party endorses as their platform" (Gillespie 350).

[10] This is a similar approach that scholars of Critical Data Studies take when analyzing "the ways in which data are generated, curated, and how they permeate and exert power on all manner of forms of life" (Iliadis et al. 2). In doing so, Critical Data Studies scholars study "data assemblages", which are "the technological, political, social and economic apparatuses and elements that constitute and frame the generation, circulation and deployment of data" (Iliadis et al. 2).

Methodology

I have yet established that the recommendation engines that I am interested in are inseparably intertwined with the bigger socioeconomic context of the platforms on which they are implemented. This means that I shall analyze the socioeconomic elements of both platforms for news and information and cultural content to get a better understanding of the differences in the public and academic debate. I have constructed the following research question:

> *What do the socioeconomic elements of recommendation engines for cultural content look like, and what do they mean in regards to the differences in academic and public debate about recommendation engines for cultural content and recommendation engines for news and information?*

In order to be able to systematically answer this research question, I have designed a set of sub-questions:

1. *What are the socioeconomic elements of Spotify in terms of ownership status, governance and business models?*
2. *How do these findings fall within the larger debate surrounding the socioeconomic context of recommendation engines for information and news?*

It is important to point out that I shall use a specific part of the analytical model that José van Dijck introduces in her work *Culture of Connectivity*. This multilayered model is designed to "understand the coevolution of social media platforms and sociality in the context of a rising culture of connectivity" (28). The design of this model is twofold: first, it concentrates on dissecting platforms as both techno-cultural constructs and socioeconomic structures. Secondly, the model is used to systematically connect this platform to the larger ecosystem of which it is a part (29). As I am interested in the socioeconomic elements of recommendation engines specifically, I shall focus on dissecting Spotify as a socioeconomic structure only. Analyzing the platform as a techno-cultural construct would be beyond the scope of this research. When dissecting Spotify's broader socioeconomic context, I shall therefore scrutinize the platform's ownership status, governance, and business models.[11]

[11] It is important to note that in this case "governance" refers to governance *by* platforms, rather than governance *of* platforms. Instead of researching the legal policies imposed upon platforms by governments, I shall research the rules that platforms have imposed upon their own users for various reasons, that range from meeting legal requirements to placating advertisers eager to associate their brands with a healthy online community (Gillespie, "Governance of and by Platforms" 255).

As Van Dijck's model consists of multiple elements from different scholarly disciplines, she argues that it is best approached with a combination of two methodologies: both actor-network theory as coined by Bruno Latour and political economy as introduced by Manuel Castells are needed in order to dissect platforms. When it comes to analyzing the techno-cultural constructs of platforms, Van Dijck uses actor-network theory to study the relations between human and non-human actors. This approach is useful when studying the relations between a platform's users and technology, as it aims to "map relations between technologies and people and tries to explain how these relations are both material and semiotic" (Van Dijck 26). However, as my focus shall be on the socioeconomic context of Spotify, my approach will solely be in the field of political economy. Sociologist Manuel Castells draws with this approach upon economics, law, and political science in order to theorize the "political-economic context in which informational networks could grow into powerful industrial players" (Van Dijck 27). Rather than focusing on interactions that users have with a certain platform or technology, political economy draws attention to an overarching institutional agency. Political economists "regard platforms and digital networks as manifestations of power relationships between institutional producers and individual consumers" (27). Castells specifically discusses the changing power relations in the global media industry as a result of new digital communication systems. According to him, "communication networks are largely owned and managed by global multimedia corporate networks", and therefore "the heart of global communication networks is connected to, and largely dependent on, corporations that are themselves dependent on financial investors and financial markets" (Castells, "Communication Power" 424). Castells' accounting of pre-existing and interlinking power structures makes this approach quite useful when it comes to studying socioeconomic elements of a digital platform as a part of a greater context.

In order to analyze platforms in this way, Castells proposes a methodological approach: "We must find the specific network configuration of actors, interests, and values which engage in their power-making strategies by connecting their networks of power to the mass communication networks, the source of the construction of meaning in the public mind" (*Communication Power* 430). In the case of my thesis specifically, this means that I shall first scrutinize the socioeconomic elements as pointed out by Van Dijck - ownership status, business models and governance - of Spotify, and after that connect those to the grander actors, interests and values that engage in their power-making strategies. In doing so, I shall also reassemble Spotify as a platform and connect it to the larger ecosystem of recommendation engines for news and information platforms. With an eye to future research, somebody could take it from here and analyze the techno-cultural aspects of the platform using actor-network theory in order to get an even more complete picture.

Methodology: Analysis

As the socioeconomic context of recommendation engines and the platforms on which they are implemented is the main focus of this thesis, I shall specifically follow the part of Van Dijck´s model that concerns the socioeconomic elements of platforms. In order to answer my first sub-question, "*What are the socioeconomic elements of Spotify in terms of ownership status, governance and business models?",* my analysis will cover all three elements separately with Spotify as a case study. After doing so, I shall follow Van Dijck's approach and reassemble Spotify as a platform again in the conclusion of this thesis. This enables me to see Spotify - and its recommendation engines - as a part of the broader academic and public debate about platforms and the alleged negative consequences that their recommendation engines have on society. In this section of the analysis, I shall answer my second sub-question: "*How do these findings fall within the larger debate surrounding the socioeconomic context of recommendation engines for information and news?".* This sub-question will enable me to zoom out further by connecting my findings from disassembling Spotify to the broader context of recommendation engines for news and information as I have discussed in the literature review in my theoretical framework. In doing so, I will follow Castells' approach of political economy as a methodology that aims to connect a platform or network to the grander actors, interests and values that engage in their power-making strategies. Finally, my findings from first disassembling Spotify's socioeconomic elements and then reassembling the platform and its recommendation engines will inform my main research question: "*What do the socioeconomic elements of recommendation engines for cultural content look like, and what do they mean in regards to the differences in academic and public debate about recommendation engines for cultural content and recommendation engines for news and information?"*. Before it is time to do so however, I shall start by scrutinizing Spotify's socioeconomic elements.

Analysis

Disassembling Spotify: Ownership status

When it comes to disassembling a platform's socioeconomic elements, Van Dijck states that a platform's ownership model is "a constructive element in its functioning as a system of production" (36). It could take on various forms: from "nonprofit, collectively owned, user-centered organizations to for-profit, corporate owner-centered enterprises" (36). The ownership status of a platform is not fixed: it can change over time and it is often influenced by the current position and future plans of a platform. The ownership status of a platform itself also influences the users' experience, as well as the image that the platform owners want the platform to emit. Moreover, "ownership status is a significant bargaining chip in the volatile ecosystem of networked media" (36). This means that a platform's ownership status is influenced by and influences the bigger ecosystem of networked media as well. Therefore, analyzing a platform's ownership status is important to get a better understanding of the current position and future direction of a platform, as well as the platform's position in the bigger ecosystem.

Music streaming service Spotify was founded in Stockholm in 2006 by Daniel Ek and Martin Lorentzon. Interestingly enough, both founders had no experience working in the music industry, but had a background in advertising technology instead. Ek and Lorentzon met through business deals in their respective previous companies, and decided to start a business together. Spotify's initial business structure was not directly straightforward: "Spotify was initially structured as a group of several firms that used a plurality of jurisdictions for owning different assets" (Eriksson et al. 42). This means that Spotify was registered as multiple companies: whereas Spotify AB was a software company, Spotify Sweden AB was another company specialized in selling advertisements in the area of "services for digital distribution of music, film, TV programs, audiobooks, games, and similar content" (42). These different companies were all owned by a bigger holding company, Spotify Technology SA, which is registered at a post-office box in Luxembourg.[12] Ek and Lorentzon still work for Spotify, with Ek being the CEO of the company and Lorentzon being responsible for developing the company's future goals and business strategy.

An invite-only beta version of Spotify was launched in 2007. This version was well received, and a year later the first public version of Spotify was launched in eight European countries. The global rollout of the music streaming service was not without problems: Spotify would often deal with local copyright collecting organizations and as a result (temporarily) retreated from various countries. As of 2021, Spotify is available in 178 countries and has 16 offices worldwide. The company was introduced to the New York Stock Exchange in 2018, and is listed in the form of American Depositary Receipts through their holding company in Luxembourg.

As Spotify is currently a public company, the question "Who owns Spotify?" is a rather difficult one to answer. According to Rolling Stone Magazine, SEC documents show that 65% of Spotify is owned by just six parties at the end of 2019: "the firm's co-founders, Daniel Ek and Martin Lorentzon (30.6 percent of ordinary shares between them); Tencent Holdings Ltd. (9.1 percent); and a run of three asset-management specialists: Baillie Gifford (11.8 percent), Morgan Stanley (7.3 percent), and T. Rowe Price Associates (6.2 percent)" (Ingham). Furthermore, two of the biggest record companies - Universal Music Group and Sony Music Entertainment - also jointly own between

[12] "This post-office box is the same one that Lorentzon uses for another holding company, which in turn owns another company registered to a post-office box in Cyprus, in which Lorentzon has placed the financial riches that he made with TradeDoubler and through which he now owns his shares in Spotfiy: a classic structure for tax avoidance" (Eriksson et al. 42).

six or seven percent of Spotify. This means that over 70% of the company is owned by just a small group of investors. The most interesting part about this ownership status however, is the combined voting power of 77.4% that Daniel Ek and Martin Lorentzon have in the company. Even though they own 'just' 30.6% of the shares, Ek also controls the entire voting power of the 9.1% stake of Tencent Holdings (Ingham). This means that both original founders still have a large influence on the company's current position and future development.

Spotify also has an elaborate history when it comes to acquisitions and company partnerships[13]. The list of companies and startups that Spotify acquired since 2013 is too long to discuss, although all acquisitions could be grouped into three general categories: the improvement of music recommendations, the improvement of (personalized) advertisements, and more recently explorations into the world of podcasts and live audio chats. The many acquisitions made by Spotify could shed light on the general direction the company is heading in, as well as provide an overview of industry trends at a given time. This is also the case when it comes to Spotify's partnerships with other companies. Whereas Spotify's list of acquisitions also includes smaller startups, their partnerships involve mostly well established companies. Most of these partnerships concern the integration of Spotify on other platforms, such as Playstation Music (Sony), Waze and Discord. Other partnerships are in relation to content, mostly in the form of podcasts through a WNYC Studios partnership and a partnership with ESPN and Netflix. Apart from podcasts, Spotify appears to be considering branching out to other media forms as well: in September 2020, Spotify signed a deal with Chernin Entertainment to produce movies and tv shows (Carman).

Disassembling Spotify: Governance

The second socioeconomic element that should be analyzed when disassembling platforms is governance. The governance of a platform is closely related to a platform's norms and values, and is connected to the legal regulations of the broader industry that a platform is a part of. As briefly mentioned before, it is important to point out that there are two types of governance: governance *of* platforms and governance *by* platforms (Gillespie, "Governance of and by Platforms" 255). In the context of this research, the concept of governance will refer to governance *by* platforms.[14] Whereas in the early days of Web 2.0 many websites and blogs had relatively little rules and were often governed by users themselves, these days many platforms are taken over by commercial

[13] The influence of business partnerships on platformization is a relatively new field of research. An important topic of research, among others, is the way in which in business partnerships mediate and shape platform power and governance. See Van der Vlist and Helmond, 2021 and De Winkel, forthcoming.
[14] This entails the rules that platforms have imposed upon their users for various reasons, that range from meeting legal requirements to placating advertisers eager to associate their brands with a healthy online community (Gillespie 2).

corporations that implement more professional forms of governance. A considerable part of this governance consists of what Van Dijck calls "social protocols", which are implicit or explicit rules to manage user activities. These social protocols are often articulated through a platform's end-user license agreements (EULAs) or Terms of Service (ToS) (Van Dijck 28).

According to Van Dijck's definition, EULAs or ToS are "a contractual relationship that users enter into every time they log on to a platform, and these contracts impose restraints and obligations" (38). Although they are a platform's connection to social norms and laws in the real world, EULAs or ToS are not laws themselves. "Like algorithms and user agency, a site's terms of service are an arena for setting and contending social norms, a struggle that may eventually affect legal rulings" (38). While EULAs or ToS are at the base of any platform, they are often left unread by users: many of them are designed to appear opaque to a platform's users because of their extreme length and inaccessible language. Furthermore, they are known to be modified by platforms quite often without users' prior consent as platform owners control EULAs and ToS entirely. Spotify's Terms of Service appear to be relatively short at first sight, with 'only' around 5800 words. This is rather minimal as compared to Facebook's ToS, which is famously said to be longer than the U.S. Constitution[15]. However, this does not mean that Spotify's ToS are instantly clear or user friendly. First of all, the ToS are only explicitly shown to a user when creating a Spotify account, which is not always the right time for users to read various pages of legal information. Secondly, the ToS are rather difficult to understand because of the detailed, legal language and use of specific jargon. Lastly, even though the ToS appear to be relatively short at first sight, they are actually full of links referring to other long pages full of regulations for specific situations. This structure makes it quite difficult to get a proper overview of all regulations that are a part of the Terms of Service. Still, it is mandatory for a potential user of Spotify to accept all Terms of Service: "If you do not agree to these Terms, then you must not use the Spotify service or access any Content".

Spotify's ToS start with a section about the service that the platform provides to its users with a brief mention of the service limitations and modifications, claiming that "our service offerings and their availability may change over time without any liability to you". After this section, a more elaborate one follows concerning "Your use of Spotify". The Spotify user is made aware of the fact that they can only use Spotify after accepting the Terms of Service, and is informed of the billing of a paid subscription, price and tax changes and cancellations. Furthermore, there is a brief mention of "User Guidelines" that a user must follow when making use of Spotify's service. This section links to a new page that contains an extensive list of things a Spotify user is not permitted to do, varying from

[15] According to The New York Times, Facebook's Privacy Policy is 5830 words long, while the U.S. Constitution is a concise 4543 words (Bilton).

reverse-engineering and crawling to providing your password and username to any other person. Strangely enough, this specific section - that only shows up when clicking on a link in the ToS - is the only section in which explicit examples of incorrect behavior are mentioned. The ToS continue with a section about content and intellectual property right, in which it is stated that a Spotify user may upload content to the Spotify Service (including Spotify's Community Forum), and that Spotify might monitor this content - although the company "has no obligation to do so". The content that users upload will also be owned by Spotify, and by uploading content the user grants Spotify the right to: "(1) Use your devices' processor, bandwidth and storage hardware (…), (2) provide advertising and other information to you, and (3) to allow our business partners to do the same." Who exactly these business partners are, is not specified. Finally, the ToS end with an elaborate legal section full of jargon.

The ToS fully apply to a user as long as they use the music streaming service - and Spotify is free to suspend their users if they breach these terms. When it comes to disputes, there are various references to the laws of the State of California: "These Terms are governed by and shall be construed in accordance with the laws of the State of California, United States of America, without regard to California's choice or conflicts of law principles." Apart from its consumers, Spotify has another group of platform users: artists that upload their music to the platform and generate their income in this way.[16] Unsurprisingly, there is a separate ToS for "Spotify for Artists". Spotify for Artists is a section of the platform that is only available to artists, in this case people that upload music to Spotify. Just like on the consumption side of Spotify, this section is only available after agreeing to the ToS. As the ToS for Spotify for Artists are similar to the general ToS, I shall not analyze them separately.

Disassembling Spotify: Business models

The final element that needs to be analyzed are a platform's business models. Business models are not static but constantly changing and evolving, just like a platform's ownership status and governance. Especially in the creative and cultural industries, the rise of the internet has had a considerable impact on the existing business models at the time.[17] The rise of music streaming was a

[16] Spotify has famously received a lot of criticism from artists that use the platform to promote their music because of their payments. Currently, Spotify pays their artists around $0.0033 per stream, which means that a song needs to be streamed 250 times for an artist to make a full dollar (Ennica).

[17] Traditionally, there were three ways of making money with creative and cultural products: "Profits derived from sales of reproduced goods (CDs, books, DVDs), profits from viewing or subscription fees (TV programs, cinema, video rental), and profits derived from advertising, which is basically the selling of audience attention juxtaposed to, or interspersed with, cultural or entertainment content" (Van Dijck 39). The rise of MP3 for example, and later on music streaming, changed the existing business models in the creative industries. These new music formats challenged the idea of buying and owning a product: rather than spending money in order to obtain a physical object such as a CD or vinyl, one would now spend money on a digital file.

particularly difficult change for businesses in the creative industries, since consumers would no longer buy a physical product or even an MP3 file: the purchased content would not even leave the platform itself anymore. With these changes also came changes in business models: platforms started experimenting with either (progressively targeted) advertising or premium models. These changes were also clearly visible in the history and current business models of Spotify.

Spotify deliberately profiles itself as a service rather than a seller of goods. The music streaming service does not only allow for access to a big database of songs, it also algorithmically curates a personalized musical experience for its users and promises to be able to select "music for every mood" (Spotify campaign 2019). This choice is in line with what Bourdieu has previously called "cultural intermediaries": "these 'need merchants', sellers of symbolic goods and services who always sell themselves as models and as guarantors of the value of their products, who sell so well because they believe in what they sell (…)" (Bourdieu 365). This concept was recently expanded by media scholar Jeremy Wade Morris. According to him, there is now an emerging layer of a new type of organizations that could be called "infomediaries": "organizations that monitor, mine and mediate the use of digital cultural products as well as the audience responses to those products via social and new media techniques" (447). This new type of infomediaries has the same general workings as Bourdieu's notion of intermediaries, except for the fact that algorithms and data mining techniques make these infomediaries even more effective and precise in curating and producing cultural content. This means that they are increasingly responsible for shaping how audiences encounter and experience cultural content (456). This renewed definition therefore allows one to see Spotify as a cultural infomediary.

In general, there are two ways in which a Spotify user can access this personalized music experience: either through signing up to Spotify's free version or paying for a subscription to the service. Remarkable here is that all of Spotify's distinctive recommendation features are also available to people with a so-called "Freemium" account. This choice reveals the level of importance given to these algorithmic recommendations by Spotify itself. When paying for a Spotify subscription however, the sound quality of music is better and users can skip songs - a basic option that is not supported in the Freemium model. The biggest difference however is the total absence of audio advertisements in the Spotify Premium subscription. Whereas in the free version users are constantly bombarded with audio advertisements in between songs and pop-up banners, advertisements of any sort will never appear for those who pay for the service. Furthermore, Freemium users are constantly urged to buy a subscription to Spotify, revealing the company's emphasis on being a subscription-based platform. This difference in (the absence of) on-platform advertising is symbolic of the impact that these different business models have on the concrete workings of a platform. When a platform is completely free, like Spotify Freemium, profits have to be generated through selling space to advertisers. When a platform is behind a paywall, most of the profits come directly from users' subscription fees. This makes it less necessary for platforms to incorporate advertisements into the service itself.

Conclusion

In the analysis section of this thesis I have answered my first sub-question by scrutinizing the socioeconomic elements of Spotify in terms of ownership status, governance and business models. However, as this is not enough to answer my main research question, I shall answer my second sub-question in this section to give further meaning to my findings. This second sub-question is as follows: "*How do these findings fall within the larger debate surrounding the socioeconomic context of recommendation engines for information and news?"*. In order to answer this question, I shall reassemble Spotify and place it in the broader socioeconomic context of recommendation engines for information and news.

Reassembling Spotify

The general structure of the ownership status of both Spotify and the field of platforms for news and information is rather similar. While Spotify is a music streaming service and could therefore be seen as a company that is a part of the cultural or creative industries, financially it is a multi-billion dollar company that is no different from other commercial companies in the area of news and information. As can be seen in my analysis, both Spotify and the platforms for news and information that I have discussed are owned by rather big companies that have only increased in size over the years. With this growth also comes the fact that the ownership structure of these companies has become more opaque: in this case, the companies are owned by an elaborate range of private investors and other companies that all have shares. When it comes to partnerships and acquisitions, Spotify could also easily be placed in the same field as companies such as Google or Facebook. Spotify's partnerships and acquisitions are often in the direction of algorithmic recommendations and advertisements and therefore look very much like the partnerships in many other industries: Spotify has even partnered with Facebook in the past. It is fair to conclude that when it comes to the socioeconomic element of ownership status, there are more similarities than differences between Spotify and platforms for news and information. Therefore, this element alone would not be enough to gain a better understanding of the differences in the debate surrounding different types of platforms and their recommendation engines. This ownership status however is an important factor in the future direction of a company - both for Spotify and platforms in other fields - as shareholders often have voting power within a company. This means that in this way, the ownership status is also connected to the general business choices that are made. This will become more clear when looking at my findings regarding the socioeconomic elements of governance and business models.

When it comes to the element of governance by platforms for both cultural content and news and information, the most important similarity is the emphasis on the Terms of Service. These terms must be accepted by the user upon entering the platform and contain an overview of the (legal) rules and regulations of a platform. As is the case with platforms for news and information, Spotify's Terms of Service have a complicated structure and are full of specific language use. While there are definitely generic similarities, the details of the Terms of Service are still different for each platform. As I have pointed out in my analysis, Spotify's Terms of Service only contain a brief mention of the monitoring of its users and their actions on the platform. This is a remarkable difference as compared to the emphasis that platforms for news and information tend to put on the governing of their users. This difference in governance is in turn related to the business models of platforms: platforms such as Facebook are freely accessible to its users and therefore more reliant on their advertising revenue. These personalized advertisements are made possible by turning all interactions on the platform into raw materials that can be used to improve targeting. This means that all interactions on the platform take in a central position, one that is different on platforms like Spotify that mostly work with a subscription-based model. Companies that are more dependent on advertising revenues also have to make sure that their platforms, as well as the content and interactions that are part of it, are suitable to attract new advertisers.

It is important to keep in mind that the lack of focus on monitoring and governing users and content in Spotify's Terms of Service does not mean that this is not happening on the platform at all. Furthermore, the fact that platforms such as Facebook seem to focus more on what their users are doing on their platform does not mean that they have more control over their users than a platform like Spotify. This is due to a difference in platform affordances: Spotify does not allow for messaging other users and only lets users share music outside of the platform, while the affordances of platforms for news and information often allow users to reply to articles and other people's comments. When the content from both platforms is further spread to other platforms or websites, it is difficult for the original platforms to keep up with and monitor the afterlife of this content. Shared music often sparks less critical debate than news articles, which means that the latter are more likely to live their own life after leaving a certain platform.[18]

The final socioeconomic element that I have focused on in my analysis are a platform´s business models. As I have mentioned before, these models are closely intertwined with the other

[18] While it can be discussed if this afterlife is still the responsibility of the platform itself, it does show that more governance on a platform does not mean that these platforms have more control over the users or content itself in the end.

socioeconomic elements of ownership status and governance. Due to increasing digitization, both the creative industries and the field of journalism have had to deal with drastically changing business models. In both cases, there was a general shift from selling physical products, such as CDs or newspapers, to selling digital goods and services. This shift meant a stronger focus on personalized recommendations as a service, both in the case of Spotify and platforms for news and information. In general, the way in which both industries reacted to this change was largely the same: while Spotify is concerned with creative content, it is still as much of a commercial company as most big platforms for news and information. Spotify was therefore as eager to discover the many opportunities of selling personalized music experiences. The underlying business models that shaped the way that these new personalized experiences looked like, however, are different for both types of platforms. While Spotify does offer a functioning Freemium model, the music streaming service is constantly urging its users to buy a Spotify subscription. Spotify currently has 165 million subscribers, and these subscriptions therefore are an important part of the service's revenue. This is in stark contrast to many services for news and information: platforms such as Facebook and Google are free to access without subscribing. Therefore, these platforms mostly rely on advertising revenues. This difference in business models has great implications for the general workings of a platform and its recommendation engines. The key to the differences in the academic and public debate about the potential negative consequences of recommendation engines for cultural content and news and information is therefore also in these business models.

Reassembling the debate surrounding recommendation engines

This seems to me a fitting moment to reiterate my main research question, which is as follows: "*What do the socioeconomic elements of recommendation engines for cultural content look like, and what do they mean in regards to the differences in academic and public debate about recommendation engines for cultural content and recommendation engines for news and information?"*. Spotify is currently in the unique position where their "key user metrics are very strongly associated with *diverse* listening" since users that are interested in a more diverse musical diet are more likely to pay for a subscription to the platform (Anderson et. al 2156). This unique position could be explained by means of their subscription-based model. Because of this model, Spotify does not only have to rely on user engagement in order to sell more advertisements, which is often the case for free platforms for news and information. When users pay for a subscription to a platform, they pay for a certain service. In the case of Spotify, they pay for a personalized music experience. This means that these recommendation engines are built to be accurate rather than engaging. They are designed to work for the user and accurately suggest the things that someone might be looking for, and in the case of Spotify this means new, diverse music. This subscription-based model therefore guarantees a certain kind of quality: if the user is paying for a service, in this case a personalized music experience that is diverse in its recommendations, a platform would ultimately profit from being as accurate with this service as possible. This is in contrast to platforms for news and information that I have discussed: as most of these platforms are freely accessible, they mostly have to rely on user engagement in order to sell more advertisements. The longer users stay on a platform and interact with it in various ways, the more advertisements can be sold. Therefore, the platform's recommendation engines are designed with the goal of engagement in mind. The guarantee of accurate recommendations disappears when a user is not paying for a specific service, and recommendation engines that are implemented on platforms that are dependent on advertising revenues are therefore mostly built to work for the platform itself rather than the user. This combination of engagement rather than accuracy, the disappearance of a guaranteed quality of recommendations and the general black-boxed nature of algorithms is a recipe for popular imaginaries such as the filter bubble, echo chamber, popularity bias or confirmation bias - imaginaries that heavily shape the critical debate surrounding these recommendation engines for news and information platforms.

As is a general consensus in the debate surrounding the potential negative consequences of recommendation engines for news and information, recommendation engines and the algorithms that they consist of are heavily intertwined with a platform's socioeconomic context, and more often than not they are built to work for the platform itself rather than for its users. As I have briefly mentioned in my theoretical framework, I expected this to be different in the field of cultural content. I have found that the socioeconomic elements of my case study Spotify are indeed different as compared to those of news and information. Not only could my findings offer a possible explanation for Spotify's unique position in which the company would profit from more diverse recommendations, but they have also highlighted the potential of scrutinizing socioeconomic elements as a way of interpreting an increasingly important debate. Adding the sphere of cultural content to this debate has further emphasized the importance of "the fluid and loaded judgement of relevance" (Gillespie 10), that seems to become even more fluid when applied to recommendation engines in different fields. Furthermore, with the increasing importance of cultural AI (Manovich) and the further rise of extensive cultural infomediaries (Morris 447) such as Spotify and Netflix this thesis will hopefully assist in the anticipation of their place in future debate surrounding the potential negative consequences of recommendation engines in general.

Finally, I have found that the nature of the content itself seems to be subordinate to these socioeconomic elements in shaping the direction of the debate. In this case, the difference between an advertisement based model or a subscription based model seems to influence the debate surrounding recommendation engines more than the difference between news and information and cultural content. These findings will hopefully offer insight into the recommended points of focus in future research: in line with Castell's works on political economy I have found that an emphasis on the grander socioeconomic context appears to be more productive than scrutinizing the actual content on platforms when making sense of this debate. Having said this however, the findings of this thesis do hint at a logical relation between specific types of content and specific business models. In general, news and information generally tend to invoke more interaction than music, and therefore seem to be more suitable for generating advertising revenues than cultural content. This could be an important given for anticipating the direction of future debate as well.

Discussion

It is not a stretch to assume that the usage and influence of recommendation engines will only increase in the future, both in the field of news and information and the creative industries. These recommendation engines and the platforms that they are a part of will increasingly continue to shape our consumption and experience of news and information and arts and culture. This means that the critical debate surrounding these recommendation engines is quite likely to become more present or intense, too. Something has to change in order to guarantee a more sustainable future of recommendation engines in all industries: a new blueprint needs to be created for recommendation engines that are able to endure or surpass this increasing criticism.

Currently, the general focus of the public debate surrounding recommendation engines seems to be on "the algorithm": an abstract, black-boxed concept that is likely to evoke well known buzzwords such as filter bubbles and echo chambers. In order to develop more futureproof recommendation engines that are less opaque and therefore less likely to be connected to said buzzwords, it is important to make more clear why certain content is shown to a user.[19] Furthermore, the workings of the algorithms themselves need to be changed too. Of course they can easily be tweaked to recommend different things, depending on what is framed as "relevant". This relevance however is more related to the grander socioeconomic context than the algorithms themselves. I would therefore argue that the general emphasis on "the algorithm" in the public debate is ineffective and even counterproductive, as socioeconomic factors mostly influence a recommendation engine's algorithms - not the other way around. As I can conclude at this point, designing futureproof recommendation engines therefore requires a holistic approach and begins

[19] An example of this is presented in the paper *Recommender Systems for Self-Actualization* (Knijnenburg et al. 2016). In this paper, the authors present a "new direction for recommender systems research with the main goal of supporting users in developing, exploring, and understanding their unique personal preferences" (11). These recommendation engines would support users in the process of decision making rather than replacing it by both stating why a certain recommendation was given and actively encouraging users to look further than standard recommendations.

with reshaping socioeconomic factors. Although I do not pretend to be actively reshaping these factors myself, this thesis does count as an exploration of this holistic approach. My scrutinization of these elements is hopefully yet another valuable step in this direction.

While I have tried my very best to write this thesis as thoroughly, critically and carefully as I could, there are still some limitations that I should address. First, in regards to the size of this thesis I have chosen to analyze Spotify as a case study of platforms for cultural content at large. While this platform is important in the field of cultural content and the creative industries, it is still one platform and therefore does not and cannot speak for the entire field. In the same spirit, I have very carefully tried to map the academic debate surrounding the potential negative consequences of recommendation engines for news and information by means of a literature review in the theoretical framework of this thesis. As it is almost impossible to capture an entire academic field, I do acknowledge that some important voices or ideas might be absent in this section.

Second, throughout this thesis I have followed Van Dijck's analytical prism for disassembling and later on assembling platforms as a part of distinct microsystems. While Van Dijck analyzes platforms in terms of techno-cultural and socioeconomic elements, I have made the choice to only focus on the latter due to the topic of this thesis. This means that I have also used only one out of the two methods that Van Dijck proposed as well: while actor-network theory works quite well in uncovering techno-cultural elements of platforms, I have only focused on political economy as put forward by Castells since this is the method that Van Dijck recommends when analyzing socioeconomic elements. Adding techno-cultural elements, and therefore actor-network theory, to this analysis of recommendation engines for news and information and cultural content would be an interesting direction for further research. Furthermore, it is important to point out that both Van Dijck's model and Castells' method are rather descriptive. This means that there is not a single roadmap to follow and the implementation of both may differ slightly depending on the topic of research and general interpretation.

Third, even though I hope that this thesis has made clear the importance of unpacking and analyzing algorithms and recommendation engines as "warm human choices" rather than "abstract, technical achievements" (Gillespie, "Relevance of Algorithms" 4), it is important to point out that this is yet a way to work around a black-boxed nature that is still very much present. As I did not have direct access to Spotify's recommendation engines, I was not able to scrutinize all factors that play a role in recommending music. This means that I might have potentially missed the recommendation engines pushing songs as part of behind-the-scenes business deals in subscriber's playlists, masked as 'neutral' suggestions. I can only speculate about those kind of situations.

I would like to end with some suggestions for future research that build upon the topic and findings of this thesis. A central topic of research that implicitly arose from my findings was that of the potential of subscription based business models or paywalls as platform gatekeepers or guarantors of content quality. The current problems and negative imaginaries that are projected on platforms for news and information would potentially be of a different size if those platforms would have a subscription-based business model that changes the meaning of "relevant recommendations" from "engagement with the platform" to "accuracy and quality". While this is just speculation based on my findings, this is an extremely interesting and arguably important topic of research that could offer valuable insight into the future of digital news consumption.

Another suggestion would be to analyze other platforms for creative content in order to gain more insight into this industry that is still largely absent in this debate about potential negative consequences of recommendation engines. Spotify is definitely a big player when it comes to cultural and creative content, but there are many other platforms that are representative of the creative industries. Furthermore, analyzing especially those platforms that contain a combination of both news and information and creative content, such as YouTube or Instagram, could lead to interesting results in both industries. Adding the second layer of Van Dijck's model and analyzing both the socioeconomic and techno-cultural elements of platforms by means of actor-network theory could lead to even more elaborate results.

Lastly, as I have pointed out, systematic change in the current design of recommendation engines at large can only be reached by changing socioeconomic factors. I have found that Tarleton Gillespie's proposal of analyzing algorithms by unpacking "the warm human and institutional choices that lie behind them" (*Relevance of Algorithms* 4) is crucial and extremely helpful when placing algorithms and recommendation engines in their socioeconomic context, as well as Castell's works on political economy that point out the power relations between institutional producers and individual consumers. By approaching recommendation engines and platforms in this way rather than in a purely technological way, the risk of missing certain overarching connections decreases. These works therefore should be at the center of future research about the socioeconomic context of recommendation engines at large. This thesis can be seen as an exploration of analyzing socioeconomic elements of recommendation engines for cultural content as a way of interpreting an increasingly important debate. As it has highlighted the potential of this approach I would highly suggest building upon this exploration with a similar approach in future research. Let us make sure that the future of recommendation engines is one of curation instead of control.

Bibliography

"About Spotify." *Spotify*, 17 Nov. 2021, newsroom.spotify.com/company-info.

Adomavicius, Gediminas et al. "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions." *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 6, 2005, pp. 734-749. *IEEE Xplore*, doi: 10.1109/TKDE.2005.99.

Aguiar, Luis et al. "The digital transformation of news media and the rise of disinformation and fake news - An economic perspective." *SSRN Electronic Journal*, 2018. *Research Gate*, doi: 10.2139/ssrn.3164170.

Anderson, Ashton, et al. "Algorithmic Effects on the Diversity of Consumption on Spotify." *Proceedings of The Web Conference 2020*, 2020, pp. 2155-2165. *ACM Digital Library,* doi: 10.1145/3366423.3380281.

Anttiroiko, Ari-Veikko. "Castells' network concept and its connections to social, economic and political network analyses." *Journal of Social Structure*¸ vol. 16, no. 11, 2019, pp. 1-18. *Exeley*, doi: 10.21307.

Apáthy, Sándor. "History of Recommender Systems: Overview of Information Filtering Solutions." *Onespire Ltd. - SAP and Management Consulting*, 20 Sept. 2021, www.onespire.net/news/history-of-recommender-systems/.

Bilton, Nick. "Price of Facebook Privacy? Start Clicking." *The New York Times*, 12 May 2010, https://www.nytimes.com/2010/05/13/technology/personaltech/13basics.html.

Bourdieu, Pierre. *Distinction: A Social Critique on the Judgement of Taste.* Translated by Richard Nice. Harvard University Press, 1984.

Boyd, Clark. "How Spotify Recommends Your New Favorite Artist". *Medium*, 2021, https://towardsdatascience.com/how-spotify-recommends-your-new-favorite-artist-8c1850512af0.

Bozdag, Engin. "Bias in Algorithmic Filtering and Personalization." *Ethics and Information Technology* vol. 15, 2013, pp. 209-227.

Bucher, Taina. "Want to be on the top? Algorithmic power and the threat of invisibility on Facebook." *New Media & Society*, vol. 14, no. 7, 2012, pp. 1164-1180. *SAGE Journals*, doi: 10.1177/1461444812440159.

Bucher, Taina. "The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms." *Information, Communication & Society,* vol. 20, no. 1, 2017, pp. 30-44. *Taylor & Francis Online,* doi: 10.1080/1369118X.2016.1154086.

Carman, Ashley. “Spotify Made a Deal to Turn Its Exclusive Podcasts into TV Shows and Movies.” *The Verge*, The Verge, 24 Sept. 2020, www.theverge.com/2020/9/24/21453026/spotify-chernin-entertainment-partnership-first-look-podcast-ip.

Castells, Manuel. *Communication Power.* Oxford University Press, 2009.

Castells, Manuel et al. *Mobile Communication and Society: A Global Perspective.* The MIT Press, 2007.

DeVito, Michael A. “From Editors to Algorithms: A values-based approach to understanding story selection in the Facebook news feed.” *Digital Journalism*, vol. 5, no. 6, 2017, pp. 753-773. *Taylor & Francis Online*, doi: 10.1080/21670811.2016.1178592.

De Winkel, Tim. “Gab as an Ecosystem of Services.” *Fringe Platforms: Researching the role of the contesting alternatives to the mainstream social media platforms*, forthcoming.

Ennica, Jacob et al. “How Much Does Spotify Pay per Stream? What You'll Earn per Song, and How to Get Paid More for Your Music.” *Business Insider*, 24 Feb. 2021, www.businessinsider.nl/how-much-does-spotify-pay-per-stream-what-youll-earn-per-song-and-how-to-get-paid-more-for-your-music/#:~:text=Spotify%20generally%20pays%20between%20%24.,distribution%20contract%20and%20listener%20base.

Eriksson, Maria et al. *Spotify Teardown: Inside the Black Box of Streaming Music.* The MIT Press, 2019.

Gillespie, Tarleton. “The Politics of ‘Platforms’”. *New Media & Society,* vol. 12, no. 3, 2010, pp. 347-364. *SAGE Journals*, doi: 10.1177/1461444809342738.

Gillespie, Tarleton. “The Relevance of Algorithms: Essays on Communication, Materiality and Society.” *Media Technologies*, edited by Tarleton Gillespie et al., The MIT Press, 2013, pp. 2-33.

Gillespie, Tarleton. “Governance of and by Platforms.” *SAGE Handbook of Social Media*, edited by Burgess, Jean et al., 2017, pp. 254-278.

Hardesty, Larry. “The History of Amazon's Recommendation Algorithm.” *Amazon Science*, Amazon Science, 20 Aug. 2020, https://www.amazon.science/the-history-of-amazons-recommendation-algorithm.

Herlocker, Jonathan L. et al. “Explaining Collaborative Filtering Recommendations.” *Proceedings of the 2000 ACM conference on Computer supported cooperative work,* 2000, pp. 241-250. *ACM Digital Library*, doi: 10.1145/358916.358995.

Hesmondhalgh, David. “Bourdieu, the media and cultural production.” *Media, Culture and Society*, vol. 28, no. 2, 2006, pp. 211-231. *SAGE Journals*, doi: 10.1177/0163443706061682.

Iliadis, Andrew et al. “Critical Data Studies: An Introduction”. *Big Data & Society*, vol. 3, 2016, pp. 1-7.

Ingham, Tim. “Who Really Owns Spotify?” *Rolling Stone*, Rolling Stone, 1 July 2020, www.rollingstone.com/pro/news/who-really-owns-spotify-955388/.

Knijnenburg, Bart et al. “Recommender Systems for Self-Actualization.” *Proceedings of the 10th ACM Conference on Recommender Systems*, 2016, pp. 11-14. *ACM Digital Library*, doi: 10.1145/2959100.2959189.

Latour, Bruno. “On actor-network theory: A few clarifications.” *Soziale Welt,* vol. 47, no. 4, 1996, pp. 369-381.

Lewis, Callum et al. *Wie is de waakhond op sociale media? Onderzoek naar de rol van traditionele media en politici op sociale media in verkiezingstijd.* Utrecht Data School, 2017.

Liddy, Elizabeth. “Natural Language Processing.” *Encyclopedia of Library and Information Science*, Marcel Decker Inc., 2001, pp. 1-15.

Mac, Ryan. “Facebook Engagement Ranking Boost and MSI, Explained.” *The New York Times*, 5 Oct. 2021, www.nytimes.com/2021/10/05/technology/engagement-ranking-boost-msi-facebook.html.

Maidel, Veronica et al. “An ontology-content-based filtering method.” *International Journal “Information Theories & Applications”*, vol. 15, 2008, pp. 303-314.

Manovich, Lev. *AI Aesthetics.* Strelka Press, 2018.

Morris, Jeremy Wade. “Curation by Code: Infomediaries and the Data Mining of Taste.” *European Journal of Cultural Studies*, vol. 18, no. 4-5, 2015, pp. 446-463. *SAGE Journals*, doi: 10.1177/1367549415577387.

Muis, Iris et al. “Liberation Begins with Stating the Facts: Framing Statistics and Information Bricolage in Geert Wilders’ Twitter Practice.” *Open Library of Humanities*, vol. 5, no. 1, 2019, pp. 1-31. *Open Library of Humanities*, doi: 10.16995/olh.320.

Nanopoulos, Alexandros et al. “How does high dimensionality affect Collaborative Filtering?” *Proceedings of the 2009 ACM Conference on Recommender Systems*, 2009, pp. 293-296. *ACM Digital Library*, doi: 10.1145/1639714.1639771.

Nguyen, Tien T., et al. “Exploring the Filter Bubble: The Effect of Using Recommender Systems on Content Diversity.” *Proceedings of the 23rd International Conference on World Wide Web - WWW ’14*, 2014, pp. 677-86. *ACM Digital Library*, doi: 10.1145/2566486.2568012.

Pariser, Eli. *The Filter Bubble: What the Internet is Hiding from You*. Penguin New York, 2011.

Ricci, Francesco et al. “Introduction to Recommender Systems Handbook.” *Recommender Systems Handbook*, Springer, 2011, pp. 1-29.

Rich, Elaine. “User modeling via stereotypes.” *Cognitive Science*, vol. 3, no. 4, 1979, pp. 329-354. *Wiley Online Library*, doi: 10.1207/s15516709cog0304_3.

Schrage, Michael. *Recommendation Engines*. The MIT Press, 2020.

Schwarz, Ori. “Facebook Rules: Structures of Governance in Digital Capitalism and the Control of Generalized Social Capital.” *Theory, Culture & Society*, 2019, vol. 36, no. 4, pp. 117-141. *SAGE Journals*, doi: 10.1177/0263276419826249.

Sheikh, Yaser et al. “On the use of computable features for film classification.” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 1, 2005, pp. 52-164. *IEEE Xplore*, doi: 10.1109/TCSVT.2004.839993(410).

Smith, Brent et al. “Two Decades of Recommender Systems at Amazon.com”. *IEEE Internet Computing*, vol. 21, no. 3, 2017, pp. 12-18. *IEEE Xplore*, doi: 10.1109/MIC.2017.72.

Van Dijck, José. “Disassembling Platforms, Reassembling Society.” *The Culture of Connectivity: A Critical History of Social Media.* Oxford University Press, 2013, pp. 24-45.

Van Dijck, José. “Facebook and the Imperative of Sharing.” *The Culture of Connectivity: A Critical History of Social Media.* Oxford University Press, 2013, pp. 45-68.

Van der Vlist, Fernando et al. “How Partners Mediate Platform Power: Mapping Business and Data Partnerships in the Social Media Ecosystem.” *Big Data & Society*, vol. 8, no. 1, 2021, pp. 1-16. *SAGE Journals*, doi: 10.1177/20539517211025061.