Analyzing Spotify's network of playlists: how tracks spread through playlists on Spotify

by:

Fenne Schoot (ANR: 1259382)

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Science in Marketing Management

Supervisor: Dr. H. Datta
Co-reader: M. J. Pachali

November 27, 2019

Management Summary

Over two billion playlists are available on Spotify, and different playlists fulfill different roles within the Spotify playlist ecosystem. The impact of these playlists on the discovery and success of new tracks and artists makes it crucial to understand how exactly these playlists interact. In this research, we therefore investigated the influence of playlists in three different categories (official Spotify playlists, major label playlists, and independent playlists) on other playlists within and across these categories. We defined the influence of one playlist (which we referred to as the *source* playlist) on another playlist (which we referred to as the *target* playlist) as the importance of a source playlist in spreading tracks to a target playlist. We observed how 49,945 newly released tracks were subsequently added to different playlists, and used a data-driven approach developed in earlier research to quantify the influence between playlists based on these observations. We then conducted a two-way ANCOVA on a sample of over 48 million source-target playlist combinations to examine the differences between playlist categories. We found a significant interaction effect between the source playlist category and the target playlist category on the influence the source playlist has on the target playlist. We concluded that while official Spotify playlists might be influenced most by other official Spotify playlists, taken all results together, the major label playlists are the ones that play the most important role within the Spotify playlist ecosystem. We therefore believe that artists and their managers should greatly consider distributing and promoting new music in corporation with a major label.

Table of Contents

1. Introduction

Currently, music streaming is more popular than it has ever been.[1] At the end of last year, revenues from streaming music platforms accounted for 75% of all recording revenues in the U.S.[2] With over 207 million active users and 96 million paying subscribers, Spotify is the biggest of these music streaming platforms.[3] The rise of music streaming platforms has dramatically changed how consumers listen to and discover new music, and with that, it has changed how new music is most successfully distributed and promoted (Ren & Kaufmann, 2017). Before the digital age, distribution and promotion revolved around getting music on the radio and in record stores; now, any successful distribution and promotion strategy revolves around *playlists*.[4]

Aguiar and Waldfogel (2018) have already shown that Spotify's most-followed playlists have the power to increase the success of new tracks and artists, and even more importantly, that they determine which new tracks and artists are discovered by consumers in the first place. This shows that it is crucial for new music to be added to these playlists, because this is how Spotify users often become aware of new tracks and artists (Aguiar & Waldfogel, 2018). Next to these few most-followed playlists created by the platform itself, playlists of different owners should also be of interest for artists and labels. For example, Spotify users can also create their own playlists. While it is nearly impossible for new artists to be added to official Spotify playlists, their chances of being added to smaller user-created playlists are higher (Goodrich, 2019). These playlists can then serve as a launch pad to other, more popular playlists; when a track is added to a few smaller playlists, Spotify's algorithms and curators pick up on this and can add the track to larger, official Spotify playlists (Goodrich, 2019). As a consequence, more users will discover the track, which might lead to more additions to user-created playlists, which Spotify's algorithms and curators might again pick up, et cetera. Furthermore, addition to these user-generated playlists can be an end goal in itself for new tracks and artists. Even though they typically have only a few followers, listening through these playlists accounts for 36% of all the streaming hours on Spotify, which is 2.5 times as much as all playlists created by Spotify together.[5] Since Spotify's pay-

[1] https://mn2s.com/news/label-services/what-is-streaming-music-changed-industry/
[2] http://www.riaa.com/wp-content/uploads/2019/02/RIAA-2018-Year-End-Music-Industry-Revenue-Report.pdf
[3] https://soundcharts.com/spotify-analytics
[4] https://medium.com/soundcharts/the-mechanics-of-the-recording-industry-6e9cbe16e7db; https://mn2s.com/news/label-services/rising-importance-of-playlists-in-music-distribution/
[5] https://soundcharts.com/spotify-analytics

out to artists is based on their numbers of streams, the user-generated playlists can increase artists' revenues just like the official Spotify playlists can ("Spotify 101", 2019).

These examples show that different playlists fulfill different roles within the Spotify playlist ecosystem, and that by being added to playlists tracks can spread further to other playlists. The impact of playlists on the discovery and success of new tracks and artists makes it crucial to understand how exactly these playlists interact. In this research, we will therefore investigate the role of different playlists and their owners in spreading tracks to other playlists.

Studying this is important for two reasons. First, gaining knowledge of the identity of influential playlists on Spotify can be extremely useful from the perspectives from viral marketing. Artists and their managers can use the results of this study to determine their promotion and distribution strategies when releasing new music, as well as weigh their decision to execute these strategies independently or sign a distribution deal with a major label or other company. Second, our results can be useful for policy makers in the music industry. The rise of online platforms has raised concerns about their power to influence market outcomes, and about how they might exercise this power in a biased way (Waldfogel, 2017). Spotify itself might not be a music producer, but the three major record labels (Sony Music Entertainment, Universal Music Group, and Warner Music Group) have substantial ownership stakes in Spotify (Aguiar & Waldfogel, 2018). Our findings will give insight into the roles different parties play on the platform, and more specifically, into the power major labels and Spotify have in choosing which tracks are spread further to other playlists.

Our research can be related to three streams of literature. First, it relates to how consumers are influenced by playlists and similar mechanisms in markets with large amounts of products. Second, it relates to the field of information diffusion in online social networks, as well as to the more specific concept of viral marketing within this field of research. Third, it relates to research on the changing role of major labels due to the impact of digital technologies.

Despite research showing that playlists and similar features can greatly influence which new tracks and artists become successful (Aguiar & Waldfogel, 2018; Sorensen, 2007; Sorensen, 2017), no extant literature has empirically investigated how playlists might influence each other. Furthermore, while there is a vast amount of literature on information diffusion and viral marketing in the context of online social networks like Facebook (for example, Bakshy, Hofman, Mason, and Watts (2011), Twitter (for example, Cha, Haddadi, Benevenuto, and Gummadi, 2010; Romero, Meeder, and Kleinberg, 2011), and Flickr (for example, Cha et al., 2009), much less is known about these concepts in the context of music streaming. We bring these two literature streams together by using concepts and methods from the field of information diffusion and viral marketing to investigate the influence playlists have on other playlists. In addition, because previous studies have reached inconsistent conclusions about the current influence of major labels on the music industry, our research sheds further light on this topic by examining the influence major labels' playlists have on Spotify.

To investigate the role of different playlists and their owners in spreading tracks to other playlists, we will use data concerning historic information about playlists and tracks on Spotify, which we obtained from the website Chartmetric.com. We will measure the influence between playlists by using the data-driven approach that was described by Saito, Kimura, Ohara, and Motoda (2010), and we will then conduct a two-way ANCOVA to determine the differences between playlists of different owners. In the next sections, we will first give more theoretical background on product discovery in markets with vast amounts of products, on information diffusion on social networks and viral marketing, and on the changing role of major labels due to digitization of the music industry. After that, we will discuss our conceptual framework and our expectations. We will then explain our data collection and preparation methods as well as the methods used for our analysis in more detail. We will report the results, and finally, we will discuss the implications and limitations of our study.

2. Literature Review

2.1 The influence of (play)lists on market outcomes and product discovery

The influence of Spotify playlists has been investigated by Aguiar and Waldfogel (2018). They found that being added to Today's Top Hits, the most-followed playlist on Spotify, generated almost 20 million additional streams for a song, and that inclusion on New Music Friday playlists substantially raised the probability of song success, including for new artists. Based on these results, they concluded that Spotify playlists influence consumers' listening choices, and as a result, have the power to raise songs' streams and stimulate song and artist discovery.

One gap in this research, however, is that they only looked at a few playlists, which were all among the most-followed playlists on Spotify and all curated by Spotify itself. The found effects might be different or less pronounced for playlists that have less followers or are curated by other companies or users. Another gap is that while Aguiar and Waldfogel (2018) emphasize the importance of being added to these playlists, they do not address the consequently arising question of *how to get on* these playlists. A third gap is that they only assessed the influence of these playlists on market-level outcomes. Other research has found that listening through user-created playlists accounts for 36% of all the streaming hours on Spotify, which is 2.5 times as much as all curated playlists together.[6] This implies that listening through playlists in the user-created longtail might be just as important as the streams generated by the playlists investigated by Aguiar and Waldfogel (2018), and that it is also important to investigate whether the addition of tracks to these playlists leads to subsequent additions to user-created playlists.

The current research fills these gaps by sampling a large number and wide range of playlists on Spotify, of different owners and popularity. We will investigate how certain playlists influence other playlists to add the same tracks, or in other words, how playlists spread tracks to other playlists. This addresses both the question of how to get on the playlists that were investigated by Aguiar and Waldfogel (2018), as well as how these playlists might spread tracks to other (user-created) playlists.

The research by Aguiar and Waldfogel (2018) has so far been the only one on the topic of playlists. But since a lot of playlists on Spotify are based on certain levels of popularity or numbers of streams (like the algorithmic "Global Top 100" or the Spotify-curated "Today's Top Hits"), these type of playlists can be compared to bestseller lists and popularity rankings. Research on this topic has been done by, among others, Sorensen (2007), who studied the impact of the New York Times bestseller list on sales of hardcover fiction titles, Cai, Chen, and Fang (2009), who studied the effect of presenting a restaurant's menu in the form of a ranking, and Tucker and Zhang (2011), who studied the effect of presenting a website serving as a directory of wedding services in bestseller list style. These studies all found that appearing on a bestseller list or ranking leads to an increase in sales and facilitates product discovery.

Like the research by Aguiar and Waldfogel (2018), however, these studies only investigate the effect of these lists on market-level outcomes, and not the effect that different lists might have on each other. We can find a first suggestion on this topic Sorensen (2017), who argued that these lists can have a self-reinforcing effect through a loop of positive feedback; if appearing on lists increases sales and facilitates product discovery, products are

[6] https://soundcharts.com/spotify-analytics

more likely to appear on lists, which again increases sales, and so forth. Newberry (2016) showed that such a self-reinforcing effect can indeed occur in the context of an online music market, but used increased prices as a cue of previous track popularity for this, and did not address this mechanism in the context of playlists.

These studies on mechanisms comparable to playlists have shown that playlists could possibly positively affect the number of streams for a track and lead to a higher chance of track discovery – which Aguiar and Waldfogel (2018) have shown to be indeed true for Spotify playlists – and that this might lead to subsequent addition to other playlists through a self-reinforcing effect. However, whether and how a track being added to one playlist indeed helps in spreading that track to other playlists, or in other words, the influence of playlists on other playlists, has not yet been empirically investigated in extant literature.

2.2 Information diffusion and viral marketing

The concept of looking at how tracks spread through playlists as a consequence of the influence playlists have on each other is related to the field of information diffusion in online social networks. More specifically within this field, the idea of finding which playlists are most effective in spreading tracks to other playlists relates to the concept of *viral marketing*, of which the goal is to target just a few influential nodes in a network and rely on them to spread the information further throughout the network (Cha et al., 2010). While there is a vast amount of literature on information diffusion and viral marketing on social networks like Flickr (for example, Cha et al., 2009), Facebook (for example, Bakshy et al., 2011), and Twitter (for example, Cha et al., 2010; Romero et al., 2011), much less is known about these concepts in the context of music streaming.

A start in this area has been made by Garg, Smith, and Telang (2011) and Ren, Cheng, Shen, and Zhu (2014), who both analyzed information diffusion on Last.fm, a music streaming platform comparable to Spotify (Garg et al., 2011). Garg et al. (2011) found that users had a positive effect on the diffusion of music to other users. Users were 6 times more likely to discover a new song and 3 times more likely to discover a new band as a result of influence by the listening behavior of other users. They concluded that the network of users had a significant power to diffuse new songs and lead to more content discovery by users. While this research just showed that users in the network were influenced by the listening behavior other users, Ren et al. (2014) took this one step further to investigate which users were the most influential. This was the first study related to finding the most influential users in the context of music streaming from an information diffusion perspective. The results demonstrated that targeting the most influential users in a music social network can lead to a more widespread diffusion of music.

The studies by Garg et al. (2011) and Ren et al. (2014) showed that it is possible to model music streaming markets as social networks through which information diffuses. However, these studies both focus on how music spreads through *users*, as a result of users directly observing the listening behavior of other users and in that way being influenced to listen to the same music. No extant research has used the concept of information diffusion in social networks to investigate how tracks spread through playlists. In addition, the idea of using viral marketing techniques for targeting playlists has not yet been investigated.

2.3 The changing role of major labels in the music industry

Bhattacharjee et al. (2007) assessed the impact of the rise of digital technologies on the music market, using data on weekly rankings of albums on the Billboard top 100 charts and album-level sharing activity captured on WinMX, and online file sharing service. They found that while major label albums still survived longer on the Billboard top 100 charts than indie label albums, indie label albums did survive longer on the charts compared to earlier periods. From this, Bhattacharjee et al. (2007) concluded that indie labels were closing the gap between them and the major labels.

More recent studies by Ren and Kaufmann (2017) and Im, Song, and Jung (2018) showed that, on the two music streaming markets they investigated, this gap has decreased even further. Ren and Kauffman (2017) analyzed more than 78,000 top chart ranking observations from music streaming service Last.fm, and Im et al. (2018) analyzed the weekly top 100 charts of the digital music market in South Korea. Both of these studies found that albums of major labels did not survive longer on the streaming charts than albums of indie labels.

Although these studies suggest a decreased disadvantage of indie labels, they only investigated the role of major versus indie labels *after* albums reached top-chart rankings, and neglected a possible disadvantage of indie labels in actually *reaching* these top charts. When Aguiar and Waldfogel (2018) analyzed a sample made up of global daily top 200 songs on Spotify (the music streaming market that is also the focus in the current research), they found that only a quarter of these tracks and just a fifth of the streams were indie label songs. In addition, they concluded that Spotify's most-followed playlists – of which we have already discussed their impact on market outcomes – still tend to promote major-label music. Furthermore, they found that playlists owned by major labels accounted for 6.7 percent of the cumulative followers of the top 1,000 most followed playlists. While this may not seem much at first sight, all the remaining playlist owners only have negligible shares, except for Spotify itself whose playlists together account for over three quarters of the top 1,000 playlists' cumulative followers (Aguiar & Waldfogel, 2018). So while the studies by Ren and Kaufmann (2017) and Im et al. (2018) showed that major label albums do not have an advantage over indie label albums anymore *once they have reached the top charts*, Aguiar and Waldfogel (2018) showed that the major labels still dominate the most popular segment on Spotify.

Datta, Knox, and Bronnenberg. (2018) focused their investigation on the individual rather than on the market level. They demonstrated that on the individual level, the adoption of Spotify led to a significant long-run shift in music consumption toward more plays, variety, and new music discovery. Based on these results, they argued that the shift from ownership to streaming could potentially level the playing field to the benefit of indie labels. However, they did not actually investigate this, and the more recent study by Aguiar and Waldfogel (2018) suggests that the gap between major and indie labels still exists.

These inconsistencies show that it is not yet clear to what extent the major labels continue to dominate or exert influence on music streaming markets, and more specifically, on Spotify. In addition, while Aguiar and Waldfogel (2018) addressed the major labels' relatively high percentage of the top 1000 playlists' cumulative number of followers, they did not investigate whether and how addition to these playlists might benefit tracks. The current study attempts to gain more knowledge on these matters by examining the influence of playlists created by major labels on other playlists on Spotify.

3. Conceptual framework

In this research, we are interested in the influence playlists have on other playlists, i.e., how the addition of a track to one playlist might lead to the addition of that track to another playlist. To investigate this, we turn to the field of social network analysis. Literature on this topic typically represents a social network as a graph, in which the nodes are the members of the network and the edges are the relationships between them. In a process called information diffusion, information spreads through nodes, along the edges between them (Guille, Hacid, Favre, and Zighed, 2013). Such a spreading process between nodes arises because of observational learning, which means that members of the network observe the behavior of other members, and can then be influenced by them to behave similarly (Garg et al., 2011). This effect can lead to a successive activation of nodes throughout the network.

Previous studies by Garg et al. (2011) and Ren et al. (2014) have already shown that music streaming markets can be represented this way as a music diffusion graph, in which the nodes are the users, and tracks diffuse through the graph because users start listening to tracks by observing the listening behavior of other users. However, research on online (music) markets has shown that the behavior of users in these markets is not only influenced by direct observations of the listening behavior of other users, but also by features provided by the platform to facilitate music or product discovery. For example, the study by Aguiar and Waldfogel (2018) has shown that playlists on Spotify have the power to influences users' listening choices and music discovery. Studies on bestseller lists – which can be likened to certain types of playlists – provide further implications that appearing on these lists influences consumers' choices (Sorensen, 2007; Cai et al., 2009; Tucker and Zhang, 2011). After having discovered a new track through a playlist, users can then add these tracks to their own playlists, which might cause other users to then also discover the track and add it to their playlist, etcetera. This way, playlists can also be influenced by playlists. This is not only the case for user-created playlists, but also for playlists created by the platform itself; Sorensen (2017) and Newberry (2016) implied that through a self-reinforcing affect, addition to lists generated by the platform can influence subsequent addition to other lists. For these reasons playlists can be influenced by playlists, and the diffusion of tracks might also be modeled as a network in which the nodes are playlists instead of users.

In this study, we thus propose to view Spotify as a network in which the nodes are playlists, and tracks spread through these playlist nodes. Figure 1 presents the conceptual framework for our research. We discuss this framework in more detail in the following subsections.

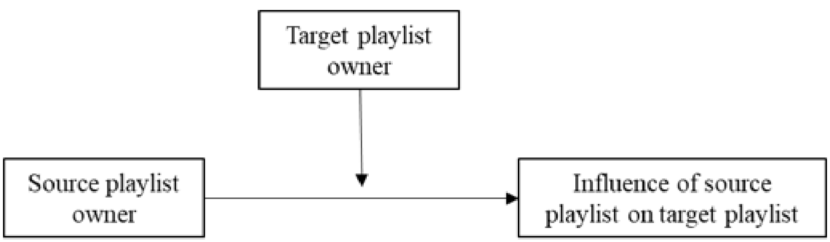

Figure 1. Conceptual framework

3.1 Influence of source playlist on target playlist

The core of our conceptual framework consists of the influence of a source playlist on a target playlist. In the field of information diffusion on social network analysis, influence has been defined in several ways. Guille et al. (2013) defined social influence as a phenomenon that individuals can undergo or exert that can also be called *imitation*, and that translates into the fact that actions of a user can induce his connections to behave in a similar way. In the context of Twitter, Bakshy et al. (2011) defined the influence of users as their ability to get their followers to further spread a particular piece of information. In the context of music social networks, Ren et al. (2014) argued that a user influences his friends, if his friends start listening to a track because they observe that the user listened to that track. However, these definitions all imply knowledge on the underlying connections between users (e.g., who follows whom, or who is friends with whom). In contrast, Saito et al. (2010) describe influence as the importance of a node in spreading information to other nodes. This definition is different in that it does not imply that the nodes are directly connected to each other.

In this research, our data does not include information on how different (owners of) playlists might be connected or observe each other's behavior; we merely have, for each track in our data, all the playlists the track is added to and the dates on which these additions occurred. We therefore base our definition on the one given by Saito et al. (2010), and define the influence of one playlist on another playlist as the importance of a playlist in spreading tracks to another playlist. Furthermore, to facilitate the discussion on the influence between playlist pairs, we refer to the playlist that spreads the track as the *source* playlist, and refer to the playlist that the track is spread to as the *target* playlist.

We propose that a source playlist s plays an important role in spreading tracks to a target playlist t (which is how we defined the influence of one playlist on another playlist) if two conditions are both fulfilled: (1) of all the tracks added to source playlist s, a large share is later also added to target playlist t, and (2) of all the tracks added to target playlist t, a large share was first added to source playlist s. We will illustrate why we propose both of these conditions to be necessary in order to quantify influence with two fictional scenarios.

First consider the following scenario, which we depict in graphical form in Figure 2 (Scenario 1). Over a certain time span, 50 tracks were added to playlist A. 35 of the tracks that were added to playlist A were later also added to playlist B, and the same amount of 35 tracks added to playlist A was later added to playlist C. We can then say that of all the tracks that were added to playlist A, a share of 70% was later added to playlist B, and an equally

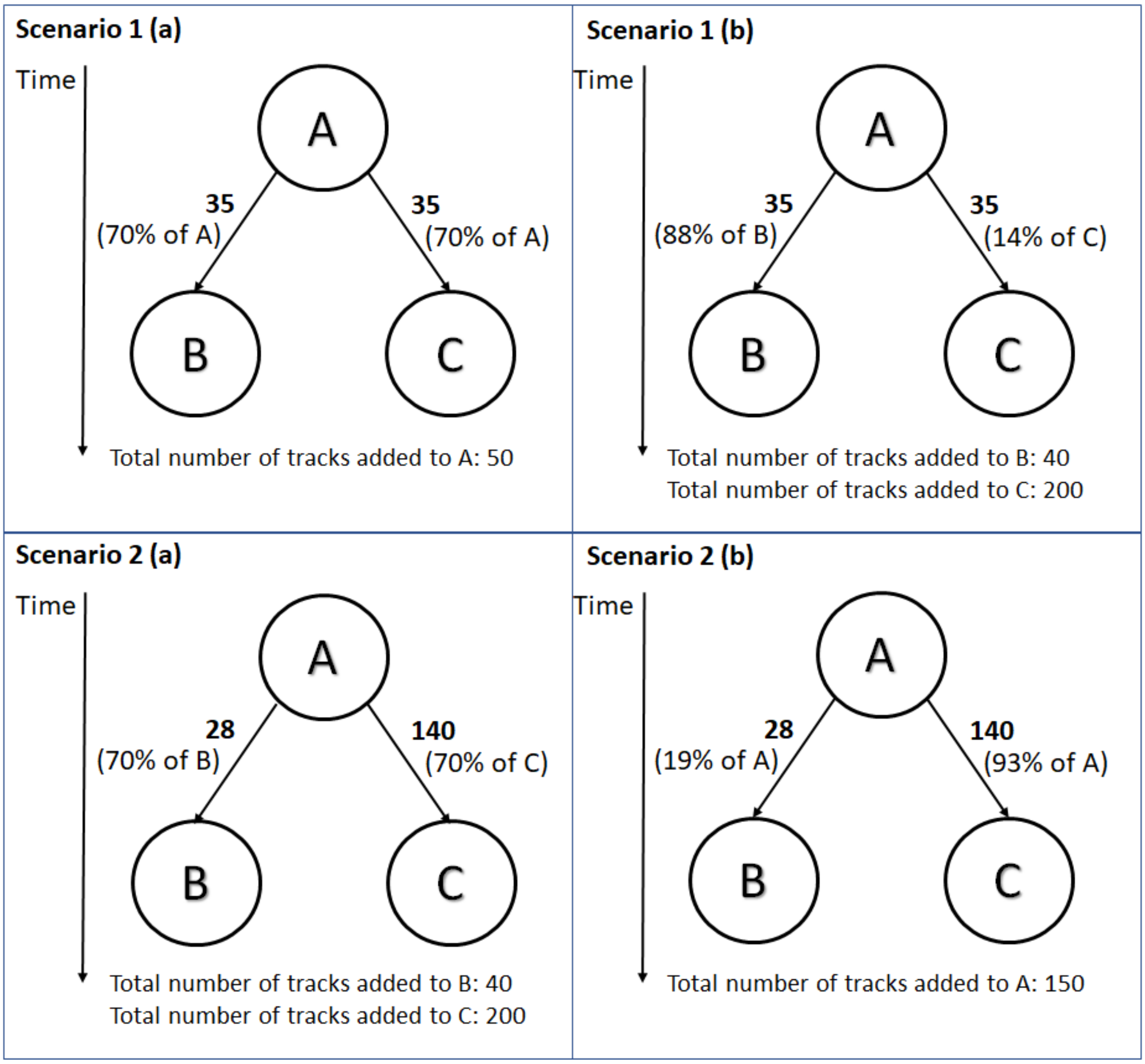

Figure 2. Two scenarios of playlist influence. A, B, and C are three different playlists. Scenario 1: over a certain time span, 50 tracks were added to A, 40 tracks were added to B, and 200 tracks were added to C. 35 tracks that were added to A were later added to B, and 35 tracks that were added to A were later added to B. Part (a) shows that we can thus say that 70% of the tracks added to A was later added to B, and that a similar percentage of 70% of the tracks added to A was later added to C. Part (b) shows that we can also say that 88% of the tracks added to B was first added to A, and that 14% of the tracks added to C was first added to A. Scenario 2: over a certain time span, 150 tracks were added to A, 40 tracks were added to B, and 200 tracks were added to C. 28 tracks that were added to A were later added to B, and 140 tracks that were added to A were later added to C. Part (a) shows that we can thus say that 70% of the tracks added to B was first added to A, and that a similar percentage of 70% of the tracks added to C was first added to A. Part (b) shows that we can also say that 19% of the tracks added to A was later added to B, and 93% of the tracks added to A was later added to C. These two scenarios illustrate that both the point of view in (a) and in (b) are needed to properly assess to what extent A influences B and to what extent A influences C.

high share of 70% was later added to playlist C (see Scenario 1 (a), Figure 2). Based on only condition (1), it would seem as if the influence of playlist A on playlist B is equal to the influence of playlist A on playlist C. However, we now reveal that over the whole time span, a total of only 40 tracks was added to playlist B and a total of 200 tracks was added to playlist

C. We can then also say that 88% of the tracks that were added to playlist B were first added to playlist A, and only 14% of all the tracks that were added to playlist C were first added to playlist A (see Scenario 1 (b), Figure 2). This additional information based on condition (2) reveals that the influence on playlist A on playlist B is greater than the influence of playlist A on playlist C, and clearly not exactly the same.

Scenario 2 in Figure 2 is based on the same reasoning, except in this case, condition (2) is the one that implies that the influence of playlist A on playlist B is equal to the influence of playlist A on playlist C (see Scenario 2 (a), Figure 2), and condition (1) reveals that there actually is a clear difference in the amount of influence (see Scenario 2 (b), Figure 2). These two scenarios show that a combination of both conditions is needed to be able to properly assess the amount of influence a playlist has on another playlist, since looking at only one of these conditions might lead to misleading results.

It is important to note that our use of the terms source playlist and target playlist is only to facilitate the discussion on who we mean to influence whom. One playlist can be a source as well as a target playlist; for example, in Figure 2, playlist A is the source playlist and playlist B is the target playlist, but if we would add more time points, playlist B might become the source playlist for another target playlist C. In addition, this figure shows that one playlist can be a source playlist for multiple target playlists, as playlist A is. Furthermore, the influence of the source playlist on the target playlist is *not* the same as the influence of the target playlist on the source playlist; we could also create a third scenario in which we determine the influence of source playlist B on target playlist A by looking at the number of tracks that were added to playlist B that were later added to playlist A, instead of the other way around.

3.2 Playlist owners

Tracks are added to playlists by their owners. To understand how playlist are influenced by other playlists thus actually means we need to understand how playlist owners decide to add tracks to their lists, and how these decisions might be influenced by other playlists and those playlists' owners. To that end, we divide playlists into three categories: official Spotify playlists (all playlists created and updated by Spotify itself), major label playlists (all playlists created and updated by major label companies), and independent playlists (all playlists created and updated by any non-Spotify and non-major label entity) (Voogt, van Doorne, and Erickson, 2018). In the following subsections we will further define and explain each category, and along with that we will form our expectations on how playlists influence each other within and across the three categories.

3.2.1 Official Spotify playlists

The first category we distinguish consists of playlists that are owned by Spotify itself. We will refer to these playlists as "official Spotify playlists". There are two main types of official Spotify playlists: *algorithmic* playlists, which are updated automatically based on algorithms, and *editor curated* playlists, which are updated manually by Spotify employees (also called Spotify's "editors' or "curators") (Aguiar & Waldfogel, 2018).[7]

Both the algorithms and the editors decide which tracks to add to playlists based on *data*.[8] Spotify editors use a strategy they call "playlisting".[9] They first test out new tracks on playlists with only small amounts of followers. Based on listening data – for example, how often the track is skipped, streamed, or added to users' own playlists – curators then assess how well a track is doing on this first list. If a track proves to be successful on such a "feeder" playlist, curators gradually give it more exposure by adding to more and more popular playlists. This way, if a track keep doing well, it can move up the playlist ladder until it reaches Spotify's flagship playlists with millions of followers.[9] This provides a first implication that tracks to a great extent spread through official Spotify playlists further to other official Spotify playlists. Furthermore, because the editors and algorithms partly use data on how often tracks are added to users' own playlists to decide which tracks to add, official Spotify playlists are also influenced by independent playlists to some extent.

Like the editor curated playlists, algorithmic playlists add tracks based on listening data – the difference is that with algorithmic playlists, the additions are done automatically based on algorithms instead of manually by curators. Some of these algorithmic playlists use algorithms based on users' listening habits, including what tracks they add to their playlists, what tracks they skip, and what tracks they share ("Spotify 101", 2019). Other algorithmic playlists, for example the Global Top 50, are solely based tracks' amounts of stream (Aguiar & Waldfogel, 2018).

---

[7] Note that in our analysis, we do not actually make a distinction between algorithmic and editor curated playlists. The purpose of mentioning the two different types here is solely to explain the mechanisms behind how tracks end up on official Spotify playlists.
[8] https://artists.spotify.com/faq/promotion#why-can't-i-submit-released-music-for-playlist-consideration
[9] For example, see Pierce (2017), Voogt et al. (2018), "Spotify 101" (2019), or https://artists.spotify.com/videos/the-game-plan/how-playlists-work.

Previous research on bestseller lists – which can be likened to some official Spotify playlists, in the sense that these can be seen as most-streamed lists – has provided evidence that appearing on these lists can lead to a self-reinforcing effect (Sorensen, 2017). What this means translated to our context is the following. Tracks are often added to official Spotify playlists based on how frequently they have been streamed. As a consequence of addition, these playlists can generate to up to almost 20 million *additional* streams for a track (Aguiar & Waldfogel, 2018). These additionally generated streams will then lead to further addition to other official Spotify playlists, which again generates even more additional streams, and thus even more additions to official Spotify playlists, et cetera. This self-reinforcing effect of official Spotify again implies that tracks spread through official Spotify playlists to other official Spotify playlists.

Considering the playlisting strategy used by Spotify, the studies on the self-reinforcing effect of bestseller lists, and the fact that it has been shown that the official Spotify playlists indeed generate a large amount of additional streams for tracks (Aguiar & Waldfogel, 2018), it seems that we can expect that addition to official Spotify playlists is of great importance for subsequent additions to other official Spotify playlists. On the other hand, both Spotify's algorithms and curators of official Spotify playlists partly base their decisions on the addition of tracks to independent playlists, and Goodrich (2019) has argued that independent playlists often serve as a launch pad for new artists to larger Spotify playlists. This means that we can also expect official Spotify playlists to be influenced by independent playlists to some extent. However, we expect this influence to be much less than that of other official Spotify playlists. In addition, because Spotify has stated that they do not treat artists differently based on their label or distributor, we expect official Spotify playlists to not be greatly influenced by major label playlists.[10] In view of all this together, we hypothesize the following:

H_1: Official Spotify playlists are influenced most by other official Spotify playlists, are influenced less by independent playlists, and are influenced even less by major label playlists.

3.2.2 Major label playlists

All three major labels (Sony Music Entertainment, Universal Music Group, and Warner Music Group) have their own playlisting brands. The major labels create and maintain their own playlists under these brand names, which are "Filtr" (Sony), "Digster" (Universal), and

[10] https://artists.spotify.com/faq/popular#does-uploading-my-music-through-spotify-for-artists-give-me-a-better-chance

"Topsify" (Warner) (Aguiar & Waldfogel, 2018). These playlist brands are among the largest sources of playlists on Spotify (Voogt et al., 2018). Of the top 1,000 most-followed playlists on Spotify, they together account for 7% of the cumulative followers (Aguiar & Waldfogel, 2018). While this may not seem much at first sight, all remaining list owners only have negligible shares, except for Spotify's editor curated lists which account for over three quarters of the cumulative followers (Aguiar & Waldfogel, 2018). After the official Spotify playlists, the major label playlists are the ones that are most visible on Spotify. Furthermore, they use a similar setup as the official Spotify playlists (Voogt et al., 2018).

In contrast to the large amount of available sources explaining Spotify's playlisting strategies and algorithms, little is known about how exactly the major labels incorporate their playlists in their promotion and distribution tactics. Of course, the reason that all three major labels have their own playlisting brands, is that it allows them another vehicle to push records on the platform without having to rely on Spotify curators. By growing their playlist follower base, they can grow the revenue brought in by their artists through these playlists (Voogt et al., 2018). Artists and their labels have access to listening data on their tracks, which enables the possibility that the major labels use a similar strategy as Spotify's "playlisting", in which they first add a track to one playlist to assess the response to this track, and use this to decide to add the track to other playlists. We expect major labels to indeed use such a strategy, which leads to the expectation that tracks spread through major label playlists to other major label playlists to a great extent, likewise to how this was the case within official Spotify playlists due to this strategy. Furthermore, we expect major label curators to use data on additions to other playlists to decide which tracks to add to their own playlists, likewise to Spotify's curators and algorithms. Because addition to an official Spotify playlist typically is a stronger signal of success than addition to an independent playlist, we expect them to be more influenced by additions to official Spotify playlists than by additions to independent playlists. We therefore hypothesize the following:

> H_2: Major label playlists are influenced most by other major label playlists, are influenced less by official Spotify playlists, and are influenced even less by independent playlists.

3.2.3 Independent playlists

The third category of playlists we distinguish comprises what we refer to as independent playlists, which are all playlists made by any non-Spotify or non-major label entity (Voogt et al., 2018). This is a very broad category, and also includes some of the biggest playlists on

Spotify created by well-known brands aimed at millions of followers, like Trap Nation and Dimitri Vegas & Like Mike (Voogt et al., 2018). However, the absolute majority of playlists on Spotify – and thus an even larger majority of playlists in this category, since the official Spotify and major label playlists are already filtered out – are those created by regular Spotify users (Goodrich, 2019). For the purpose of this research, we will therefore focus on these user-created playlists. While the purpose of official Spotify playlists and major label playlists is to promote tracks to a large audience, most users simply create playlists for their own personal enjoyment ("Spotify 101", 2019). Users typically use playlists to organize their music, and as utilities to play their favorite tracks (Voogt et al., 2018).

Since users naturally first need to be *aware* of a track before they can even consider adding it to their playlists, the playlists through which consumers discover new tracks and artists have a large amount of influence on which tracks users add to their playlists. Aguiar and Waldfogel (2018) provided evidence that the official Spotify playlists have the power to determine which tracks and artists users discover. Research on bestseller lists by Sorensen (2007), Cai et al. (2009), and Tucker and Zhang (2011) found additional evidence that lists created by the platform facilitate discovery by consumers. Furthermore, the official Spotify playlists are more visible on the platform than other playlists, and have the most followers (Voogt et al., 2018). If playlists are less visible and listened to less, then it is also likely that they have less influence on other playlists. Based on these observations, we expect independent playlists to be greatly influenced by the official Spotify playlists.

On the other hand, in spite of the observations about the role of official Spotify playlists in new music discovery, many of the streams generated by the official Spotify playlists are "drive-by" streams; users are often not very engaged when listening to these playlists, and are less likely to turn into active, long-term fans ("Spotify 101", 2019). In contrast, independent playlists might typically have less followers, but the followers that these playlists *do* have might be more engaged with the playlists. The reason for expecting this to be the case is that these playlists are not promoted by Spotify in any way and are barely visible, which makes it is likely that their followers discovered these playlists only because they are friends (or friends of friends) of the playlists' owners. Richardson and Domginos (2001) argued that consumers' choices are influenced more by the opinions of their peers and friends than by influentials. Similarly, Leskovec, Adamic, and Huberman (2007) argued that consumers are more interested in what friends buy than what an anonymous person buys, are more likely to trust their opinion, and are more influenced by their actions. In addition, Garg et al. (2011) showed that on Last.fm, users' track discovery was heavily influenced by the listening behavior of their connections, even when the ties between them were weak.

Based on these arguments, we expect that independent playlists can greatly influence other independent playlists. This stands in contrasts to our previous expectations on the influence of independent playlists on major label playlists and Spotify playlist, which we only expected to be small. We thus hypothesize the following:

H_3: Independent playlists have more influence on other independent playlists, than they have on both Spotify playlists and on major label playlists.

Despite this, we expect that the visibility of playlists on the platform is a stronger determinant of influence than the connections that might exist between users. After the official Spotify playlists, the major label playlists are the most visible and most-followed playlists (Voogt et al., 2018; Aguiar & Waldfogel, 2018). We therefore hypothesize the following:

H_4: Independent playlists are influenced most by official Spotify playlists, are influenced less by major label playlists, and are influenced even less by other independent playlists.

4. Data

4.1 Sample collection

To start our sample collection process, we used the website Everynoise.com to obtain a list of all albums that were released on Spotify in 2017. We took a stratified sample from these albums, by first dividing the albums into ten buckets based on the popularity of the artists that made the album. We measured artists' popularity by their position on a popularity ranking compiled by Everynoise.com. The first bucket contained albums by artists ranked in the top 10% of most popular artists, the second bucket contained albums by the next 10% most popular artists, etc. We then took an equally sized random sample from each bucket, so that we would obtain an equal distribution in our data of more and less popular artists.[11]

We then proceeded to use the website Chartmetric.com to collect additional information about these albums. For every track on every album, we sent a request to the application protocol interface (API) from Chartmetric.com to return all playlist positions the track previously had had and currently had at the time the data was retrieved, which was at the end of May, 2019. Because some albums were not tracked by Chartmetric.com, we ended up with

[11] As an exception, the sample from the bucket containing albums by the top 10% most popular artists was twice as big as the samples from other buckets. This is due to another study using the same raw data set, and does not influence our analysis.

data on 31,925 albums. After removing 4,064 duplicate rows (i.e., rows that were exact copies of other rows in the data set, due to the Chartmetric API returning the same information twice), our data set comprised 600,074 rows. Each row corresponded to a unique combination of a track and the addition of that track to a playlist. For each track, artist, album and playlist, we also retrieved additional meta data from Chartmetric's API (e.g., number of playlist followers, release date, and album label).

4.2 Data preparation and variable operationalization

The unit of analysis in our study is the addition of a track to a playlist. Hence, each row in our data set corresponded to a unique track-playlist combination. We will subsequently refer to the addition of a track to a playlist as a "playlist addition". Furthermore, each row in the data set contained additional information about the characteristics of the track and playlist. We will now proceed to describe how we constructed our final data set from the raw data. An overview of all variables and their operationalizations can be found in Table 1.

4.2.1 Tracks

We identified tracks by their unique Chartmetric ids, which we obtained using Chartmetric's API. We used ids rather than clear-text names because different tracks could have the same track name (for example, the track name "Home" occurred for 31 tracks in de data, but these were all different tracks by different artists).

We obtained albums' release dates using Charmetric's API. 13,235 tracks in the data came from albums that were *not* released in 2017. Since we only sent requests to the API for albums released in 2017, the API had returned these tracks incorrectly. We therefore removed them from the data set.

For each track, we collected the name of the label under which the track's album was released. We used this label name to create a binary variable, indicating whether a track was released by a major label or by an independent (indie) label. To that extent, we used a list consisting of the three major labels (Sony Music Entertainment, Warner Music Group, and Universal Music Group) and all their relevant subsidiaries/sublabels. All tracks released by labels on that list were coded as released by a major label, and all tracks released by labels *not* on this list were coded as released by an indie label.[12]

[12] After we categorized playlists' owners (as explained in section 4.2.2), we saw that 3,532 tracks that we coded as released by indie labels had been added to playlists owned by major labels. The first explanation for this

Table 1

Variable Operationalization

Variable	Operationalization
Tracks	
Track id	Unique identifier for each different track.
Track label	Whether a track was released and/or distributed by a major label, or released and distributed by an indie label.
Number of playlist additions	The total number of playlists that a particular track was added to.
Playlists	
Playlist id	Unique identifier for each different playlist.
Playlist category	Category to which a playlist belongs, based on its owner. We distinguish between official Spotify playlists (playlists owned by Spotify), major label playlists (playlists owned by a major label, i.e., by either Sony, Universal, or Warner), and independent playlists (playlists owned by any non-Spotify and non-major label entity).
Playlist addition date	The date that a particular track was added to a particular playlist.
Number of playlist additions	The total number of tracks that were added to a particular playlist.

4.2.2 Playlists

We identified playlists by their unique ids that we received from the Chartmetric API, likewise to how we identified unique tracks.

We categorized each playlist as being either an official Spotify playlist, a major label playlist, or an independent playlist (Voogt et al., 2018). Official Spotify playlists (playlists that were owned by Spotify itself) could be recognized by being tagged as "Official" in the data received from the Chartmetric API. Major label playlists (playlists that were owned by a major label) could be recognized by looking at playlists' owner names, which we also received from the Chartmetric API. The major labels Sony, Universal, and Warner operate their own playlists under the names of "Filtr", "Digster", and "Topsify", respectively (Aguiar & Waldfogel, 2018). We therefore classified playlists as major label playlist if their owner name was or contained one of these three words. All remaining playlists that did not fall in either of these two categories were classified as independent playlist (playlists that were owned by any non-Spotify or non-major label entity).

---

phenomenon is that the list of sublabels we used for classifying tracks as released by a major or an indie label was not up-to-date (for example, the previously independent label Spinnin' Records was bought by Warner Music in 2017 (van den Eerenbeemt, 2017), but was not on the list). Furthermore, some artists do not have an official record label deal with a major label company, but do have a *distribution* deal (Fowlkes, 2019). This means that while some tracks are not *released* by a major label, they *are distributed* by one. The appearance of "indie label" tracks on playlists owned by major labels thus implies that these tracks in fact do have a contract or agreement with a major label company. For these two reasons, we changed the classification of these 3,352 tracks appearing on playlists owned by major labels to "major label" tracks.

4.2.3 Playlist additions

To be able to investigate how tracks spread through playlists on Spotify, we need each track's complete set of playlist appearances *including the right chronological order.* We therefore obtained from Chartmetric's API for each track the date it was added to each particular playlist. For some tracks these dates suggested that the track was added to a playlist before the track was even released, which is evidently incorrect. We therefore removed the 4,400 tracks for which this was the case from the data.

Some tracks were added to the same playlist multiple times. If this was the case, we kept only the addition with the chronologically earliest date and removed later additions of the same track to the same playlist from the data. This is based on what is called the monotocity assumption in network analysis, which suggests that an activated node cannot be deactivated, and can thus only be activated once (Guille et al., 2013).

Finally, for each particular playlist, we counted the number of tracks that were added to the playlist; and for each particular track, we counted the number of playlists that the track was added to.

4.3 Descriptive statistics

We will now provide a general description of the final data set, of which a summary can be found in Tables 2 and 3 (Appendix).

4.3.1 Tracks

The final data set consisted of 306,191 observations of playlist additions of 49,945 different tracks. We observed all playlists tracks were added to from their release date until the time of data retrieval, which was at the end of May, 2019. On average, each track was added to 6 playlists (*SD* = 13.0), with a minimum of 1 and a maximum of 360 playlist additions for one track. Most tracks were only added to a few playlists (see Figure 3); only 14% of all tracks in the data were added to more than 10 playlists. Table 2 lists the five tracks in the data that were added to the most different playlists.

We classified tracks as released by either a major label or an indie label. 36% of the tracks in our data (*N* = 17,832) was released and/or distributed by a major label, and 64% (*N* = 32,113) was released and distributed by an indie label. Each major label track was on average added to 9 playlists (*SD* = 18.4), while each indie label track was on average added to only 4 playlists (*SD* = 8.0). Furthermore, the maximum number of playlist additions for one major

Table 2

Top 5 tracks that were added to the most playlists

	Track name	Artist name(s)	Number of playlists track has been added to
1	Biking	Frank Ocean, JAY Z, Tyler, The Creator	360
2	Party on the West Coast	Matoma, Snoop Dogg, Faith Evans, The Notorious B.I.G.	347
3	Spirit	J Hus	236
4	First F**k	6LACK, Jhené Aiko	235
5	Pa Ti	Bad Bunny	214

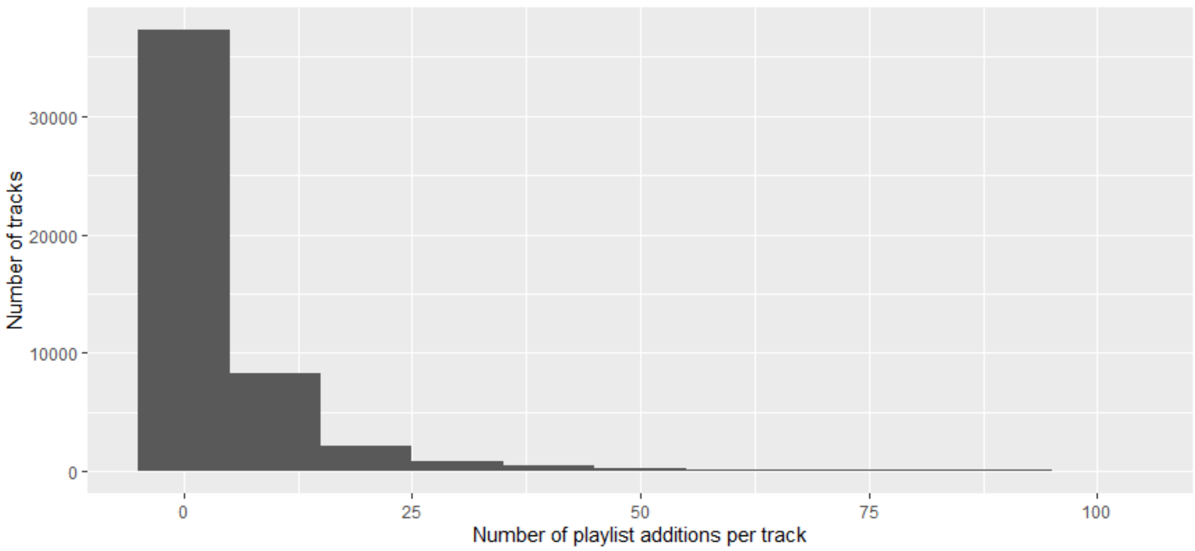

Figure 3. Distribution of number of playlist additions per track (up to 100 playlist additions per track)

label track was 360, while the maximum number of playlist additions for one indie label track was 200.

4.3.2 Playlists

Tracks were added to a total of 57,255 different playlists, with an average of 5 additions per playlist (*SD* = 12.4). The minimum number of additions to one playlist was 1, and the maximum number of additions to one playlist was 877, i.e., we observed 877 additions of tracks to this playlist.

Only 6% (*N* = 3,126) of the playlists in our data was owned by Spotify itself, and only 4% (*N* = 2,361) of the playlists was owned by the major labels; 90% (*N* = 51,768) of the playlists was owned by independents. Strikingly, while only a small fraction of the playlists in the data was owned by Spotify, these playlists together accounted for 71% of the cumulative number of followers of all playlists in the data.

Figure 4 shows that almost two thirds of the tracks added to official Spotify playlists were major label tracks. In other words, Spotify adds almost twice as many major label tracks as indie label tracks to its playlists. This already provides evidence that official Spotify playlists are biased towards adding major label tracks.

5. Methodology

In this research, we want to investigate the role of different playlists in spreading tracks to other playlists. In order to investigate this, we proposed to model Spotify as a network in which the nodes are playlists, and tracks can spread from playlist to playlist throughout the network. We defined the influence of one playlist in this network (which we refer to as the *source* playlist) on another playlist in the network (which we refer to as the *target* playlist) as the importance of the source playlist in spreading tracks to the target playlist. We based our definition on the research by Saito et al. (2010), in which they defined influential nodes as the nodes in a network that play an important role in spreading the information to many other nodes, and that they called "super-mediators".

Most research on influence in information diffusion is "model-driven", and requires information on the connections between nodes (Saito et al., 2016). However, our data does not include information on how different (owners of) playlists might be connected or observe each other's behavior. Saito et al. (2010) developed a "data-driven" approach to identify these super-mediators for which no information on connections between nodes is required, but that

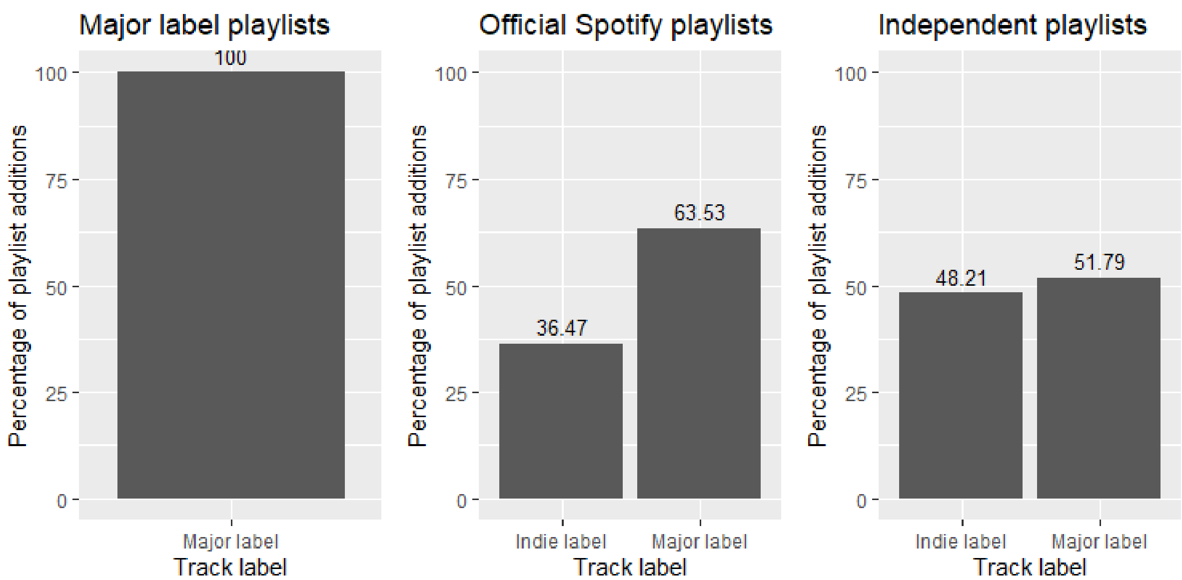

Figure 4. Percentage of playlist additions that were major label tracks or indie label tracks, of the total of all playlist additions in one playlist category.

just requires abundant observed information diffusion sequences – of which we have almost 50,000.

Because of our similar definition of influence as that in the research of Saito et al. (2010), and because of the data-driven nature of their approach for which we have more than sufficient data, we base our method of measuring influence between playlists on the method used by Saito et al. (2010). However, we change this method slightly, since Saito et al. (2010) were interested in finding which nodes were important in spreading information to as many other nodes as possible, while we are interested in finding which nodes are important in spreading information to specific other nodes.

5.1 Measuring influence

We defined the influence of a source playlist *s* on a target playlist *t* as the importance of playlist *s* in spreading tracks to playlist *t*. A source playlist that is important in spreading a track to a target playlist is a source playlist *s* that often occurs in diffusion sequences *before* the target playlist *t*, and that almost never occurs in diffusion sequences *after* playlist *t* or in diffusion sequences that do not include *t* at all (Saito et al., 2010). In other words, this means that if a source playlist *s* influences a target playlist *t*, we would observe in our data that (1) diffusion sequences including addition to *s* frequently also later include addition to *t*, and (2) diffusion sequences including addition to *t* frequently also add an earlier addition to *j*. We can relate this to the notions of *recall* and *precision*, two widely used measures in the field of information retrieval and binary classification (Saito et al., 2016). The observation in (1) resembles a high precision, and the observation in (2) resembles a high recall. We can calculate the precision $P(s,t)$ as

(1) $$P(s,t) = \frac{N(t|s)}{N(s)},$$

where $N(t|s)$ stands for the number of diffusion sequences that included addition to s before t, and $N(t)$ stands for the total number of diffusion sequences in the data that included addition to *t*. We can calculate the recall $R(s,t)$ of tracks added to *t* that were first added to *s* as

(2) $$R(s,t) = \frac{N(t|s)}{N(t)},$$

where $N(t|s)$ again stands for the number of diffusion sequences that included addition to *s* before *t*, and $N(t)$ stands for the total number of diffusion sequences that included addition to *t*.

In our conceptual framework, we explained why it is useful to look at both the share of tracks added to the source playlist that is later added to the target playlist as well as the share of tracks added to the target playlist that were first added to the source playlist (Figure 2) when determining the influence of the target playlist on the source playlist. Notably, the formula for precision in Eq. (1) is the same as how one would calculate the share of tracks added to a source playlist *s* that were later added to a target playlist *t*, and the formula for recall in Eq. (2) is the same as how one would calculate the share of tracks added to a target playlist *t* that was first added to a source playlist *s*. This resemblance provides further understanding of why this method is useful to measure influence.

We have now developed the intuition that the combination of the precision $P(s,t)$ and the recall $R(s,t)$ is indicative of the amount of influence that playlist *s* has on playlist *t*. We therefore calculate the influence $F(s,t)$ of playlist *s* on playlist *t* as the *harmonic mean* of these two measures:

(3) $$F(s,t) = 2 * \frac{P(s,t) * R(s,t)}{P(s,t) + R(s,t)}.$$

The reason for using the harmonic mean rather than using the arithmetic mean is the following. Consider a scenario (1) in which the precision has a value of 0.4 and the recall has a value of 0.6, and a scenario (2) in which the precision has a value of 0.01 and the recall has a value of 0.99. The arithmetic mean is equal for scenario (1) and (2), while the harmonic mean is 0.48 for scenario (1) and 0.098 for scenario (2). For this reason, we feel that the harmonic mean is more indicative of the influence between playlists.

5.2 Analyzing influence

For each possible playlist combination (*s*,*t*) of all playlists in the data, we calculated the influence $F(s,t)$ of *s* on *t* in the following way. First, we counted the number of diffusion sequences $N(s)$ that included *s*, the number of diffusion sequences $N(t)$ that included *t*, and the number of diffusion sequences $N(t|s)$ that included *s* before *t*. Next, we calculated the precision $P(s,t)$ of *s* for *t* using the formula in Eq. (1) and the recall $R(s,t)$ of *s* for *t* using the formula in Eq. (2). We then calculated the influence $F(s,t)$ of *s* on *t* as the harmonic mean of $P(s,t)$ and $R(s,t)$, using the formula in Eq. (3).

By calculating the influence of all possible combinations of playlists in the data, we were able to determine each playlist's influence on each of all the other playlists in the data. We then conducted a two-way ANCOVA to determine whether a statistically significant difference was present between the influences of different source playlist owners on different target playlist owners, while controlling for the number of tracks added to the source playlist and the number of tracks added to the target playlist. Due to computer memory limitations, this ANCOVA was conducted on a sample of 48.993.000 playlist combinations (7.000 distinct playlists). The two control variables (the number of tracks added to the source playlist and the number of tracks added to the target playlist) were added because they were used in calculating the influence of the source playlist on the target playlist.

6. Results

In this research, we were interested in how tracks spread through different playlists on Spotify. We defined the influence of a source playlist *s* on a target playlist *t* as the importance of the source playlist *s* in spreading tracks to target playlist *t*, which is high if (1) a large share of the tracks that are added to source playlist *s* will later added to target playlist *t*, and (2) a large share of the tracks that are added to target playlist *t* were first added to source playlist *s*. We conducted a two-way ANCOVA on a sample of over 48 million source-target playlist combinations to examine how this influence is affected by the source playlist owner and the target playlist owner, while controlling for the number of tracks added to the source playlist and the number of tracks added to the target playlist. Source playlist owner and target playlist owner both included three levels (official Spotify, major label, independent). The results of the two-way ANCOVA are presented in Table 3 (Appendix).

We found a significant and positive effect for both covariates, which were the number of tracks added to the source playlist ($F(2, 49{,}000{,}000) = 191{,}400{,}000$, $p < .001$), and the number of tracks added to the target playlist ($F(2, 49{,}000{,}000) = 4{,}378{,}000$, $p < .001$). This indicates that the influence of a source playlist on other playlists increases as the number of tracks added to the source playlist increases (see Figure 1 in the Appendix), and that the extent to which a target playlist is influenced by other playlists increases as the number of tracks added to the target playlist increases (see Figure 2 in the Appendix). This may be the case because a playlist owner that adds more tracks is more prone to discovering new tracks to add, and is therefore more influenced by other playlists. In turn, its influence on other playlists may then also increase because of a spillover effect in new music discovery (Garg et al., 2011).

The main effects for source playlist owner and target playlist owner were both statistically significant at the .001 significance level, but more importantly, the interaction effect between source playlist owner and target playlist owner was significant, $F(4, 49{,}000{,}000) = 44.94$, $p < .001$. To further analyze this interaction, we used Tukey's HSD post-hoc comparisons. For each playlist category separately, we will first discuss how tracks end up on these playlists (i.e., the extent to which they are influenced by playlists of the same and different owners), and subsequently, how these playlists then spread tracks to other playlists (i.e., the extent to which they exert influence on other playlists of the same and different owners). Tables 4 and 5 (Appendix) present the results of the post-hoc comparisons as well as the means and standard deviations of all source and target playlist owner combinations. Figure 5 and 6 present a graphical overview.

6.1 Official Spotify playlists

6.1.1 How do tracks spread to official Spotify playlists?

In H_1, we hypothesized that official Spotify playlists are influenced most by other official Spotify playlists, are influenced less by independent playlists, and are influenced even less by major label playlists. We found partial support for this hypothesis. The results of Tukey's HSD post-hoc comparisons revealed that with respect to Spotify target playlists, the differences between all three source playlist categories are indeed significant. However, Figure 5a shows that while official Spotify playlists are indeed influenced most by other official Spotify playlists, they are influenced more by major label playlists than by independent playlists.

6.1.2 How do tracks spread further from official Spotify playlists?

Figure 6a shows the influence of official Spotify source playlists on different target playlists. In this case, however, Tukey's HSD post-hoc comparisons revealed that none of these differences was significant. In spreading tracks to other playlists, the influence of official Spotify source playlists on other target playlists appeared to be unaffected by the owner of the target playlist; after a track is added to an official Spotify playlist, it will spread in a similar magnitude to other official Spotify playlists as to major label playlists and to independent playlists.

6.2 Major label playlists

6.2.1 How do tracks spread to major label playlists?

In H_2, we hypothesized that major label playlists are influenced most by other major label playlists, are influenced less by official Spotify playlists, and are influenced even less by

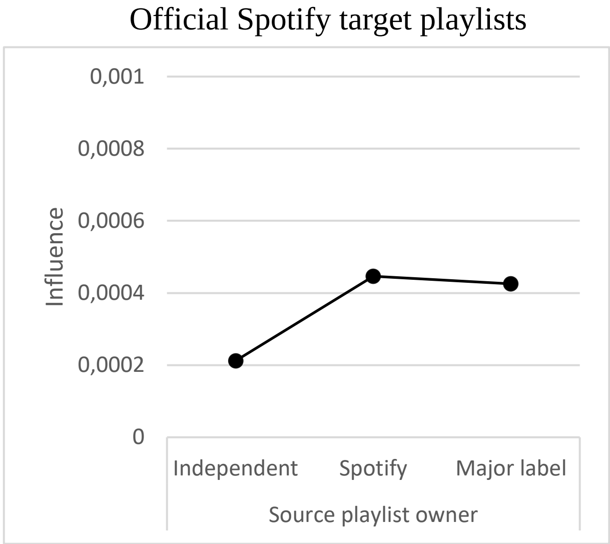

Figure 5a. Extent to which Spotify target playlists are influenced by different source playlists.

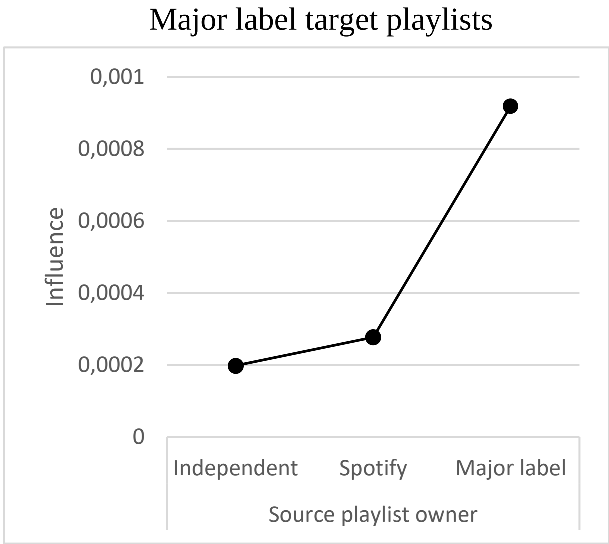

Figure 5b. Extent to which major label target playlists are influenced by different source playlists.

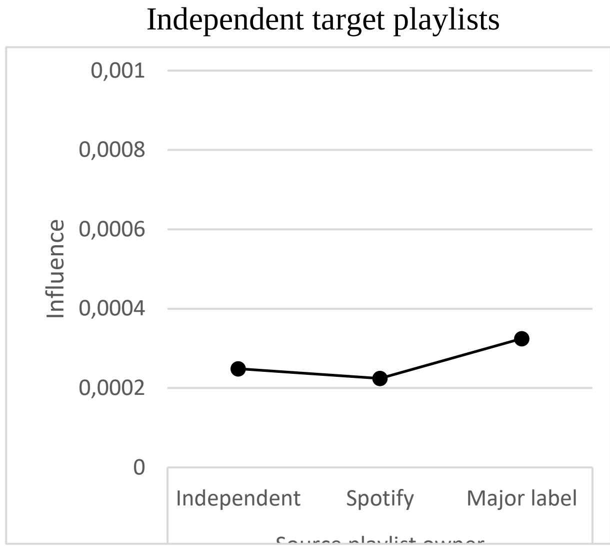

Figure 5c. Extent to which independent target playlists are influenced by different source playlists.

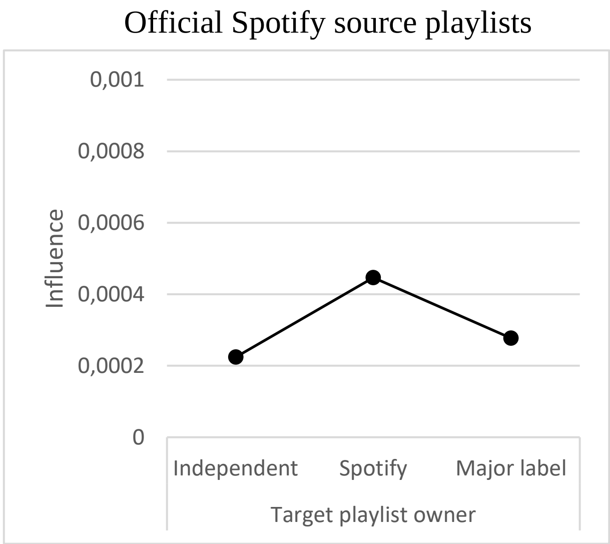

Figure 6a. Influence of official Spotify source playlists on different target playlists.

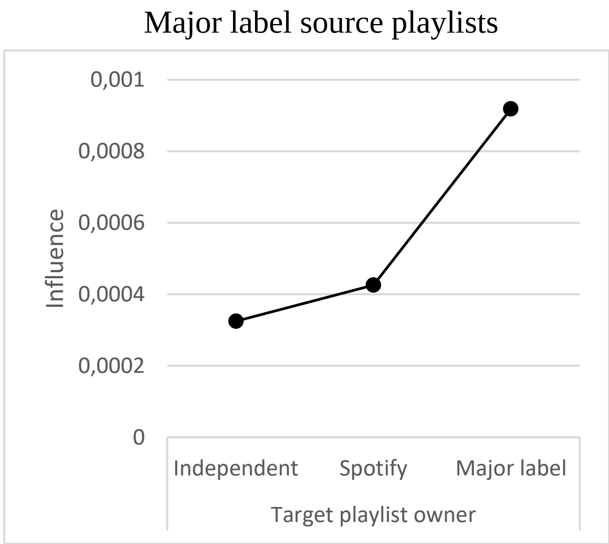

Figure 6b. Influence of major label source playlists on different target playlists.

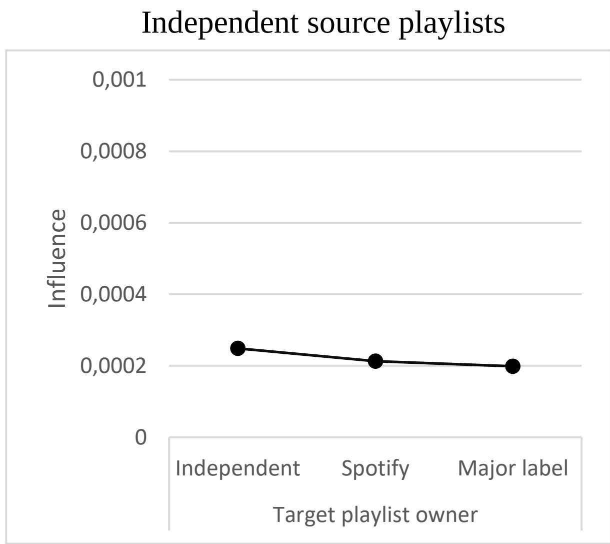

Figure 6c. Influence of independent source playlists on different target playlists.

independent playlists. Indeed, we found significant differences between all three source playlist categories, and Figure 5b shows that the order from most to least influential as stated in this hypothesis was indeed found.

6.2.2 How do tracks spread further from major label playlists?

Figure 6b shows the influence of major label source playlists on different target playlists. The influence of major label playlists on other major label playlists was significantly higher than on the other two target playlist categories. However, the difference between the influence of major label playlists on independent playlists and on official Spotify playlists was not statistically significant. Major label playlists are thus especially effective in spreading tracks to other major label playlists, and equally less effective in spreading tracks to independent playlists and to official Spotify playlists.

6.3 Independent playlists

6.3.1 How do tracks spread to independent playlists?

In H_4, we hypothesized that independent playlists are influenced most by official Spotify playlists, are influenced less by major label playlists, and are influenced even less by other independent playlists. In partial support of this hypothesis, we indeed found that independent playlists are influenced significantly less by independent playlists than by major label playlists (see Figure 5c). However, instead of being influenced the *most* by official Spotify playlists, independent playlists were found to be influenced significantly *less* by official Spotify playlists than by both other playlist categories. We expected that the visibility and amounts of followers of Spotify playlists would cause them to be more influential than the other playlists, but apparently, we underestimated the magnitude of the increased engagement and peer influence effects.

6.3.2 How do tracks spread further from independent playlists?

Figure 6c shows the influence of independent source playlists on different target playlists. In H_3, we hypothesized that independent playlists have more influence on other independent playlists than they have on both other playlist categories. Tukey's HSD post-hoc comparisons revealed that the influence of independent source playlists on other independent source playlists was indeed significantly higher than their influence on both other playlist categories. Moreover, we also found a significant difference between the influence of independent playlists on Spotify playlists and on major label playlists.

7. Discussion

In this research, we examined the influence of playlists on other playlists within and across three playlist categories (official Spotify playlists, major label playlists, and independent playlists). We defined the influence of a source playlist on a target playlist as the importance of the source playlist in spreading tracks to the target playlist. We argued that this influence is high if (1) a large share of the tracks that are added to the source playlist are later added to the target playlist, and (2) a large share of the tracks that were added to the target playlist were first added to the source playlist.

7.1 Who influences whom?

Our results showed that the influence source playlists have on other playlists depends on the owner of the of the source playlist as well as on the owner of the target playlist.

Figure 5 shows, for each target playlist category separately, the extent to which playlists in that category are influenced by each of the source playlist categories. For each of the target playlist categories, significant differences were found between all three source playlist categories, but the order from most to least influential differed across the target playlist categories. When the goal is to spread tracks to official Spotify playlists, other official Spotify playlists are most important. When the goal is to spread tracks to one of the other two playlist categories, however, major label playlists are most important. Independent playlists are least important, except when spreading tracks to other independent playlists; in that case, official Spotify playlists are even less influential than the independent playlists.

Figure 6 shows the results from the point of view of source playlist owners. The influence of official Spotify source playlists is not affected by the category of the target playlist, since no statistical differences were found between the three categories. Independent playlists, on the other hand, have the most influence on other independent playlists, significantly less influence on official Spotify playlists, and significantly even less influence on major label playlists. Furthermore, major label playlists have the most influence on other major label playlists, but significantly less influence on both independent playlists and major label playlists.

7.1 Managerial implications

Previous research has made clear that when releasing new music, addition to official Spotify playlists should be an important goal for artists and their managers (Voogt et al., 2018). These playlists determine which tracks and artists are discovered by consumers, and significantly boost their success (Aguiar & Waldfogel, 2018). Our research shows that tracks are mainly spread to these official Spotify playlists through other Spotify playlists. This means that spending time and resources on maintaining a positive relationship with Spotify (e.g., through pitching to playlist curators, or by signing a licensing agreement) is likely to pay off.

But even more importantly, we have several findings that point to a great impact caused by the presence of the major labels on Spotify. Official Spotify playlists might be influenced *most* by other official Spotify playlists, but they are also to a great extent influenced by major label playlists. The major label playlists themselves, in turn, are mainly influenced by other major label playlists. Appearing on major label playlists thus greatly increases the likelihood that tracks will spread further to official Spotify playlists, and as a result, increases the probability of the tracks' success. Moreover, in our descriptive analysis we found that tracks released by major labels were on average added to 9 playlists, while tracks released by indie labels were on average added to only 4 playlists. In addition, we found that Spotify is greatly biased towards adding tracks released by major labels to its own playlists.

Taken together, these results suggest that releasing new music through major labels (or signing a distribution deal with them) will lead to more exposure through the major label playlists, a higher chance of being added to official Spotify playlists, and as a result of both these facts more additions to all playlist categories overall. While being an independent artist might have its appeal, we believe that distributing and promoting new music in corporation with a major label will not only lead to increased success, but may even be a necessity to attain enough visibility on the platform to be discovered by consumers in the first place.

7.2 Limitations and future research

A potential limitation in our study is that our definition of influence may mistakenly attribute influence to what in reality is a sequence of independent events. For example, it might be the case that a source playlist A does not actually influence a target playlist B to add the same tracks, but that playlist A and playlist B just often add the same tracks due to similar taste, or due to using the same data to base their playlisting decisions on. In other words, do we really measure *influence*, or do we measure the *correlation* between two playlists' track additions? Anagnostopoulos, Kumar, and Mahdian (2008) developed a test for distinguishing between influence and correlation in social networks, which they called the *edge-reversal* test. The idea behind this test is that in the case of influence, tracks spread in a specific direction along an edge between two playlists (for example, tracks spread from a source playlist A to a target playlist B), and reversing the edge (to where B is the source playlist that spreads tracks to target playlist A) would change the estimated amount of influence. In the case of correlation, on the other hand, who adds the track first does not matter, and reversing the edge would not make a difference; the number of tracks that spread from playlist A to playlist B would, after a large number of observations, on average be the same as the number of tracks that spread from playlist B to playlist A. In our case, reversed edges clearly lead to different results. For example, the average influence of major label playlists on independent playlists is significantly higher than the average influence of independent playlists on major label playlists. Since we calculated these two averages using the exact same playlist pairs but in a reversed relation to each other, this implies that we indeed measure influence and not just correlation. However, it might also be the case that major label playlists simply tend to be faster in adding tracks to their playlists after their release date than independent playlists are, and that this is what causes the difference between the two averages. In the research by Ren et al. (2014) the time when a user first listened to a track indeed correlated with the influence of the user in spreading the track to others. Further research should investigate the issue of influence versus correlation more in-depth by applying the edge reversal test, while additionally controlling for the number of days after tracks' release dates different playlists owners add tracks to their playlists, which is readily available in our data.

Our work is the first to investigate how Spotify playlists influence each other and therefore leaves several additional opportunities for further research. First, further research could make a more detailed distinction between playlist owners. We treated the category of independent playlists as if this solely consisted of user-created playlists, while it in reality also comprises playlists created by artists, independent record labels, non-music companies, and more. In addition, playlists can be distinguished on a number of characteristics other than their owner; for example, more detailed categories could be created that also distinguish between algorithmic and editor curated Spotify playlists, between charts and mood-based lists, between playlists with recent and older music, etcetera. A lot of this information is already available in the data used in our study, so further research can easily investigate the effects of these variables using the same data and methods.

Finally, different tracks might spread through playlists differently. In our data preparation, we already made the distinction between major label tracks and indie label tracks, but we did not yet use this as a variable in our analysis. Further research should be undertaken to investigate the differences between these tracks.

References

Aguiar, L., & Waldfogel, J. (2018). Platforms, promotion, and product discovery: Evidence from Spotify playlists. *National Bureau of Economic Research, No. w24713*.

Anagnostopoulos, A., Kumar, R., & Mahdian, M. (2008). Influence and correlation in social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 7-15). ACM.

Bakshy, E., Hofman, J. M., Mason, W. A., & Watts, D. J. (2011). Everyone's an influencer: quantifying influence on twitter. In *Proceedings of the fourth ACM international conference on Web search and data mining* (pp. 65-74). ACM.

Bhattacharjee, S., Gopal, R. D., Lertwachara, K., Marsden, J. R., & Telang, R. (2007). The effect of digital sharing technologies on music markets: A survival analysis of albums on ranking charts. *Management Science, 53*(9), 1359-1374.

Cai, H., Chen, Y., & Fang, H. (2009). Observational learning: Evidence from a randomized natural field experiment. *American Economic Review*, *99*(3), 864-82.

Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, K. P. (2010). Measuring user influence in twitter: The million follower fallacy. In *fourth international AAAI conference on weblogs and social media*.

Datta, H., Knox, G., & Bronnenberg, B. J. (2018). Changing their tune: How consumers' adoption of online streaming affects music consumption and discovery. *Marketing Science*, *37*(1), 5-21.

Domingos, P., & Richardson, M. (2001). Mining the network value of customers. *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 57-66). ACM.

Fowlkes, K. (2019). *Why sign a distribution deal?* Retrieved from https://medium.com/@esqfowlkes/why-sign-a-distribution-deal-b9fd3870d62e

Garg, R., Smith, M. D., & Telang, R. (2011). Measuring information diffusion in an online community. *Journal of Management Information Systems*, *28*(2), 11-38.

Goodrich, G. (2019). *Meet Music Tastemakers of 2019: Independent Playlist Curators.* Retrieved from https://medium.com/playlist-push/meet-musics-tastemakers-of-2019-independent-playlist-curators-c56d52681257

Guille, A., Hacid, H., Favre, C., & Zighed, D. A. (2013). Information diffusion in online social networks: A survey. *ACM Sigmod Record*, *42*(2), 17-28.

Huberman, B. A., Romero, D. M., & Wu, F. (2008). Social networks that matter: Twitter under the microscope. *arXiv preprint arXiv:0812.1045*.

Im, H., Song, H., & Jung, J. (2018). A survival analysis of songs on digital music platform. *Telematics and Informatics, 35*(6), 1675-1686.

Leskovec, J., Adamic, L. A., & Huberman, B. A. (2007). The dynamics of viral marketing. *ACM Transactions on the Web (TWEB), 1*(1), 5.

Newberry, P. W. (2016). An empirical study of observational learning. *The RAND Journal of Economics, 47*(2), 394-432.

Pierce, D. (2017). *The secret hit-making power of the Spotify playlist.* Retrieved from https://www.wired.com/2017/05/secret-hit-making-power-spotify-playlist/#

Ren, J., Cheng, Z., Shen, J., & Zhu, F. (2014). Influences of influential Users: An empirical study of music social network. In *Proceedings of International Conference on Internet Multimedia Computing and Service* (p. 411). ACM.

Ren, J., & Kauffman, R. J. (2017). Understanding music track popularity in a social network.

Romero, D. M., Meeder, B., & Kleinberg, J. (2011). Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on Twitter. *Proceedings of the 20th international conference on World wide web* (pp. 695-704). ACM.

Saito, K., Kimura, M., Ohara, K., & Motoda, H. (2010). Discovery of super-mediators of information diffusion in social networks. In *International Conference on Discovery Science* (pp. 144-158). Springer, Berlin, Heidelberg.

Saito, K., Kimura, M., Ohara, K., & Motoda, H. (2016). Super mediator–A new centrality measure of node importance for information diffusion over social network. *Information Sciences, 329*, 985-1000.

Sorensen, A. T. (2007). Bestseller lists and product variety. *The journal of industrial economics, 55*(4), 715-738.

Sorensen, A. T. (2017). Bestseller Lists and the Economics of Product Discovery. *Annual Review of Economics, 9*, 87-101.

Spotify 101: Understanding the Spotify Playlist Landscape. (2019). Retrieved from https://playlistradar.com/spotify-101-understanding-the-spotify-playlist-landscape/

Tucker, C., & Zhang, J. (2011). How does popularity information affect choices? A field experiment. Management Science, 57(5), 828-842.

Van den Eerenbeemt, M. (2017, September 7). Warner Music koopt Nederlands dancelabel Spinnin' Records, inclusief conflict Martin Garrix. *De Volkskrant.* Retrieved from https://www.volkskrant.nl/cultuur-media/warner-music-koopt-nederlands-dancelabel-spinnin-records-inclusief-conflict-martin-garrix~b86d322f/

Voogt, B., Yau, J., van Doorne, T., & Erickson, E. (2018). *Spotify Marketing Strategy.* Retrieved from https://heroic.academy/spotify-marketing-strategy/

Waldfogel, J. (2017). How digitization has created a golden age of music, movies, books, and television. *Journal of Economic Perspectives, 31*(3), 195-214.

Appendix

Table 1

Descriptive statistics for tracks

Label	N tracks	N playlist additions	Playlist additions per track
Major label	17,832 (35.7%)	165,377	*M* = 9.3, *SD* = 18.4
Indie label	32,113 (64.3%)	140,814	*M* = 4.4, *SD* = 8.0
Total	49,945	306,191	*M* = 6.1 *SD* = 13.0

Table 2

Descriptive statistics for playlists

Category	N playlists	N playlist additions	N additions per playlist
Official Spotify	3,126 (5.5%)	18,279	*M* = 5.9, *SD* = 11.4
Major label	2,361 (4.1%)	9,666	*M* = 4.1, *SD* = 5.8
Independent	51,768 (90.4%)	278,279	*M* = 5.4, *SD* = 12.7
Total	57,255	306,191	*M* = 5.4, *SD* = 12.4

Table 3

ANCOVA results

Predictor	Sum of Squares	*df*	Mean square	*F*	*p*
Source playlist owner	2	2	1	585.50	.000***
Target playlist owner	1	2	0	160.50	.000***
Number of tracks added to source playlist	386,149	1	386,149	191,400,000.00	.000***
Number of tracks added to target playlist	8,833	1	8,833	4378000.00	.000***
Source playlist owner*Target playlist owner	0	4	0	44.94	.000***

*** *p* <.001

Table 4

Means and standard deviations for influence of source playlists on target playlists, categorized by source playlist owner and target playlist owner

	Target playlist owner					
	Independent		Spotify		Major label	
Source playlist owner	*M*	*SD*	*M*	*SD*	*M*	*SD*
Independent	2.49·10-4	9.59·10-3	2.12·10-4	8.88·10-3	2.49·10-4	8.54·10-3
Spotify	2.24·10-4	8.36·10-3	4.46·10-4	1.29·10-2	2.77·10-4	9.62·10-3
Major label	3.25·10-4	1.13·10-2	4.25·10-4	1.27·10-2	9.18·10-4	1.92·10-2

Note. *M* and *SD* represent mean and standard deviation, respectively.

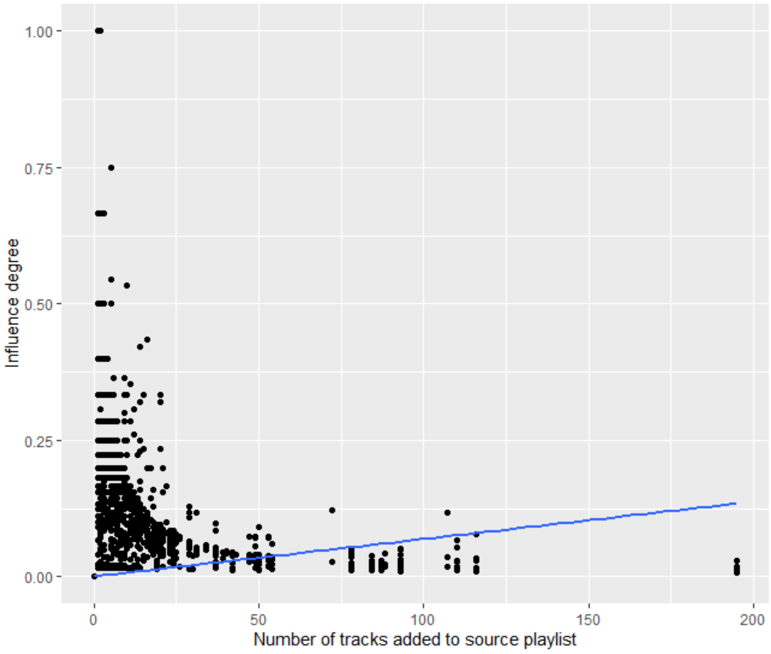

Figure 1. Relation between the number of tracks added to source playlists and their influence on other playlists.

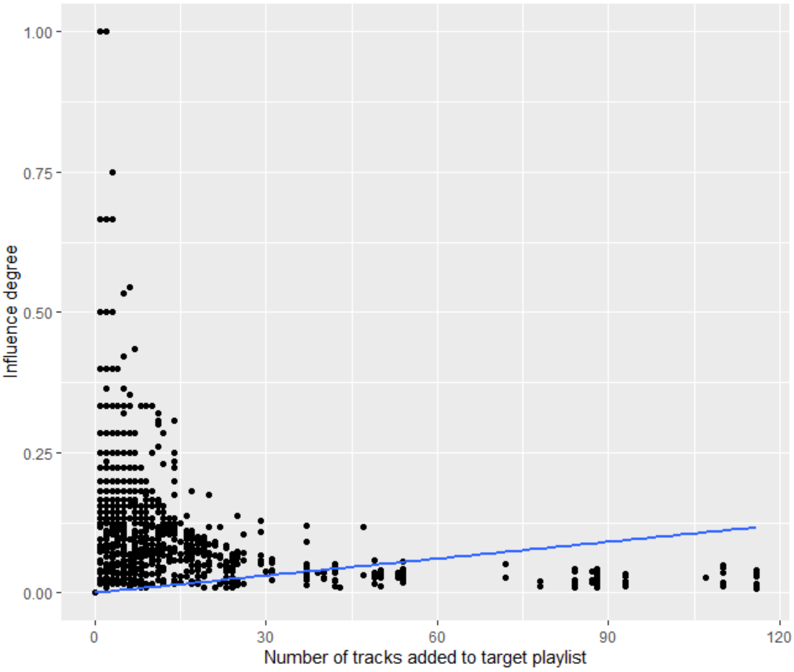

Figure 2. Relation between the number of tracks added to target playlists and their influence on other playlists.

Table 5

Results of post hoc comparisons using Tukey's HSD

	Source playlist owner	Independent			Major label			Spotify		
	Target playlist owner	Independent	Major label	Spotify	Independent	Major label	Spotify	Independent	Major label	Spotify
Source playlist owner	Target playlist owner									
Independent	Independent									
	Major label	.00***								
	Spotify	.00***	.00***							
Major label	Independent	.00***	.00***	.00***						
	Major label	.00***	.00***	.00***	.00***					
	Spotify	.00***	.00***	.00***	.99	.00***				
Spotify	Independent	.00***	.00***	.00***	.00***	.00***	.00***			
	Major label	.92	.04*	.99	.00***	.00***	.00***	.08		
	Spotify	.06	.00***	.00***	.00***	.00***	.00***	.97	.07	

* $p < .05$

*** $p < .001$