CESifo WORKING PAPERS

10120
2022

November 2022

Selection, Consumption, and New Music Exploration in an Online Social Network: A Dyadic Approach

Johannes Loh

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University’s Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10120*

Selection, Consumption, and New Music Exploration in an Online Social Network: A Dyadic Approach

Abstract

We study peer influence in an online social network on a platform where consumers purchase music albums. They can follow their peers and become informed about their consumption choices. In particular, we are interested in how this affects consumers' exploration of new music that exhibits unfamiliar attributes (e.g. artist, genre, or instrumentation). Our empirical analysis contains two parts: First, we analyze how the formation of new dyads in the network depends on consumer-peer similarities in their preference for certain album attributes. This affects music exploration because it determines which peer purchases consumers are exposed to. Second, conditional on the determinants of dyad formation, we investigate how within-dyad information flows affect consumers' purchase decisions, and in particular their exploration of unfamiliar attributes. Our analysis produces three key findings: First, preference similarities between consumers and peers are the strongest predictor of the formation of dyads. This likely stifles consumers' exploration of new music because it limits their exposure to unfamiliar attributes. Second, we find a strong positive peer effect of consumers observing peer purchases after the formation of a dyad. Third, this effect is stronger for albums from unfamiliar artists, but weaker for those that exhibit unfamiliar horizontal attributes (e.g. its genre). Together, this suggests that new music exploration in online social networks is limited and subject to nuance.

JEL-Codes: D120, D830, L820, L860.

Keywords: online social networks, consumption choices, peer influence, music consumption.

Johannes Loh
BI Norwegian Business School
Oslo / Norway
johannes.loh@bi.no

October 1, 2022

Selection, Consumption, and New Music Exploration in an Online Social Network: A Dyadic Approach

Johannes Loh
BI Norwegian Business School
johannes.loh@bi.no

October 1, 2022

1 Introduction

Online social networks enable consumers to observe the choices and behavior of their peers (Aral & Walker, 2011; Dewan et al., 2017). In the context of online shopping, this conveys information about product attributes and quality, which makes it an important driver of consumption choices (Hu et al., 2019). Many e-commerce platforms therefore make increasing use of digital technologies to foster interactions between their users and to facilitate the salience of their purchase decisions (Qiu et al., 2021). A key mechanism through which peer behavior influences consumption choices is observational learning (Cai et al., 2009; Katz & Lazarsfeld, 2017): By observing peers making purchases through an online network, consumers learn about the existence of certain products as well as their quality and attributes.

Prior research has produced rich insights about how this shapes consumption across a wide variety of empirical contexts, such as movies (Moretti, 2011), restaurant dining (Cai et al., 2009; Qiu et al., 2018) and kidney transplants (Zhang, 2010). In addition, scholars have explored several sources of heterogeneity like network structure (Ameri et al., 2019; Zhang et al., 2018), product attributes (Wang et al., 2013; Zhu & Zhang, 2010), and different consumer characteristics (Aral & Walker, 2012; Tucker, 2008). These studies provide valuable insights and consider a wide spectrum of determining factors. However, they also produce some contradictory findings, suggesting that the effectiveness of learning from peer choices depends on additional factors.

In this paper, we study peer influence in music consumption in an online social network, and in particular the *exploration* of new music. We are interested in if and how the effectiveness of observational learning varies by consumers' familiarity with its horizontal product attributes. Music is highly differentiated, with different genres, instrumentation, and other stylistic elements, and consumers do not value each equally. Moreover, digital technologies increased both the consumption and supply of music, and niche products in particular (Aguiar & Waldfogel, 2016, 2018), which can create benefits from variety for consumers (Brynjolfsson et al., 2010). However, the sheer amount of available options confronts them with an overload of information (Anderson & de Palma, 2012; Dinerstein et al., 2018). In addition, its experience good nature makes utility from music consumption uncertain (Datta et al., 2018) and exploration costly (Bronnenberg, 2015). As a result, the information gained from observing peer choices may be particularly useful in this context. However, with the exception of a few studies that point to distributional effects on product sales (Salganik et al., 2006; Tucker & Zhang, 2011; Zhu & Zhang, 2010), evidence for the role of peer influence in exploration is sparse.

We use a dyad-level perspective and study two aspects of consumer behavior in online social networks that can influence their exploration of new music. First, we study determinants of the formation of consumer-peer dyads. The choice to follow a certain peer will determine which purchases can be observed through the network. In particular, when a consumer forms a dyad with a peer who exhibits similar preferences towards horizontal attributes, exposure to products that constitute an exploration opportunity may be limited. Second, and conditional on peer selection, we study how observed peer choices within a dyad affect the consumer's subsequent purchase decisions, and the exploration of unfamiliar music in particular. Here, we posit that this lets the consumer learn about the existence of a particular album, and that it conveys information about its quality. Hence, observing peer purchases has both an *awareness* and a *signaling* effect. We argue that both depend on the consumer's familiarity with the product's attributes, which in turn determines the strength of peer influence. Together, we therefore ask three research questions:

1. How does consumer-peer similarity in preferences for product attributes drive the formation of dyads in the network?
2. What is the causal peer effect of observing peer purchases on consumers' choices?
3. How does this peer effect differ by their familiarity with the product's horizontal attributes, i.e. the extend to which it constitutes an exploration opportunity?

We investigate these aspects in context of Bandcamp, a digital music platform where consumers purchase albums from independent musicians. The platform maintains a social network where consumers can follow their peers and become informed about their purchases. Detailed information about albums' horizontal attributes (e.g. genre and instrumentation) and consumers' purchase histories lets us track the development of their preferences over time and identify exploration opportunities. Moreover, in combination with information about the evolution of the network over time we evaluate how consumer-peer similarities drive dyad formation, which peer purchases are observed, and how this drives consumption choices on the platform.

Our empirical analysis contains two parts: First, we investigate consumers' decisions to follow a certain peer within a discrete choice modeling framework (section 4). We find that similarities in the preference for music artists as well as horizontal attributes are the strongest predictors for the formation of new dyads in the network. This speaks to homophilous tie formation dynamics that limit consumers' exposure to unfamiliar attributes, thus stifling their exploration of new music.

Second, conditional on peer selection, we analyze how information flows within each dyad affect consumption choices (section 5). We leverage information about the sequence of tie formations, peer purchases, and consumer choices to estimate a causal peer effect. To study exploration, we then let this effect vary by the consumer's familiarity with the attributes of the observed peer purchase. We find a strong positive effect. Specifically, the probability that a consumer purchases the same album as a peer they are connected to is 2.61 times higher when they are able to observe that purchase (compared to not being able to). Moreover, this effect is *stronger* for albums from unfamiliar artist, but *weaker* for those that exhibit unfamiliar horizontal attributes. This implies that consumers do discover new music through the network, which however corresponds to their horizontal preferences. Hence, the degree to which information flows in online social networks promote new music exploration is limited and subject to nuance.

2 Related Literature

Much evidence exists about the role of peer influence in consumption decisions. It has been demonstrated in a variety of empirical contexts, such as the box office performance of movies (Moretti, 2011), Anime watching (Ameri et al., 2019), restaurant discovery (Qiu et al., 2018), cellphone purchases (De Matos et al., 2014), Facebook app usage (Aral & Walker, 2011), adoption of a paid music subscription service (Bapna & Umyarov, 2015), and music listening behavior (Dewan & Ramaprasad, 2014; Dewan et al., 2017; Salganik et al., 2006). When information about a product or service is limited or unavailable otherwise, observing the behavior of others can shape consumers' beliefs and affect their choices. That is, *social* or *observational learning* can help them infer a product's quality before consuming it themselves (Banerjee, 1992; Bikhchandani et al., 1992). This learning effect has received broad empirical support. In an experiment using an articial music market, Salganik et al. (2006) shows that earlier consumption choices affect later ones when they can be observed. In a field experiment with restaurant goers, Cai et al. (2009) demonstrate that showing information about the five most popular dishes increased their demand by 13 - 20%, with a stronger effect for customers with less information ex ante. Moretti (2011) show a positive learning effect in the context of movies. Those that do surprisingly better than expected in their opening week will experience increased box office performance in subsequent weeks, because consumers adjust their beliefs about their quality. In the context of the kidney transplants, Zhang (2010) shows that rejections by earlier patients send a negative quality signal to subsequent ones on a waiting list, making them reject them at higher rates, too. Similarly, Qiu et al. (2018) and Hu et al. (2019) show that consumers learn from their friend's actions in the context of restaurant choice and switching wireless carriers, respectively.

While the existence of peer influence is well-documented, the strength of observational learning can vary. Hence, a bulk of research studies sources of heterogeneity. First, consumer and peer characteristics have received some attention. In the context of a video messaging tool in an investment bank, Tucker (2008) shows that only managers and employees in boundary spanning positions exert meaningful influence, but ordinary employees do not. Similarly, Aral & Walker (2012) identify different groups of Facebook users with varying degrees of both influence and susceptibility, and Wang et al. (2013) show that the relative influence of experts and popular individuals varies by product type.

Second, several characteristics of the consumer's network have been studied. Moretti (2011) and Wang et al. (2018) investigate the role of network size in the context of movie going and book reviews, respectively, with diverging findings. The former finds a stronger influence in larger networks, while the latter find that it is stronger in smaller networks. In addition, multiple studies make a distinction between observing choices of direct network ties and popularity information from the whole network. In the context of an online music community Dewan et al. (2017) show that "proximity" and popularity influence act as substitutes, and that the former dominates the latter. In contrast, Ameri et al. (2019) show that information from the whole network has a stronger learning effect than information from direct ties in an anime watching community. Zhang et al. (2018) show that direct ties have a strong positive influence in the adoption of "caller ringback tones" in a cellular network, while the direction of the effect from indirect ties depends on network size: it is positive in larger, and negative in smaller networks. Finally, Zhang & Godes (2018) demonstrate heterogeneous effects between weak (i.e. unidirectional) and strong (bidirectional) network ties in a online community about books.

Third, several product characteristics have received attention. Qiu et al. (2021) develop a theoretical model with a distinction between vertically (i.e. the quality dimension) and horizontally differentiated products, which they subsequently test in a lab experiment. They find that learning from strangers is stronger for vertically differentiated, and learning from friends is stronger for horizontally differentiated products. Wang et al. (2013) find heterogeneous learning effects for technology and fashion products. For the former learning from experts is more effective, and for the latter learning from more popular individuals. Others have considered distributional aspects. In their experiment, Salganik et al. (2006) find that peer influence reinforces the "long tail" distribution of music adoption, which implies a stronger effect for more popular products. In contrast, Zhu & Zhang (2010) find a stronger effect of online consumer reviews on less popular video games. Similarly, Tucker & Zhang (2011) find a stronger effect for niche or "narrow-appeal" products in a field experiment from a website for wedding service vendors, which they attribute to consumers attaching a relatively higher quality premium compared to "broad-appeal" products.

Together, these studies demonstrate various sources of heterogeneity in peer influence and observational learning. However, the aspect of consumers' experience and familiarity with certain products or product types has not received much attention yet, even tough it should be an important determinant of the strength of peer influence. We argue that this is to a large extent determined by the degree of new information the consumer receives through the network: Information about *familiar* product types may be redundant either because the consumer already knows it or can easily acquire it elsewhere. In contrast, a social network may be the primary source of information about *unfamiliar* product types. In turn, this has implications for consumers' exploration of new or unfamiliar products.

In the context of music, consumers have learning opportunities other than observing peers' behavior, which can also vary by product type. Datta et al. (2018) study how a shift from owning to streaming music affects consumer's listening behavior. Consistent with a decrease in exploration costs, they find that those that adopt streaming increase both the quantity and diversity of music they listen to. Kretschmer & Peukert (2020) study the quasi-random (un-)availability of music on German YouTube. They find that it increased music sales for new artists and mainstream music in particular, indicating that YouTube presents an inexpensive discovery channel. Aguiar & Waldfogel (2021) study the effect of songs being added to and dropped from Spotify playlist with nuanced findings. While most playlists favor songs from U.S. based major labels, those dedicated to showcasing smaller artists benefit songs from indie labels more. In all, these studies show that providing less costly modes of discovery can nudge consumers toward exploration of new music.

In this paper, we are interested in how observing peer purchases in an online social network affects consumers' choices, and their exploration of new (to them) horizontal product attributes in particular. We adopt a dyadic approach in that we study (i) how new ties are formed in the network, and (ii) the effect of the information flows from the peer to the consumer within a dyad. Both should impact music exploration. First, selecting into a dyad determines which types of peer purchases a consumer is exposed to. If the peer exhibits similar horizontal preferences, then the albums they purchase will also reflect the focal consumer's preferences. Hence, opportunities for new music exploration are potentially reduced. Similar dynamics have been considered by political scientists who study "echo chambers" on social media platforms: Users engage in network ties that reflect their own beliefs and, as a result, reinforce them (Barberá et al., 2015; Colleoni et al., 2014; Dubois & Blank, 2018). In addition, Ameri et al. (2022) show that homophily is an important determinant of friendship formation in an anime watching community. Second, after a dyad has been formed, the consumer is able to observe the peer's purchases with two effects on her choices: She learns about the existence of an album – an *awareness* effect. In addition, it updates her belief about its quality – a *signaling* effect. We argue that both vary between albums that exhibit more or less familiar horizontal attributes.

We contribute to the literature in three ways: First, study the formation of dyads and its implications for subsequent consumption choices. While especially the role of homophily has been discussed in previous work, we explicitly analyze and discuss how peer selection relates to the exploration of products with unfamiliar horizontal attributes. Second, we contribute to the discussion around heterogeneity in peer influence by looking at information flows within a dyad and considering consumers' familiarity with product attributes. Here, we highlight how their experience relates to the usefulness of the information received through the network and how it affects product exploration at the individual level. Third, we contribute to the literature about digitization and music consumption by demonstrating the role of online social networks as a channel for new music discovery.

3 Setting and Data

3.1 Empirical Setting: Bandcamp

We study an online social network on Bandcamp, a digital music platform that has been established in 2009. On the platform, consumers purchase albums or songs as digital download or in physical form. Consumers can, but do not have to, set up an account. Musicians can also sell other merchandise such as T-shirts. In contrast to larger alternatives such as Spotify or Apple Music, Bandcamp has a distinct focus on independent musicians, hence it mainly features lesser-known and amateur artists who are not signed with a major label or, in most cases, any label at all. In addition, Bandcamp allows for limited[1] streaming of songs, but artists primarily sell complete albums. While creating a presence and uploading music to Bandcamp is free, its primary source of revenue is a share of ten to 15 percent of of each sale on the platform.[2]

=== Figures 1, 2, and 3 here ===

The focus on independent music is a clear differentiating factor for the platform, and it mainly targets a consumer base of music "aficionados". It stresses the possibility for fans to "directly support the artists they love"[3] and it maintains a blog[4] featuring daily articles about genre developments, artist and album spotlights, and other related content. In addition, album pricing contains a "pay-what-you-want" element in that consumers typically have the possibility to pay

[1] If a consumer has not yet purchased an album, they can still stream songs for a limited time and at reduced quality.

[2] https://get.bandcamp.help/hc/en-us/articles/1500006084082-What-are-Bandcamp-s-fees-

[3] https://bandcamp.com/about

[4] https://daily.bandcamp.com/

more than the asking price - which they do in more than 40 percent of transactions according to the platform.[5] An additional source of revenues comes from "Bandcamp for Labels"[6] which gives (independent) labels the possibility to set up dedicated pages and centrally manage associated artists. It also features guides for artists on aspects like pricing.

In early 2013, Bandcamp underwent a significant redesign, which stressed the community element of the platform. Under the label "Bandcamp for Fans"[7] it introduced a range of new features. First, it allows consumers to register a "fan" account which equips them with a profile page (see Figure 1). Second, it introduced collection pages that allow these registered consumers to showcase their purchases in their profile. In addition, album pages now contain a "supported by" section with all registered consumers who added it to their collection (see Figure 2). Third, it introduced social network functionality by allowing registered consumers to connect with their peers. This connection resembles Twitter in that they "follow" others, hence establishing a unidirectional tie unless followed back.

Together, these new features[8] introduced several new ways for consumers to connect with others and to discover music. While consumers can access any registered peer's album collection via their profile page, following them enables two additional channels to observe peer purchases. First, Bandcamp sends them a "digest" email about it, which occurs immediately. Second, it introduced a "fan feed" containing information about these purchases along other information such as relevant new album releases (see Figure 3). In addition, the platform facilitates the formation of new network ties via two elements. Most channels includes algorithmic recommendations for peers to follow, which primarily draw on common albums, artists, and network connections. In addition, the aforementioned "supported by" section lets consumers discover new potential ties via album pages.

In this study, we are interested in two aspects of these features. First, we study how similarities between consumers and potential peers drive the establishment of new ties. The channels of peer discovery likely nudge consumers towards forming ties with peers who exhibit a high degree of similarity in their horizontal preferences. This will influence their subsequent exploration of new horizontal product attributes because it determines with type of peer purchases they are exposed to. Second, conditional on the observed and unobserved characteristics of the consumer-peer dyad, we study how the ability to observe peer purchases affects subsequent consumption patterns. In particular we analyze if they are more likely to

[5] https://get.bandcamp.help/hc/en-us/articles/360007802534-What-pricing-performs-best-

[6] https://bandcamp.com/labels

[7] https://blog.bandcamp.com/2013/01/10/bandcamp-for-fans/

[8] Note that they introduced a few additional features which are not directly related to our research questions.

purchase the same album as the peer, and to what extent this effects varies by their familiarity with the album's horizontal attributes. As such, we are interested in how it affects their exploration of new music.

3.2 The Data

Using webscraping techniques, we acquired data on all albums, purchases, and the social network from Bandcamp in March 2020. First, our data set contains 3,226,898 albums by 994,150 artists that have been uploaded to the platform since its launch. We collected information about its price and currency, the number of songs, its duration, and release date. We use tags as indication for its horizontal attributes, such as its genre, instruments, and other stylistic elements. Artists assign them themselves, and the platform uses them to classify albums under certain genres and sub-genres. While we do not have information about artists' countries of origin, the currency in which they offer their albums gives an indication. The vast majority of artists is located in the U.S. (59.78%), the EU (18.55%), the UK (8.61%), or Canada (5.19%). Speaking to the high presence of amateur artists on the platform, 49.17% of albums are offered for free. For all others, the average price varies by the used currency. For those offered by U.S. artists it is 6.15 USD.

Second, we collected data on registered consumers adding an album to their collection. Usually, this happens via the purchase of an album, but there is another possibility. Artists can produce a limited amount of vouchers for free downloads which they use as promotional tools, for instance to send to music critics. Still, going forward, we refer to such a collection addition as purchase or sale, as this is by far the most common way. In addition, this means that we only observe purchases by registered consumers, but not others. While we consider this a data limitation[9], we are not overly concerned here. We study peer influence in the social network and only registered consumers can follow others. Next, we extracted an exact timestamp for each purchase, which lets us track consumers' complete purchase histories. In total, our data set contains 32,849,290 purchases made by 1,468,631 registered consumers who appear with their first purchase.[10] At the time of our data collection the average consumer had bought 22.39 albums. However, this distribution is skewed, and the median consumer had bought 6 albums. On the other side of the transaction, 55.05% of all albums in our data set do not exhibit any

[9] As we do not have information about purchases by unregistered consumers, the calculation of albums sales here has to be understood as a lower bound. A potential bias may be in place if more successful albums attract more interest from unregistered consumers (e.g. through word-of-mouth from other sources). Then, the understatement of sales figures is more pronounced for more successful albums.

[10] Note that it is possible for consumers to create an account on Bandcamp without ever making a single purchase. As a result of our data collection process we are unable to obtain information about them.

sales. For those that do, the average album sold 22.67 copies. Again, the distribution is skewed with a median of 4 sales.

Finally, we obtained information about the social network. Our data set contains a total of 2,188,872 ties between two consumers on the platform and an exact timestamp of its formation. This lets us track how the network developed over time. A total of 272,300 consumers follow at least one peer, or 18.54% of all registered who made a least one purchase. Out of those who are part of the network, the average consumer follows 191.11 peers, and is followed by 163.61 others. However, these distributions are skewed, and the consumer at the median only follows 42 and is followed by 23 others. In addition, most ties are not reciprocated, with consumers following each other in 22.9% of all cases.

=== Figure 4 here ===

In combination with the complete purchase history, this lets us track which peer purchases each consumer is exposed to over time, and whether or not they also purchased these albums afterwards. Figure 4 shows the development of available albums, album sales, and new network ties over time.

3.3 Exploration, Horizontal Product Attributes, and Consumer Preferences

In this study, we define exploration as the purchase of products (albums) that exhibit attributes that are unfamiliar to the focal consumer. Here, we consider two types of such *exploration opportunities*: First, we consider the purchase of an album from an unfamiliar or new (to the consumer) artist, and we use information about consumers' purchase histories to identify this. Second, we consider the purchase of an album with unfamiliar horizontal attributes such as its genre, instrumentation, or other stylistic elements. The two are similar but distinct as a "new" artist may well exhibit musical characteristics the consumer is familiar with.

We use tags associated with albums to operationalize horizontal attributes. While assigned by the artist themselves, the platform maintains a wide range of normalized tags it suggests during the album upload and uses for further genre classifications. Still, it is possible for artists to assign any character string as a new tag. As a result, our data set contains a total of 737,728 unique tags, with the vast majority being attached to a single album.[11] Here, we use only a small subset of all tags for two reasons. First, the vast majority does not provide any useful information. Second, it provides a means of dimensionality reduction which is necessary to make

[11] The high number is a result of artists typing in their own tags, which often make reference to a very small town, similar bands, and inside jokes. In addition, many are typos.

our calculations feasible. As a starting point, we selected the top 200 most assigned tags. While only a small fraction of all unique tags, we find that 99.21% of albums in our sample exhibit at least one of them. Not all of them represent stylistic musical attributes, but make reference to a location instead and we manually removed them. In the end, we used a total of 156 tags representing genre information and other stylistic elements for our operationalization of albums' horizontal attributes and consumer preferences. Examples are "electronic", "rock", "rap", and "instrumental".[12] With this set, we construct our representations for horizontal attributes and consumer preferences.

First, to represent an album's horizontal attributes, for each we constructed a vector that indicates which tags are associated with it. We call this the *attributes vector*:

$$\mathbf{C_j} = \begin{pmatrix} c_{j1} & c_{j2} & \cdots & c_{jM-1} & c_{jM} \end{pmatrix} \tag{1}$$

with $c_{jm} \in \{0, 1\}$ indicating if tag m is associated with album j. For example, if the element representing "rock" appears as a one in the vector, it would indicate that the album is part of this genre. However, this approach goes beyond such simple genre classification by allowing for albums to span multiple genres, and by indicating additional attributes like its instrumentation.

Second, we use information about consumers' sequence of album purchases to capture their preferences for these attributes over time. Specifically, we start by summing up the attributes vectors of all albums she owns at a given point in time, yielding another 156-dimensional vector, $\mathbf{C_{it}} = \begin{pmatrix} c_{it1} & c_{it2} & \cdots & c_{itM-1} & c_{itM} \end{pmatrix}$, with $c_{itm} = \sum_t c_{jm}$ indicating the number of times consumer i has "consumed" tag m at time t. Then, we express the elements in $\mathbf{C_{it}}$ as shares by dividing each element by the number of total tag occurrences, arriving at our representation of consumer preferences:

$$\mathbf{P_{it}} = \begin{pmatrix} p_{it1} & p_{it2} & \cdots & p_{itM-1} & p_{itM} \end{pmatrix}, \tag{2}$$

with $p_{itm} = \frac{c_{itm}}{\sum_m c_{itm}}$ representing the share of each tag in a consumer's *preference vector*. As such, it is bound between zero and one and higher values mean that a consumer has purchased many albums associated with that tag, indicating that she has a relatively high preference for the horizontal attribute it represents. Table 1 contains an simplified example to illustrate our operationalization of attributes and preferences.

=== Table 1 here ===

We use these vector representations to construct two key measures we use in our analyses. First, in Section 4 we use the distance in preferences between a focal consumer and a potential

[12] Table A9 in the appendix contains the complete list of tags and the number of albums they are assigned to.

peer she may follow to evaluate how ex ante consumer-peer similarity drives the formation of new network ties. Second, in Section 5 we construct the distance between a focal consumer's preferences and the attributes of a potential album she may buy to evaluate if peer influence is stronger or weaker for purchase decisions involving products that constitute an exploration opportunity.

For both, we calculate the *soft* cosine distance (Sidorov et al., 2014) between the respective vectors. In contrast to the "regular"" cosine distance, this version allows for the possibility that certain elements of a vector are more similar to one another than others. We believe this to be appropriate in our case. For example, the tag "metal" likely represents stylistic elements that are more similar to "rock" than "jazz". In addition, some tags represent sub-genres such as "alternative-rock" or "underground-hip-hop". Therefore, we construct a measure for the similarity between two tags k and l using information about how often they are assigned to the same album. The idea is that tags that are more commonly co-assigned exhibit similar musical characteristics. We calculate this measure as

$$s_{mn} = \frac{|m \cap n|}{|m \cup n|}, \tag{3}$$

where $|m \cap n|$ is the number of albums that exhibit both tags m and n, and $|m \cup n|$ is those that exhibit either of the two. It is bound between zero and one with higher values indicating a higher conceptual similarity[13] between two tags. Similar approaches can be found in Leung (2014) and Kovacs & Hannan (2015). The soft cosine distance is then calculated as

$$d(\mathbf{A}, \mathbf{B}) = 1 - \frac{\sum_{m,n} s_{mn} \mathbf{A}_m \mathbf{B}_n}{\sqrt{\sum_{m,n} s_{mn} \mathbf{A}_m \mathbf{A}_n} \sqrt{\sum_{m,n} s_{mn} \mathbf{B}_m \mathbf{B}_n}}, \tag{4}$$

yielding a measure between zero and one, where larger values indicate a greater distance between two N-dimensional vectors $\mathbf{A}$ and $\mathbf{B}$. In Sections 4 and 5 we use this to calculate the distance in preferences between consumer-peer dyads and to what extent an album exhibits horizontal attributes that are distant from a consumer's preferences.

4 Peer Selection

Networks evolve endogenously and determine which peer purchases consumers observe. This has implications for the exploration of new music in particular. If consumers mainly follow peers that have similar preferences, they will be mainly exposed to peer purchases of albums

[13] By construction, the similarity is symmetric between two tags, i.e. m will be as similar to n as n is to m.

with similar horizontal attributes to those they would also purchase absent peer influence. Thus, the homophilous formation of dyads would stifle the exploration of new music. On the flip-side, consumers may choose to follow peers with the intent of new music exploration. In this case, we would observe increased formation of dyads between consumers that exhibit dissimilar preferences, exposing them to peer purchases of albums with unfamiliar horizontal attributes, and promoting exploration. In addition, platform design plays a role. Bandcamp uses algorithmic recommendations to facilitate the formation of new ties. They commonly highlight the commonalities between the consumer and the recommended peer, making reference to common albums, artists, or network ties. As a result, the platform draws on consumers' past behavior to nudge them towards following peers, reinforcing homophilous network evolution. Together, dyad formation in our setting is determined by consumer preferences and biases (e.g. homophily vs. exploration intention) as well as these platform nudges. The relative strength of these forces determines whether or not, and to what extent, the stage of peer selection promotes or stifles new music exploration.

In this section, we analyze how different elements of similarity between consumers and potential peers drive the formation of new ties. We use a discrete choice model framework to analyze peer selection in our context. Typically, they are used to model consumer choice among a set of products (Train, 2009). The purchase decision is then determined by her observable characteristics (e.g. demographic information) and product attributes (e.g. price), and she chooses the utility-maximizing alternative. We follow the same logic and apply it to our context. A consumer who participates in the network has to decide which peer to follow, and will choose the alternative that yields the greatest expected benefits in terms of her exploration needs.

4.1 Choice Model and Estimation

We model the choice of consumer i to follow peer $k \in \mathbb{C}_{it}$ at time t using a linear utility specification, where $\mathbb{C}_{it}$ denotes i's choice set. In principle, this includes all other consumers on the platform that the consumer does not already follow at time t. However, estimating such a model would not be feasible because it would require the modeling of a repeated choice (over t) among the full set of these alternatives. Our data contains 1,468,631 individual consumers, hence the computational requirements for estimating models with the full choice set at different points in time would be too great.[14] Hence, we take steps to reduce the size of the choice set

[14]For reference, consider modeling a repeated choice of all 272,300 consumers who participate in the network among all other 1,468,630 consumers on the platform at twelve different points in time. The final dataset would contain $272,300 \times 1,468,630 = 399,907,949,000$ potential dyads, and a total of $399,907,949,000 \times 12 = 4,798,895,388,000$ observations.

and the complexity of the model.

First, we reduce model complexity by only considering those points in time at which a consumer actually began to follow another peer. In so doing, we simplify the model along the time dimension compared to a repeated choice over additional periods where no peer was chosen. In addition, this approach abstracts away from the outside option of *not following* anybody. Second, we further reduce the choice set by only considering a random subset of 50 alternative peers the consumer did not choose to follow at that point in time. Similar approaches can be found in Nicolle (2020) and Train et al. (1987). Together, choice sets under consideration here include a total of 51 alternatives, consisting of the actually chosen peer and a random draw of consumers that had been active on the platform at that point in time, but that the consumer has not yet been following.

Consumer i then follows peer $k \in \mathbb{C}_{it}$ that maximizes her utility from doing so. That is, she chooses the peer whom she expects to provide her with the most relevant information flows in terms of her exploration needs. The utility is given by

$$U_{ikt} = \mathbf{X}_{it}'\beta + \mathbf{Y}_{kt}'\gamma + \mathbf{Z}_{ikt}'\delta + \epsilon_{ikt} = V_{ikt} + \epsilon_{ikt}. \tag{5}$$

Here, $\mathbf{X}_{it}$ and $\mathbf{Y}_{kt}$ are vectors of observable characteristics of consumer i and peer k at time t, respectively, and $\mathbf{Z}_{ikt}$ contains measures of similarity between the two. Further, ϵ_{ikt} captures unobserved dyad-specific utility shocks, for example i and k meeting outside the platform. Here, we are interested in studying how these factors impact the probability that i follows k at time t, hence that k is the utility-maximizing alternative in $\mathbb{C}_{it}$. Following the discrete choice literature (Train, 2009), this can be expressed as

$$\pi_{ik} = \frac{e^{V_{ikt}}}{\sum_{l \neq k} e^{V_{ilt}}}. \tag{6}$$

We run conditional logit regressions to analyze this choice. The dependent variable is a dummy indicating that consumer i is choosing to follow peer k at time t ($\text{IsDyad}_{\text{ikt}}$). By construction of the choice set, this is the case for one in 51 alternatives. We also include choice fixed effects in the regressions. Hence the variation we explore stems from differences between the 51 alternatives in each choice set, but not differences between these choices. This controls for unobserved time trends as well as consumer characteristics at this point in time. As a result, the term $\mathbf{X}_{it}'\beta$ is dropped due to collinearity with the choice fixed effects. The vector of consumer-peer similarity, $\mathbf{Z}_{ikt}$, contains the following variables: We construct the preference distance between i and k based on the tags they consumed up until t. Using their preference vectors and equation 4, we calculate this as

$$\text{PreferenceDistance}_{\text{ikt}} = d(\mathbf{P_{it}}, \mathbf{P_{kt}}).$$

This is a measure for the differences between i's and k's preferences towards horizontal product attributes that draws from their purchase histories on the platform. It is bound between zero and one, with greater values indicating more dissimilar preferences. In addition, we use the number of commonly owned albums ($\text{CommonAlbums}_{\text{ikt}}$), common unique artists they purchased albums from ($\text{CommonArtists}_{\text{ikt}}$), and common network ties ($\text{CommonTies}_{\text{ikt}}$). For the latter we do not distinguish between followees and followers (doing so yields similar results). These measures capture additional aspects of consumer-peer similarities for artists and other consumers on the platform. In addition, they directly factor into the platform's recommendations of potential peers to follow. Next, the vector of peer characteristics, $\mathbf{Y}_{kt}$, contains the total number of albums he owns ($\text{PeerAlbumCount}_{\text{kt}}$), unique artists he purchased these from ($\text{PeerArtisitCount}_{\text{kt}}$), and other peers he is following ($\text{PeerFollowees}_{\text{kt}}$) and followed by ($\text{PeerFollowees}_{\text{kt}}$). They capture aspects of a potential peer's "quality". For instance, if he exhibits a larger number of followers already, this may signal his good reputation to a consumer deciding whether or not to also follow him. Likewise, a higher number of owned albums may signal experience and musical competence. In addition, these variables correlate with our similarity measures – e.g. a higher number of unique artists for the peer would mechanically increase the number of common artists. Hence, we control for such conflating scale effects. Lastly, we include a measure of preference concentration. We calculate the Herfindahl-Hirschman Index (HHI) of k's preference vector, $\mathbf{P}_{\text{kt}}$, to capture the variety of horizontal characteristics he has consumed in the past ($\text{PeerPreferenceHHI}_{\text{kt}}$,[15]) which may be a determinant of peer selection. For instance, if a consumer seeks to explore unfamiliar horizontal attributes, she may choose to follow a peer that exhibits a broad range in his album collection.

4.2 Results

4.2.1 Descriptive Statistics

Table 2 contains summary statistics for the variables used in our conditional logit regressions. We provide separate statistics for peers and dyads that had actually been chosen by a consumer and the 50 random alternatives that have not. We will not discuss each individual variable here, but we highlight some aspects that preface our estimation results. First, in terms of peer

[15] This is calculated as $\sum_m p_{ktm}^2$.

characteristics, those who are actually followed by the focal consumer ($\text{IsDyad}_{\text{ikt}} = 1$) exhibit considerably higher averages for the number of owned albums, unique artists purchased from, as well as others they are following and followed by, compared to the random draw of alternatives ($\text{IsDyad}_{\text{ikt}} = 0$). This suggests that consumers tend to follow those peers who are relatively more active on the platform - both in terms of product consumption and participation in the consumer network - at higher rates. Second, in terms of dyad characteristics, we see that those consumer-peer pairs that actually form a dyad show a considerably lower average preference distance, and higher average numbers of common albums, artists, and ties, compared to those who do not. Hence, consumers tend to follow relatively similar peers at higher rates, which is consistent with homophily in dyad formation and platform nudges in that direction.

=== Table 2 here ===

4.2.2 Conditional Logit Results

The results from our conditional logit regressions are presented in Table 3. To compare the economic significance of each variable, we transformed them into standard deviation units, and we report average marginal effects. Hence, for each variable, the reported coefficient represents a percentage point change in the predicted probability of dyad formation associated with an increase of one standard deviation unit. Model 1 contains only peer characteristics, and we gradually add our consumer-peer similarity measures over Models 2 to 7. In Model 8, the full set of variables enters the regression, and we discuss its results here. First, we turn to the role of peer characteristics, Y_{kt}. We find that a higher number of albums ($\hat{\beta} = 0.0056$, $p < 0.001$) and unique artists ($\hat{\beta} = 0.0127$, $p < 0.001$) are both associated with an increased dyad formation probability. This is consistent with the idea that a larger album collection signals competence and experience, hence a greater benefit from following that peer. In addition, a greater number of unique artists (with the same number of albums) constitutes greater music variety in the peer's collection, indicating that the prospect of observing a broader set of artists with peer purchases increases the likelihood of dyad formation. This is also in line with the negative coefficient on the peer's preference concentration ($\hat{\beta} = -0.1024$, $p < 0.001$), indicating that a more diverse set of horizontal attributes in his album collection is positively associated with dyad formation. In terms of a peer's participation in the consumer network, only the number of his followers is positively related to dyad formation ($\hat{\beta} = 0.1346$, $p < 0.001$), but not how many he is following ($\hat{\beta} = -0.0001$, $p > 0.1$). This is consistent with the notion that a greater number of followers signals a good reputation. Intuitively, it also makes sense that the number of followees does not matter as much – in the end, the consumer is interested in receiving information from the peer, but does not seem to care as much where the peer receives his information from.

=== Table 3 here ===

Second, we turn to the dyad characteristics. We find that a higher preference distance is negatively related to dyad formation ($\hat{\beta} = -0.2303$, $p < 0.001$), and higher number of common albums ($\hat{\beta} = 0.0056$, $p < 0.001$), artists ($\hat{\beta} = 0.1906$, $p < 0.001$), and network ties ($\hat{\beta} = 0.0828$, $p < 0.001$) are each positively related. Together, this indicates that consumer-peer similarities drive the formation of new dyads. This is again consistent with homophily and platform recommendations.

=== Figure 5 here ===

Next, we are interested in the relative economic significance of our peer and dyad variables. Due to our choice set construction and the inclusion of choice fixed effects, it is important to note that the estimated coefficients here only reflect differences between the 51 alternatives in each set. Hence, comparing effect sizes here serves to qualitatively document these *internal* differences. Figure 5 plots the estimated coefficients. First, we see the greatest overall effect sizes for the consumer-peer preference distance ($\hat{\beta} = -.2303$) and the number of common artists ($\hat{\beta} = .1906$). This suggests that similar preferences, both in terms of horizontal attributes and artists, are the strongest predictor of dyad formation. Second, network ties seem to play an important role. Both the coefficients on the number of common ties ($\hat{\beta} = .0828$) and a peer's followers ($\hat{\beta} = .1346$) imply strong relationships with dyad formation. This likely reflects that the platform recommends potential peers to follow based on common ties. In addition, it shows that a greater number of followers signals a good reputation of the potential peer.

Finally, a peer's preference concentration is a strong predictor of dyad formation ($\hat{\beta} = -.1024$), suggesting that a greater diversity of horizontal attributes increase its probability. This adds nuance to the overall picture because it contrasts the notion that preference similarity plays an important role. The former suggests that consumers seek variety in observed peer purchases, and the latter that they seek peer purchases that correspond to their preferences. There are two potential explanations for this tension: For one, the strong influence of consumer-peer similarities may primarily reflect their process of discovering potential peers to follow. Platform recommendations predominantly contain similar peers. Moreover, consumers can discover peers via album or profile pages, and which of those they visit is likely informed by their purchase histories and network ties. Both discovery modes should then nudge consumers towards more similar peers. In contrast, the role of variety in a peer's album collection may then reflect dyad formation preferences *conditional* on peer discovery. That is, out of the relatively similar set of potential peers to follow consumers are exposed to, they prefer those who provide a greater variety.

These findings also contain important implications for new music exploration in the network. We show that consumer-peer similarities are the strongest predictor for dyad formation. This implies that information flows in the network mainly exposes consumers to peer purchases that correspond to their horizontal preferences. As a result, selection dynamics should rather reinforce these preferences and stifle the exploration of unfamiliar horizontal attributes and artists. In contrast, our findings also indicate that consumers are variety-seeking to an extent. Together, there may be a tension between relatively high exposure to similar horizontal attributes and a desire to explore unfamiliar ones.

5 Peer Influence in Consumption and Exploration

In a next step, we study how consumers' ability to observe their peers' purchases affects album consumption and their new music exploration. Therefore, we now shift our focus to what happens *after* a dyad has formed. As such, our arguments going forward should be understood as *conditional* on the determinants of a consumer's choice to follow a particular peer. Here, we are interested in two questions. First, does observing a peer purchasing an album increase the likelihood that the consumer also purchases that particular album? As we point out in detail below, within each consumer-peer dyad, the key variation we explore is across peer purchases, some of which the consumer can observe (peer influence should play a role), and some of which she cannot. And second, how does this effect differ between albums that constitute an exploration opportunity (i.e. that exhibit unfamiliar horizontal attributes or are from unfamiliar artists) and those that do not.

5.1 Conceptual Framework

Following another enables a consumer to observe his purchases. We expect this to have two effects on her consumption behavior. First, it has an *awareness effect*: It may inform her that the album exists in the first place. Second, and conditional on learning of an album's existence, it has a *signaling effect*: The fact that her peer (that she selected into following) purchased the album conveys information about its quality. We develop a simple conceptual framework to derive predictions about how the awareness and signaling effects relate to album consumption.

First, we analyze how it varies between peer purchases she can observe and those she cannot, i.e. the *main effect* of peer influence in consumption. Second, we analyze heterogeneity in this effect between exploration opportunities and other albums.

5.1.1 Album Discovery and Awareness Effect

Before making a purchase decision, consumers have to first learn about an album's existence. Consumers have many opportunities to learn of the existence of an album other than the social network on the platform. They can get this information from genre magazines and websites, going to concerts, other streaming platforms (e.g. Spotify, Apple Music), or advertising. In addition, they can simply browse Bandcamp (which would incur search cost). Hence, it is possible that a consumer learns about an album's existence through such *private discovery channels*. Here, we therefore assum that, the private discovery – i.e. without observing a peer purchase – of album j by consumer i happens with probability $\pi_{ij} > 0$.

The social network simply adds another discovery channel for the consumer. If a peer she follows purchases the album, the platform uses several means to inform her about its existence. Hence, the consumer observes the peer purchase of album j with probability $\rho_{ij} > 0$ and learns of its existence. Note that this happens with a positive probability because it is of course possible that the consumer ignores or misses the peer purchase. Then, the combined probability of consumer i discovering album j is $\rho_{ij} + (1 - \rho_{ij})\pi_{ij}$. The *awareness effect* simply captures the difference between relying on private discovery only and having the additional channel through the network, $\rho_{ij} + (1 - \rho_{ij})\pi_{ij} - \pi_{ij}$. It is therefore given by

$$\text{Awareness Effect} = (1 - \pi_{ij})\rho_{ij} > 0. \tag{7}$$

The effect is increasing in ρ_{ij} and decreasing in π_{ij}. Intuitively, this means that peer purchases become *less* important for discovery, the *more* effective a consumer's private channels are. Intuitively, this means that information flows through the network are relatively more useful for albums that the consumer does not discover easily on her own. Arguably, this is the case for those that constitute an exploration opportunity. Consider an album that exhibits attributes the consumer is familiar with. It is either produced by an artist she already knows, or it may be part of a genre she has listened to in the past. Therefore, her private discovery channels should be more effective: She is more likely to read magazines or websites that include coverage of the album, streaming sites are more likely to include it in listening recommendations to her, and the artist may inform her about the album release directly, e.g. in the form of a newsletter or via social media. In contrast, the same private discovery channels are likely not in place for albums that constitute an exploration opportunity. Consequently, π_{ij} should increase in the consumer's familiarity with the album's attributes. As a result, information flows about albums that constitute an exploration opportunity are more useful for the consumer and the awareness effect should be stronger.

5.1.2 Purchase Decision and Signaling Effect

After becoming aware of an album's existence, the consumer decides whether or not to purchase it. First, consider an album the she discovers through her private discovery channels, i.e. without observing a peer purchase. As an experience good, the utility from doing so is uncertain ex ante and can be expressed as

$$u_{ij} = f(\mathbf{X}_j, m_{ij}) + \epsilon_{ij}, \tag{8}$$

where $\mathbf{X}_j$ is a vector of album j's observable characteristics (e.g. its length, number of tracks, and sales up until this point), and m_{ij} captures consumer i's familiarity with its horizontal attributes, or its match with her preferences. The term $\epsilon_{ij} \sim N(0, \frac{1}{m_{ij}})$ is a random utility shock that is drawn from a distribution with a mean of zero and a variance that depends on m_{ij}. This captures the ex ante uncertainty about consumption utility, which we assume to depend on i's familiarity with the album's horizontal attributes. The intuition is that she can form a more accurate prior about the utility she receives for albums that correspond to her preferences. For example, being knowledgeable in a certain genre should enable her to more easily discern the album's quality. The decision to purchase the album is then based on her prior, $\mathbb{E}(u_{ij}) = f(\mathbf{X}_j, m_{ij})$, and she will do so if it exceeds the album price, $\mathbb{E}(u_{ij}) > p_j$.

Now consider the case where she observes a peer purchase of that album through the network. In addition to its observable characteristics, $\mathbf{X}_j$, and the consumer's familiarity with its horizontal attributes, m_{ij}, this sends a signal about its quality. After all, her peer has purchased the album. Hence, we assume that the consumer attaches a higher expected quality to the album. Hence, the signal can be expressed as

$$S_{ijk} = f(\mathbf{X}_j, m_{ij}) + \Delta + \mu_{ijk}, \tag{9}$$

with Δ representing the increase in the expected utility. Similar to equation 8, $\mu_{ijk} \sim N(0, \frac{1}{\tau_{ik}})$ is a random variable, which captures the precision of the signal. It depends on τ_{ik} which can be interpreted as the trust consumer i puts into peer k's judgment. Intuitively, a larger τ_{ik} means that the consumer is more confident that the signal is an accurate representation of the utility from consuming j.

Having observed the peer purchase of j, consumer i's decision is now based on her posterior:

$$\mathbb{E}(u_{ij}|S_{ijk}) = \omega_m \cdot f(\mathbf{X}_j, m_{ij}) + \omega_\tau \cdot [f(\mathbf{X}_j, m_{ij}) + \Delta], \tag{10}$$

with $\omega_m = \frac{m_{ij}}{m_{ij}+\tau_{ik}}$ and $\omega_\tau = \frac{\tau_{ik}}{m_{ij}+\tau_{ik}}$ representing the relative weights the attaches to her prior and the signal, respectively. This means that the relative relevance of the two is determined by her familiarity with the album's horizontal attributes, m_{ij}, and the trust she puts into the peer's judgment, τ_{ik}. The higher m_{ij} she more she relies on her own prior, and the higher τ_{ik} the more she relies on the signal. The *signaling effect* of observing peer purchases is determined by the difference between the posterior and the prior, $\mathbb{E}(u_{ij}|S_{ijk}) - \mathbb{E}(u_{ij})$, and is given by

$$\text{Signaling Effect} = \omega_\tau \Delta > 0. \tag{11}$$

The effect increases in both ω_τ and Δ. The higher the relative trust in the peer's judgment and the greater the increase in expected utility, the stronger the signaling effect. Again, this bears implication for heterogeneity between albums that present an exploration opportunity for the consumer and those that do not. With $\omega_\tau = \frac{\tau_{ik}}{m_{ij}+\tau_{ik}}$ it follows that $\frac{\partial \omega_\tau \Delta}{\partial m_{ij}} < 0$. Hence, the signaling effect is decreasing in the consumer's familiarity with the album's horizontal attributes. The intuition is simple: For albums that exhibit familiar attributes or that are from familiar artists, she can more easily discern the quality on her own. This means that the information conveyed through the network does not add much insight above her own judgment. In contrast, the consumer has to rely more on the signal with albums that constitute an exploration opportunity. As she is unable to discern their quality easily, she has to rely more on the judgment of her peer. Hence, the signaling effect should be stronger for albums with unfamiliar attributes.

5.2 Empirical Framework

The aim of our empirical analysis is two-fold: First, we intend to provide a causal estimate of peer influence in the network on Bandcamp, i.e. how observing peer purchases affects consumers' decisions to also purchase a particular album. We call this the *main peer effect.* Second, to study how this affects consumers' exploration of unfamiliar music we then explore heterogeneity in the main peer effect by products that present such an exploration opportunity compared to those that do not. Recall that we consider albums that are from unfamiliar artists or that exhibit unfamiliar horizontal attributes as such.

5.2.1 Empirical Challenges

Peer effects are notoriously difficult to estimate (Angrist, 2014). In our study, we face three key challenges. First, we seek to study how information flowing from a peer to a consumer influences the behavior of the latter. However, the direction of influence is not necessarily clear in networks such as the one we study. As a result, the consumer may be influencing the peer's behavior instead, driving the correlations in our analysis - an issue known as the "reflection problem" (Manski, 1993). We address this issue by leveraging the unidirectional nature of ties in the network – consumers *follow* a peer, as opposed to becoming *friends* on the platform. Still, many peers in our data follow the consumer back, establishing a reciprocal link with information flowing in both directions. To remedy this, we exclude all of these reciprocal ties from our analysis and focus on those that are unidirectional.

Second, in section 4 we show that consumer-peer similarities drive the formation of new dyads, hence homophily may bias our results if not addressed (Wang et al., 2018). In particular, two issues arise. The endogenous nature of dyad formation means that the choice of following a peer may reflect past and subsequent consumption choices on the platform. This can entail unobserved between-dyad differences and bias our results if not addressed. In addition, similarities in purchase behavior may be the result of consumer-peer similarities in preferences, even in the absence of peer influence – they may simply like the same music. This would lead to an overestimation of peer effects. We address these issues in two ways. We use a dyad-level approach which exploits the panel structure and detailed time-varying information of our data: our data contain the exact sequence of purchases by consumers and their peers, as well as the time of the dyad formation. Details on our identification strategy follow below. In addition, we use dyad fixed effects throughout, i.e. our estimates do not reflect differences between dyads. Essentially, all of our estimates can be understood as conditional on the reason why a consumer follows a certain peer. Most importantly time invariant common preferences between the consumer and the peer are averaged out.

Third, we study whether or not a focal consumer purchases a certain album after their peer purchased it. The issue is that both doing so not necessarily reflects peer influence, but may also be driven by album characteristics. For instance, it may be the result of its high quality. To address this, we include album fixed effects throughout, controlling for such unobserved confounders.

5.2.2 Identification Strategy

We begin by explaining how we identify the main peer effect of observing peer purchases. The unit of analysis in our framework is the consumer-peer dyad. We are interested in how the information flowing from a focal peer to a focal consumer affects the latter's behavior. Specifically, we analyze how this affects the probability that a consumer purchases a certain album after her peer did. We call this a *follow purchase*. Hence, the unit of observation is each purchase made by the peer in the dyad. The dependent variable is a dummy that indicates if a follow purchase took place or not. The main independent variable of interest is a dummy indicating whether or not the consumer is able to observe her peer's purchases. Ideally, we would run an experiment where they are randomly hidden or shown to the consumer. Absent such a random intervention, for us the closest feasible approximation is to exploit the detailed time-varying information in our data. In particular, we can distinguish between peer purchases that happened before the dyad is formed and those that happen afterwards. Hence, we capture the consumer's ability to observe her peer's purchases via a dummy indicating those that happen after the dyad has been formed. Figure 6 shows an illustrative example of this research design.

=== Figure 6 here ===

This design serves to alleviate issues arising from homophily that would otherwise bias our results. The idea is the following: In the period before the dyad is formed, the consumer cannot yet observe peer purchases. Of course, it is still possible that the consumer purchases the same album as her peer. However, in the absence of information flows between them, the consumer discovered the album through her private discovery channels and the follow purchase can be attributed to common preferences. In contrast, in the after period, the consumer is able to observe peer purchases. Hence, when we detect a follow purchase here, it is the result of a combination of common preferences *and* causal peer influence. To then back out the peer effect we estimate the difference in probabilities of a follow purchase between the before and after periods. In addition, we add dyad fixed effects throughout to control for between-dyad differences. Hence, the variation we explore stems from within each dyad, but at different points in time. A similar approach is used by Wang et al. (2018), and the estimate represents a causal effect of observing a peer purchase (the peer effect) that sits on top of common preferences.

In addition, we investigate how the main effect varies by albums that present an exploration opportunity compared to those that do not. We use two variables to indicate this. To capture the exploration of unfamiliar horizontal product attributes, we construct the distance between the consumer's preference vector at that time and the attributes vector of the album purchased by the peer. Hence, using equations 1, 2, and 4 we calculate the preference distance between consumer i and album j at time t as

$$\text{PreferenceDistance}_{\text{ijt}} = d(\mathbf{P_{it}}, \mathbf{C_j}).$$

This is measure between zero and one, with higher values indicating a higher degree of unfamiliarity with the horizontal preferences.

In addition, we construct a dummy indicating whether or not the consumer had purchased an album from the artist before ($\text{NewArtist}_{\text{ijt}}$). Hence, it captures the exploration of new artists. This is related to but distinct from the preference distance as such an album may still exhibit familiar horizontal attributes. Hence, the two measures are complementary in how they represent the exploration of new music.

Note that our identification strategy relies on two assumptions. First, we assume that the purchase behavior of the peer does not vary between the before and after periods, which could otherwise drive our results. Second, in principle it is possible that the consumer and peer develop more similar preferences over time, which could both spark the formation of the dyad and would bias our estimates. In this case, what we detect as exploration would simply reflect consumers' preference evolution that happens without peer influence. However, our data suggest that both aspects are not an issue in our regression samples, and we provide supportive evidence in appendix A.1.

5.3 Sample and Estimation

Our sample contains 27,951,139 peer purchases in 848,397 dyads. In its construction, we only consider peer purchases that take place in the three months before and after the formation of the dyad. The key variation we explore happens over time, and a longer time window would contain increasingly conflating factors.[16] In addition, we implement a "donut" around the dyad formation date by excluding peer purchases that happen in the three days before and after the dyad formation. The reason is that consumers can discover peers via album pages. As a result, it is likely that a consumer purchases an album, then directly searches for others who also bought it on the album page (see Figure 2). Hence, this follow purchase is then likely the event that causes the dyad formation. By implementing a donut, we avoid including such chains of events in our sample. We also exclude those purchases where the consumer bought the album before the peer, which makes a follow purchases impossible. Moreover, we exclude dyads that are either

[16] Using one, six, or twelve months before and after produces similar results.

outliers in the number of peer or consumer purchases, or for which all peer purchases fall either in the before or after period. Lastly, we exclude albums that are sold at a very high price (above the 99$^{\text{th}}$ percentile).[17]

We add a range of control variables throughout. We control for the total number of peers the focal consumer follows (we use its logarithm, $\text{Followees}_{\text{it}}$). A higher number means that she is exposed to more peer purchases, which may limit the attention she can pay to individual ones. This would impact the awareness effect, as ρ_{ij} would likely be smaller. We also construct a measure of the consumer's preference concentration ($\text{PreferenceHHI}_{\text{it}}$) as the HHI of the elements in $\mathbf{P_{it}}$. On the one hand, this controls for her preference for variety. On the other, our measure for the preference distance is sensitive to the number of non-zero elements in the vectors, which we address this way. Further, we control for the album sales up until the focal peer purchase (we use its logarithm, $\text{AlbumSales}_{\text{jt}}$).[18] The more popular an album becomes, the more prominently it likely features in private discovery channels, hence impacting the awareness effect through π_{ij}. Lastly, we add a dyad-specific linear time trend throughout to control for unobserved within-dyad trends. Together, we estimate the following model:

$$\begin{aligned}\text{FP}_{\text{ijk}} = \beta_0 + \beta_1\text{After}_{\text{ikt}} + \beta_2(\text{After}_{\text{ikt}} \times \text{PreferenceDistance}_{\text{ijt}}) + \beta_3(\text{After}_{\text{ikt}} \times \text{NewArtist}_{\text{ijt}}) \\ + \beta_4\text{PreferenceDistance}_{\text{ijt}} + \beta_5\text{NewArtist}_{\text{ijt}} + \beta_X\mathbf{X}_t + \alpha_{ik} + \alpha_j + t_{ik} + \epsilon_{ijk}.\end{aligned}$$

Here, FP_{ijk} is a dummy indicating a follow purchase, i.e. that consumer i purchases album j after peer k. $\mathbf{X}_t$ is the vector of time-varying controls, α_{ik} and α_j are dyad and album fixed effects, respectively, t_{ik} is the dyad-specific linear time trend, and ϵ_{ijk} is the econometric error term.

5.4 Main Results

5.4.1 Descriptive Statistics

Table 4 contains summary statistics for our regression sample. We report them for the whole sample as well as for the before and after periods separately. Within our six-month time span, follow-purchases happen for 1.79% of all peer purchases, and the rate is around twice as high after a consumer is able to observe them.[19] The after period represents 50.47% of all observations.

[17] Our results are robust to the inclusion of these outliers.

[18] Popularity information can also be considered an aspect of peer influence. However, we are interested in the dyad-level information flows only.

[19] For reference, some industry experts estimate average click-though rates for search ads of 1.91% and for display ads of 0.35% in the context of Google AdWords (see https://blog.hubspot.com/agency/google-adwords-benchmark-data).

Hence, we have a balanced sample, even though peer purchases naturally happen at their own leisure.

Looking at exploration opportunities, albums from new artists (for the consumer) represent 94.41% of all observations. It does not differ between the before and after periods, which adds credence to our identification strategy – consumer-peer similarity seems to not increase over time. In addition, the high share of albums by new artists suggests that information flows through the network indeed bear the opportunity of new music discovery. Figure 7 plots how share of new and familiar artists between consumers' purchases and those by their peers. This contrasts a consumers' actual consumption patterns with the information they are exposed to through the network. While around half of all albums consumers purchase are from familiar artists, they are much more exposed to peer purchases of new artists albums. This illustrates that the network presents consumers with plenty opportunities to explore new artists.

=== Figure 7 here ===

Next, The consumer-album preference distance exhibits an average of 0.5493 and does not vary between the two periods. Again, figure 8 shows the distribution of the preference distances for those albums consumers purchase on the platform, and those they are exposed to through the network. Consumers mostly purchase albums that completely reflect their preferences, i.e. that exhibit a distance of zero. In contrast, their distance to albums purchased by their peers predominantly lie at the high end of the spectrum, and only very few fully correspond to consumers' preferences. This illustrates that the network exposes them to opportunities to explore new horizontal attributes as well.

=== Figure 8 here ===

Moreover, the average price of the albums in our sample is 4.37 USD with a maximum of 17.98 USD (the 99$^{\text{th}}$ percentile), and 29.03% are offered for free. Again, these are not different between the before and after periods. Finally, concerning the control variables, on average consumers follow 74.98 peers, which is naturally higher in the after than the before period, and albums sold an average of 152.65 copies in our sample.

5.4.2 Estimation Results: Main Peer Effect

As our regressions include two sets of fixed effects, and we are interested in interaction terms in parts of our analysis, we estimate linear probability models throughout. Table 5 contain the results for the main peer effect, i.e. we do not explore heterogeneity at first. Models 1 to 5 contain different combinations of fixed effects, and Model 6 contains the full specification including control variables. The main coefficient of interest, After_{it}, gives the percentage point difference in the probability of a follow purchase between the before and after period. In the full model, the probability is 0.0188 *percentage points* higher after the dyad has been formed. This can be interpreted as the *absolute* increase in the probability of a follow purchase. In addition, we can calculate its *relative* increase using the share of follow purchases in the before-period, which can be understood as the *base rate* at which follow purchases occur. In the full regression sample this base rate is 0.0117. Hence, our estimate corresponds to a *relative* probability increase of 161%.[20] In other words, a consumer is 2.61 times more likely to purchase an album after her peer when she is able to observe that peer purchase (compared to not being able to). Together, these findings provide evidence for the existence of sizable peer influence in the social network on Bandcamp.

=== Table 5 here ===

Further, we are interested in how this pattern unfolds over time. Thus, we run an additional model where we use dummies representing the 13 weeks before and after the dyad formation instead of our after dummy. That is, a week dummy equals one when a peer purchase has taken place in that week. We plot the estimated coefficients in Figure 9. The probability of a follow purchase stays relatively constant in the before period, but increases sharply just after the dyad has been formed. Then, it again steadily decreases. This pattern suggests a rather immediate peer effect – information flows seem most relevant for consumers just after they establish the dyad, but it then dwindles over time.

=== Figure 9 here ===

5.4.3 Estimation Results: Exploration

Table 6 contain results from our analysis of consumers' exploration of new music. Hence, we investigate potential heterogeneity in the main peer effect between albums that constitute an exploration opportunity and those that do not. First, Models 1 to 3 contain the results for albums by new artists. In Model 1 the interaction between a dummy indicating this and the after dummy is negative. The probability increase in the after period is 15.39 percentage points

[20] This is simply calculated via $\frac{0.0188}{0.0117} \times 100 = 161$.

lower for albums from new artists. Hence, the *absolute* increase in the probability of a follow-purchase is lower for new artists. However, this negative relationship can be attributed to the stark differences in the base rates. For albums from new artist the base rate is only 0.28 %. For albums from familiar artists it is much higher, with 14.7 %. This shows that consumers already purchase albums from familiar artists at relatively high rates before they form a dyad with the peer.

Therefore, we are primarily interested in differences between the *relative* change in probabilities. Hence, we perform a sample split in Models 2 and 3. The absolute increase in probability of a follow-purchase is 0.66 percentage points for new artists, and it is 14.43 percentage points for familiar artists. However, taking into account the difference in base rates, the picture is reversed. For new artists we find a relative increase of 234.9 %, and for familiar artists it is 98 %. Together, we find a positive peer effect for albums from both new and artists albums, but it is relatively stronger for new artists. This is in line with expectations: Consumers rely more on information flows through the network in their discovery and subsequent purchase of albums from new artists, i.e. those that constitute an exploration opportunity. In contrast, they purchase albums from familiar artists at high rates even without information from the network, which speaks to the notion of effective private discovery channels being in place.

=== Table 6 here ===

Next, Models 4 to 6 investigate heterogeneity along the preference distance between the consumer and the focal album. As such, it represents the exploration of unfamiliar horizontal attributes, such as its genre. Results show a similar pattern to the new artist analysis. We estimate a negative coefficient for the interaction with the after dummy in model 4 of table 6, hence the percentage point increase in the follow purchase probability is lower for a higher preference distance. Again, this can be partly attributed to base rate differences. Hence, we again perform a sample split in Models 5 and 6. We call observations above the median of the sample distribution a *high* and those below a *low* preference distance (PD).

Base rates again differ between the two: it is 0.583% for albums with a high PD, and 1.73% for albums with a low PD. That is, follow purchases in the before period are less likely for exploration opportunities. In models 5 and 6 of table 6 we find a lower absolute increase in the probability of a follow purchase for albums with a high PD (model 6, 0.76 percentage points) than those with a low PD (model 5, 2.82 percentage points). Here, this means that the increase in the relative probability is also higher for albums with a low PD (162.9 %) than those with a high PD (130 %). This means that the peer effect is *weaker* for albums that constitute an exploration opportunity, both in absolute and relative terms. This contradicts our expectations and contrasts the results concerning familiarity with artists. A possible explanation is that the quality premium a consumer attaches to observed peer purchases of albums with unfamiliar horizontal attributes is not sufficient to offset a potential mismatch with her preferences.

Together, these findings paint a nuanced picture of peer influence in new music exploration. They suggest that consumers are nudged towards exploring new artists, but not necessarily unfamiliar horizontal attributes, such as genres. While follow-purchases increase across the board, the signaling effect may not be as relevant as the awareness effect: Consumers are willing to engage with new artists, but preferably those that correspond to their preferences for stylistic elements, and therefore do not impose a disutility from a mismatch with tastes.

5.5 Additional Analyses

5.5.1 Mechanism Test: Awareness and Signaling Effects, and the Role of Album Price

We[21] posit that observing peer purchases through the network increases the likelihood of a follow purchase though an awareness and a signaling effect. While we are unable to clearly disentangle the two empirically, the album price can serve as an indication for their relative importance. The idea is the following: As the album price is unlikely to affect album discovery, it should not be related to the awareness effect. However, consumers only purchase an album if their expected utility exceeds its price. As the signaling effect should increase a consumer's expected utility, we may expect her reservation price to also increase. Thus, if the effect of observing peer purchases on consumer's purchase probability is greater for more expensive albums, this indicates that the signaling effect matters. If, however, this is not the case, then the awareness effect dominates.

=== Table 7 here ===

Therefore, we run an additional set of regressions where we explore heterogeneity in the main peer effect by album price. Results are presented in Table 7. Models 1 to 3 analyze the role of the price directly. Model 1 contains its interaction with the after dummy, and we split the sample by prices above and below the sample median in Models 2 and 3. We find that the

[21] The appendix contains additional robustness checks and lines of inquiry. In appendix A.2 we explore how different dimensions of heterogeneity may play a role. It adds a few more nuances, but is consistent with our main findings. In appendix A.3 we propose an alternative identification strategy, in which we use a time shift approach to construct control observations. Results are consistent with our main analysis both qualitatively and in terms of effect sizes. Finally, we test the robustness our main results with a battery of different samples, all providing qualitatively consistent results.

absolute percentage point increase in the probability of a follow purchase is greater for lower prices, but that the relative increase is greater for higher prices. Again, this disconnect can be attributed to a higher base rate for lower prices. This suggests that consumers are nudged towards albums with a higher price, speaking to the notion that the signaling effect matters. In Models 4 to 6 we investigate heterogeneity between free and non-free albums, providing similar findings. The relative probability increase is stronger for non-free albums (model 5, 181.6%) than free albums (model 6, 125.2%), suggesting that observing peer purchases nudges them towards paid products, proving additional evidence for a signaling effect.

5.5.2 Interplay of Selection and Peer Effect Strength

In section 4 we show that consumer-peer similarities in preferences for horizontal attributes are the strongest predictor of dyad formation in the network. We argue that should limit consumers' exposure to exploration opportunities. Our data suggest that this is indeed the case. Figure 10 plots the distribution of the consumer-album preference distance for two types of dyads. The left panel shows the distribution for dyads where the consumer and peer exhibit a low (i.e. below the sample median) distance between their preference vectors at the time of the dyad formation. That is, it represents a dyad formation where the two had similar horizontal preferences. The right panel shows dyads, where the distance between their preferences vectors is high (i.e. above the sample median), representing dissimilar horizontal preferences. Comparing the two clearly shows that dyads with higher consumer-peer similarity at the time of dyad formation are connected to less exposure to exploration opportunities through the network.

=== Figure 10 here ===

Next, we may also expect that the consumer-peer similarity at the time of the dyad formation impacts the strength of the peer effects. We therefore run two additional sets of regressions. Table 8 reports the results using a sample of dyads with a high level of consumer-peer similarity (i.e. low consumer-peer preference distance) at the time of its formation. In contrast, Table 9 uses a sample with a low level of consumer-peer similarity. For both types of dyads, the patterns are similar to our previous analyses: (i) we find strong main peer effects, (ii) consumers are nudged towards the exploration of new artists, but not unfamiliar horizontal attributes (e.g. genre), and (iii) they are nudged towards paid albums, as opposed to free ones.

However, we also find that the effect sizes are generally stronger in our sample of more similar dyads. This suggests that consumers are more inclined to purchase an album after observing a purchase by a similar peer, compared to a dissimilar one. There are two potential explanations for this: First, they may attach a higher "premium" (Δ in equation 9) to the expected utility when observing a similar peer. Second, they may put a higher level of confidence in their "judgment" of an albums quality. In other words, as they liked similar music in the first place, the consumer may attribute a higher level of musical competence to a similar peer than a dissimilar one. In any case, this further reinforces the selection dynamics we find in section 4. Not only are consumers exposed to fewer exploration opportunities when forming dyads with similar peers, but they are also more inclined to purchase the same album after observing a peer purchase.

=== Tables 8 and 9 here ===

6 Discussion and Conclusion

We study peer influence in music consumption and exploration in an online social network. In particular, we explore heterogeneity along consumers' experience or familiarity with horizontal product attributes, and discuss implications for the exploration of unfamiliar product types. Using a dyad-level approach, we first analyze how similarities between consumers and potential peers determine the formation of new network ties, and what this implies for new music exploration. Then, *conditional* on the determinants of peer selection, we study how within-dyad information flows affect consumption choices after a dyad has formed. Here, we empirically analyze how the strength of this peer influence varies by the consumer's familiarity with an album's horizontal attributes (artist, genre, and other stylistic elements), i.e. the extend to which it constitutes an exploration opportunity. We argue that consumption choices are affected by an awareness and a signaling effect of observing peer purchases. That is, consumers learn about the existence of an album, and they update their beliefs about its quality. Both depend on the consumer's familiarity with the focal album's horizontal attributes, which in turn determines the potential usefulness or redundancy of the information conveyed through the network, and therefore the strength of peer influence.

Our analysis of the social network on Bandcamp produces three key findings: First, consumer-peer similarities in their preference for horizontal attributes and artists are the strongest predictors for the formation of dyads. This bears important implications for the exploration of new music: Selecting into a dyad determines which types of peer purchases a consumer is exposed to. Our results suggest that consumers predominantly select into information regimes that reinforce *existing* preferences by limiting their exposure to products that represent exploration opportunities (i.e. those that exhibit unfamiliar attributes). However, in addition to these homophilous dynamics we also find evidence for variety-seeking selection behavior: Consumers also seem to prefer to follow peers with more diverse album collections, ceteris paribus. A potential explanation for this tension is that peer discovery mechanisms on the platform nudge consumers towards more similar peers, which may be at odds with their desire for exposure to more diverse product attributes. Hence, elements of platform design may not be perfectly aligned with consumer preferences in the network we study.

Second, conditional on the determinants of peer selection in a particular dyad, we find a strong positive peer effect of observing peer purchases through the network. In particular, we investigate how the ability to do so affects the probability that a consumer purchases the same album as their peer. We find that it is 2.61 times higher when the consumer is able to observe that peer purchase (compared to not being able to). This speaks to the existence of sizable peer influence in the social network we study.

Third, while always positive, we find considerable differences in the strength of this peer effect between albums that constitute exploration opportunities for the consumer and those that do not. Consistent with our conceptual framework, we find a considerably *stronger* effect for albums by unfamiliar artists. In contrast, we find that the effect is *weaker* for albums that exhibit unfamiliar horizontal attributes, such as the genre. This paints a nuanced picture about new music exploration: Consumers seem to discover new artists through the network, but preferably those that correspond to their horizontal preferences towards stylistic attributes. A potential explanation is the relative importance of the awareness and signaling effects of observing peer purchases. Information flows inform consumers about the existence of artists they did not know before (awareness effect). However, the quality premium they attach to the album as a result of the peer purchase (signaling effect) may not be sufficient to offset a potential mismatch with horizontal preferences. Hence, the awareness effect is likely more relevant than the signaling effect in the network we study.[22] [23]

The findings contain practical implications for marketing professionals and managers of digital platforms and online social networks. First, we show that information flows in a network create awareness for products and have a positive effect on subsequent consumption. However, this mode of product discovery may not be efficient: Even though most consumers in the network select into following peers with similar preferences, most peer purchases they then observe are

[22] Our analysis of album price in section 5.5.1 provides evidence that there exists a signaling effect in addition to an awareness effect. Still, we interpret our findings about heterogeneity along a consumers' familiarity with an album's attributes as evidence that the latter is more important than the former.

[23] Our finding that the peer effect is stronger in dyads between consumers and peers with similar preferences (see section 5.5.2) adds further support for that notion: It implies that consumers expect to like albums more if they discover them through ties where the preference mismatch is likely less severe.

of albums that exhibit unfamiliar horizontal attributes (see Figure 8). Thus, the social network disproportionately exposes consumers to information flows which are less likely to spark follow purchases. In turn, this implies that most information obtained through the network may not be very useful for consumers. Second, our findings suggest that peer influence is strongest for products by unfamiliar artists that exhibit familiar horizontal attributes, which implies that information flows in the network mainly raise awareness for these albums. However, this can also be achieved via relatively simple recommendation engines that draw on consumers' purchase histories and then promote artists that they have not yet explored, but that exhibit similar genre tags. In that sense, our findings imply that social networks on e-commerce platforms act as substitutes rather than complements for algorithmic recommendations, but without the same degree of control over desirable distributional shifts in sales on the platform as a whole.

We make contributions to the literature on peer influence in online social networks. First, we highlight consumers' past experience and familiarity with certain (horizontal) product attributes as a determinant of the strength of peer influence. We argue and show that this determines the usefulness of the information flowing through the network, and in particular provide a nuanced picture of how this affects the exploration of new or unfamiliar product types. Therefore, we contribute to the discussion about heterogeneity in peer influence both by product type (Qiu et al., 2021; Tucker & Zhang, 2011; Wang et al., 2013; Zhu & Zhang, 2010) and consumer characteristics (Aral & Walker, 2012; Tucker, 2008). In addition, to the best of our knowledge, this is the first study to consider new product exploration as the main outcome of interest. Second, we discuss and provide evidence about how the stage of peer selection already determines the exploration of new music via the type of peer purchases consumers are exposed to. This creates new insights to the discussion about how network evolution dynamics relate to consumption patters on digital platforms (Ameri et al., 2022). In addition, we demonstrate differences in the peer effect strength (after tie formation) depending on the similarity between the consumer and peer in a dyad. Hence, peer selection not only affects what information flows consumers are exposed to, but also how useful they perceive them to be. Third, we provide insights about the mechanisms that underlie peer influence in online social networks. While the idea of awareness and signaling effects of observing peer behavior is not new, we do provide empirical evidence about their relative importance in driving consumption choices. Finally, we also contribute to the literature about music consumption in the digital economy by highlighting online social networks as a channel for discovery. In line with prior work we regard this as a reduction in exploration costs (Bronnenberg, 2015). The analysis by Datta et al. (2018) is closest in terms of the outcomes under study. Like them, we investigate implications for new music exploration at the individual level. It also complements work that uncovers a differential impact of digital technologies between broad- and narrow-appeal music (Aguiar & Waldfogel, 2016, 2021; Kretschmer & Peukert, 2020). In particular, we show that consumers are reluctant to deviate from their preferences, which suggests limits to the benefit for niche products.

Our study contains several limitations. First, we study music "aficionados" on a platform that put into place a social network with the aim of facilitating exploration. This may limit the external validity of our findings. For instance, users of larger streaming platforms may exhibit a lower disposition towards exploration than the consumers under study here. Hence, future research should provide additional insights from different empirical contexts, ideally with a set of consumers that more accurately reflect "mainstream" consumption behavior. Second, due to our empirical design we only study consumers that participate in the social network. This presents an additional limitation of the external validity, as they likely represent a group of consumers that have a disposition towards exploring new music. Third, we study cultural products, which are horizontally differentiated and experience goods. Hence, our findings are unlikely to apply to other product types that do not share these characteristics. Fourth, we are only able to show suggestive evidence about the mechanisms (awareness and signaling effects). Future research may make this distinction explicit, perhaps in the form of an experiment. And finally, we rely on observational data in our estimation of peer effects. While we are confident in our empirical framework, it is still important to consider that we do not strictly separate peer influence from confounding factors such as common preferences. Instead, similar to Wang et al. (2018) we identify peer influence that sits "on top" of such time-invariant confounders. In addition, in our analysis of peer selection we neither attempt nor claim to uncover causal effects. Rather, we aim to document and uncover patterns that are relevant for peer influence in music exploration in an more description manner.

References

Aguiar, L. & Waldfogel, J. (2016). Even the Losers Get Lucky Sometimes: New Products and the Evolution of Music Quality Since Napster. *Information Economics and Policy*, 34, 1–15.

Aguiar, L. & Waldfogel, J. (2018). Quality Predictability and the Welfare Benefits from New Products: Evidence from the Digitization of Recorded Music. *Journal of Political Economy*, 126(2), 492–524.

Aguiar, L. & Waldfogel, J. (2021). Platforms, power, and promotion: Evidence from spotify playlists. *The Journal of Industrial Economics*, 69(3), 653–691.

Ameri, M., Honka, E., & Xie, Y. (2019). Word of mouth, observed adoptions, and anime-watching decisions: The role of the personal vs. the community network. *Marketing Science*, 38(4), 567–583.

Ameri, M., Honka, E., & Xie, Y. (2022). From strangers to friends: Tie formations and online activities in an evolving social network. *Journal of Marketing Research*.

Anderson, S. P. & de Palma, A. (2012). Competition for Attention in the Information (Overload) Age. *RAND Journal of Economics*, 43(1), 1–25.

Angrist, J. D. (2014). The perils of peer effects. *Labour Economics*, 30, 98–108.

Aral, S. & Walker, D. (2011). Creating Social Contagion through Viral Product Design: A Randomized Trial of Peer Influence in Networks. *Management Science*, 57(9), 1623–1639.

Aral, S. & Walker, D. (2012). Identifying Influential and Susceptible Members of Social Networks. *Science*, 337(6092), 337–341.

Arndt, J. (1967). Role of product-related conversations in the diffusion of a new product. *Journal of Marketing Research*, 4(3), 291–295.

Banerjee, A. V. (1992). A simple model of herd behavior. *The Quarterly Journal of Economics*, 107(3), 797–817.

Bapna, R. & Umyarov, A. (2015). Do Your Online Friends Make You Pay? A Randomized Field Experiment on Peer Influence in Online Social Networks. *Management Science*, 61(8), 1902–1920.

Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science*, 26(10), 1531–1542.

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5), 992–1026.

Bronnenberg, B. J. (2015). The Provision of Convenience and Variety by the Market. *RAND Journal of Economics*, 46(3), 480–498.

Brynjolfsson, E., Hu, Y., & Smith, M. D. (2010). Research Commentary — Long Tails vs. Superstars: The Effect of Information Technology on Product Variety and Sales Concentration Patterns. *Information Systems Research*, 21(4), 736–747.

Cai, H., Chen, Y., & Fang, H. (2009). Observational Learning: Evidence from a Randomized Natural Field Experiment. *American Economic Review*, 99(3), 864–882.

Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal of Communication*, 64(2), 317–332.

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.

De Matos, M. G., Ferreira, P., & Krackhardt, D. (2014). Peer Influence in the Diffusion of iPhone 3G over a Large Social Network. *MIS Quarterly*, 38(4), 1103–1134.

Dewan, S., Ho, Y.-J., & Ramaprasad, J. (2017). Popularity or proximity: Characterizing the nature of social influence in an online music community. *Information Systems Research*, 28(1), 117–136.

Dewan, S. & Ramaprasad, J. (2014). Social media, traditional media, and music sales. *MIS Quarterly*, 38(1), 101–122.

Dinerstein, M., Einav, L., Levin, J., & Sundaresan, N. (2018). Consumer Price Search and Platform Design in Internet Commerce. *American Economic Review*, 108(7), 1820–1859.

Dubois, E. & Blank, G. (2018). The echo chamber is overstated: The moderating effect of political interest and diverse media. *Information, Communication & Society*, 21(5), 729–745.

Hu, M. M., Yang, S., & Xu, D. Y. (2019). Understanding the social learning effect in contagious switching behavior. *Management Science*, 65(10), 4771–4794.

Katz, E. & Lazarsfeld, P. F. (2017). *Personal influence: The part played by people in the flow of mass communications*. Routledge.

Katz, M. L. & Shapiro, C. (1994). Systems Competition and Network Effects. *Journal of*

Economic Perspectives, 8(2), 93–115.
Kovacs, B. & Hannan, M. (2015). Conceptual Spaces and the Consequences of Category Spanning. *Sociological Science*, 2, 252–286.
Kretschmer, T. & Peukert, C. (2020). Video Killed the Radio Star? Online Music Videos and Recorded Music Sales. *Information Systems Research*, 31(3), 776–800.
Leung, M. D. (2014). Dilettante or Renaissance Person? How the Order of Job Experiences Affects Hiring in an External Labor Market. *American Sociological Review*, 79(1), 136–158.
Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60(3), 531–542.
Moretti, E. (2011). Social Learning and Peer Effects in Consumption: Evidence from Movie Sales. *Review of Economic Studies*, 78(1), 356–393.
Nicolle, A. (2020). Are Consumers Myopic? Evidence from Handset and Mobile Services Choices. *Working Paper*.
Qiu, L., Chhikara, A., & Vakharia, A. (2021). Multidimensional observational learning in social networks: Theory and experimental evidence. *Information Systems Research*, 32(3), 876–894.
Qiu, L., Shi, Z., & Whinston, A. B. (2018). Learning from your friends' check-ins: An empirical study of location-based social networks. *Information Systems Research*, 29(4), 1044–1061.
Salganik, M. J., Dodds, P. S., & Watts, D. J. (2006). Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market. *Science*, 311, 854–856.
Sidorov, G., Gelbukh, A., Gomez-Adorno, H., & Pinto, D. (2014). Soft similarity and soft cosine measure: Similarity of features in vector space model. *Computacion y Sistemas*, 18(3), 491–504.
Train, K. E. (2009). *Discrete Choice Methods with Simulation*. Cambridge university press.
Train, K. E., McFadden, D. L., & Ben-Akiva, M. (1987). The demand for local telephone service: A fully discrete model of residential calling patterns and service choices. *The RAND Journal of Economics*, (pp. 109–123).
Tucker, C. (2008). Identifying formal and informal influence in technology adoption with network externalities. *Management Science*, 54(12), 2024–2038.
Tucker, C. & Zhang, J. (2011). How does popularity information affect choices? A field experiment. *Management Science*, 57(5), 828–842.
Wang, C., Zhang, X., & Hann, I.-H. (2018). Socially Nudged: A Quasi-Experimental Study of Friends' Social Influence in Online Product Ratings. *Information Systems Research*, 29(3), 641–655.
Wang, J., Aribarg, A., & Atchadé, Y. F. (2013). Modeling choice interdependence in a social network. *Marketing Science*, 32(6), 977–997.
Zhang, B., Pavlou, P. A., & Krishnan, R. (2018). On direct vs. indirect peer influence in large social networks. *Information Systems Research*, 29(2), 292–314.
Zhang, J. (2010). The Sound of Silence: Observational Learning in the U.S. Kidney Market. *Marketing Science*, 29(2), 315–335.
Zhang, Y. & Godes, D. (2018). Learning from online social ties. *Marketing Science*, 37(3), 425–444.
Zhu, F. & Zhang, X. (2010). Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics. *Journal of Marketing*, 74(2), 133–148.

Figures

Figure 1 Consumer profile including information about peers they follow

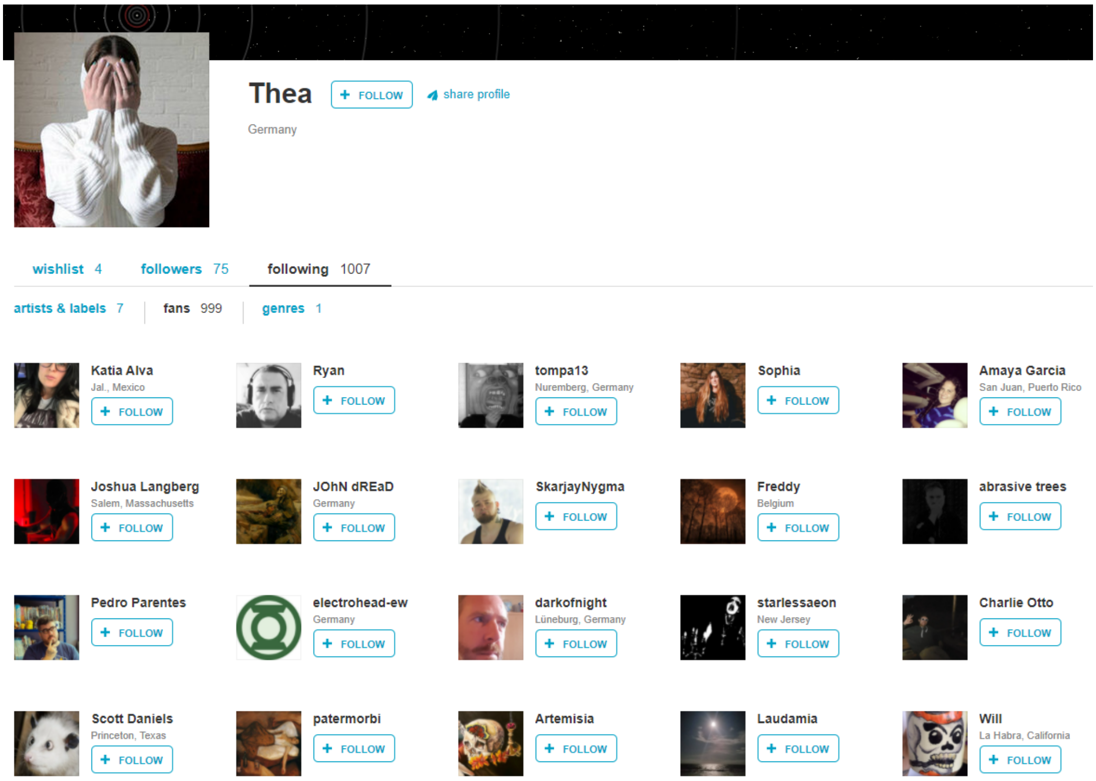


Figure 2 Album page including "supported by" section

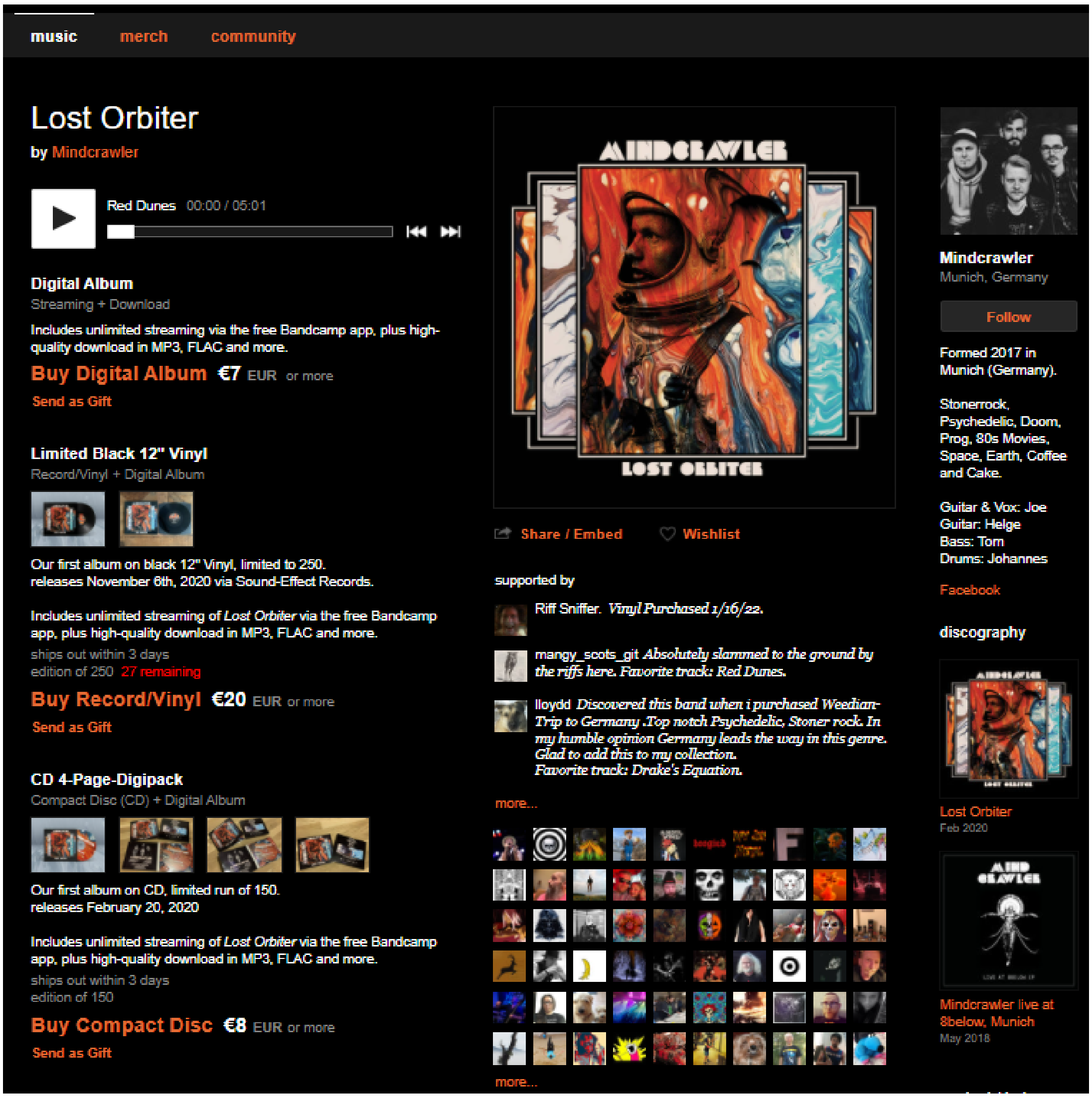


Figure 3 Fan feed including peer purchase information

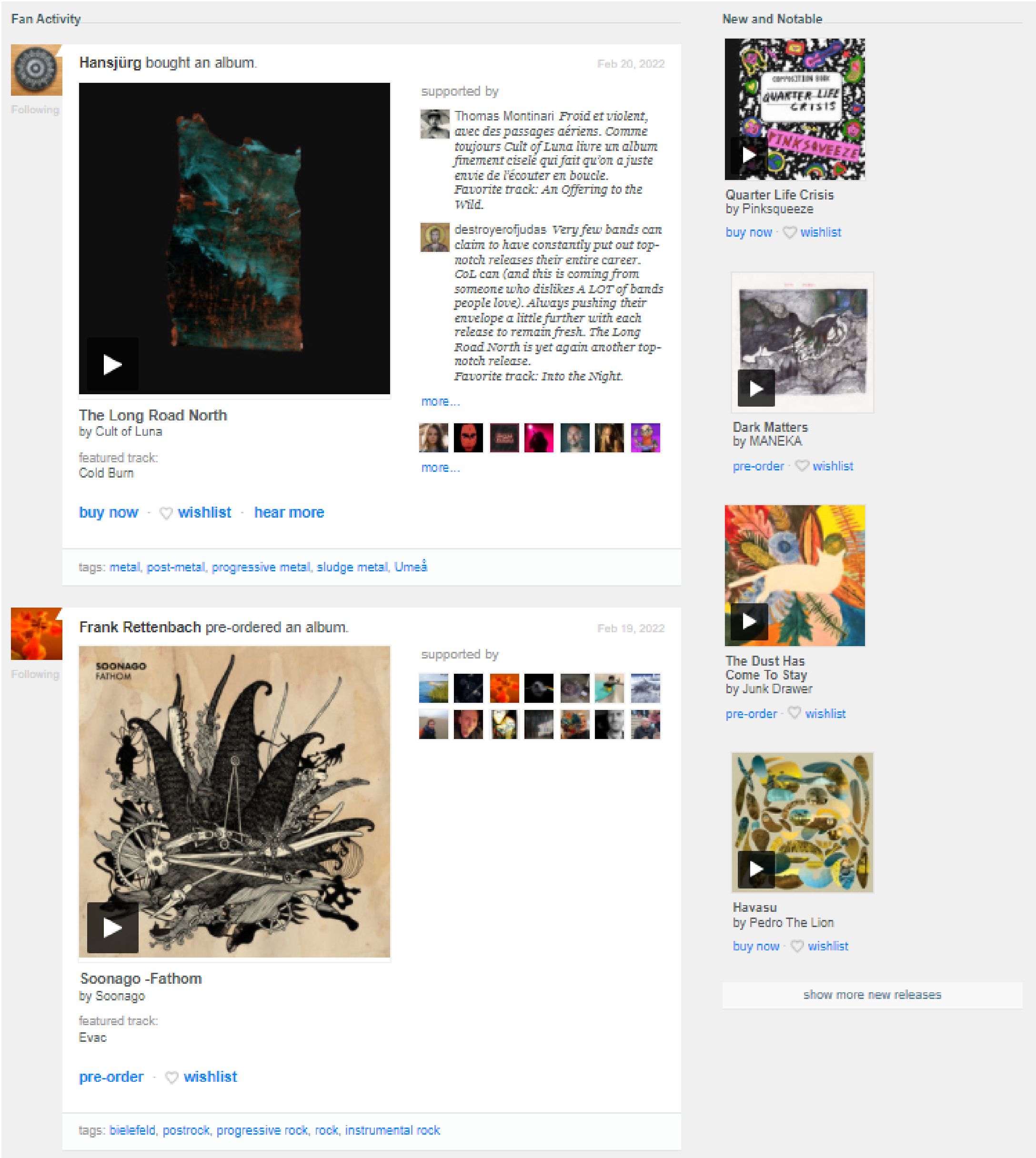
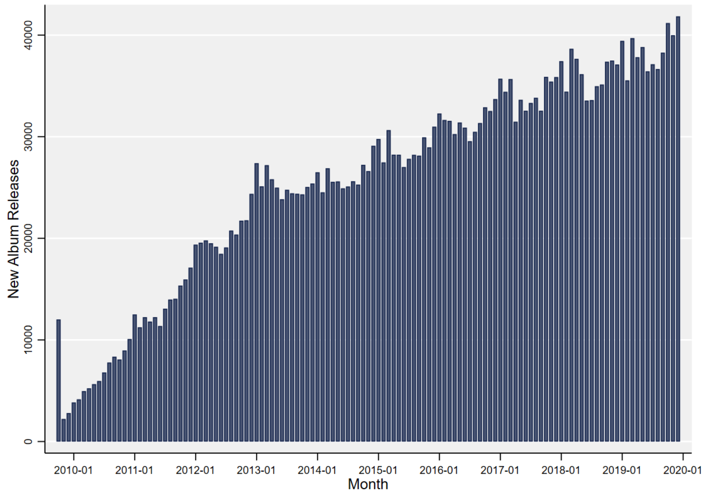
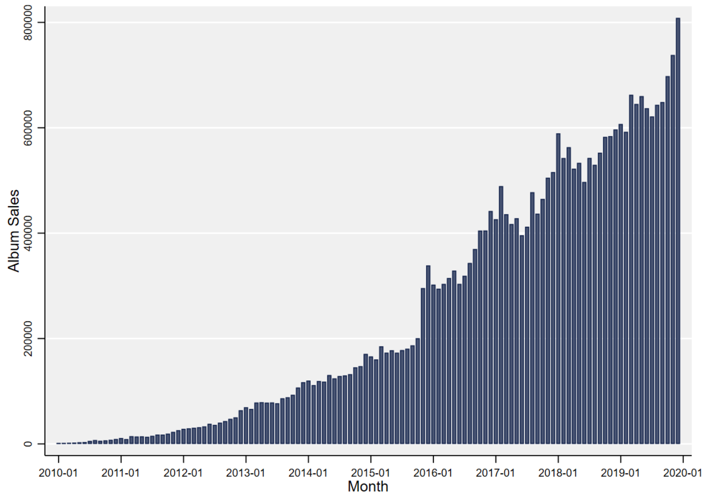


Figure 4 Platform Developments over Time

(a) Album Releases



(b) Album Sales



(c) New Network Ties

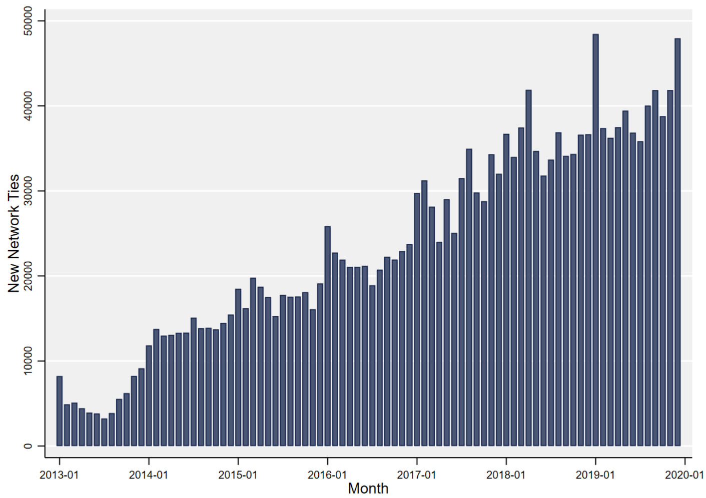


Figure 5 Peer Selection: Estimated Coefficients

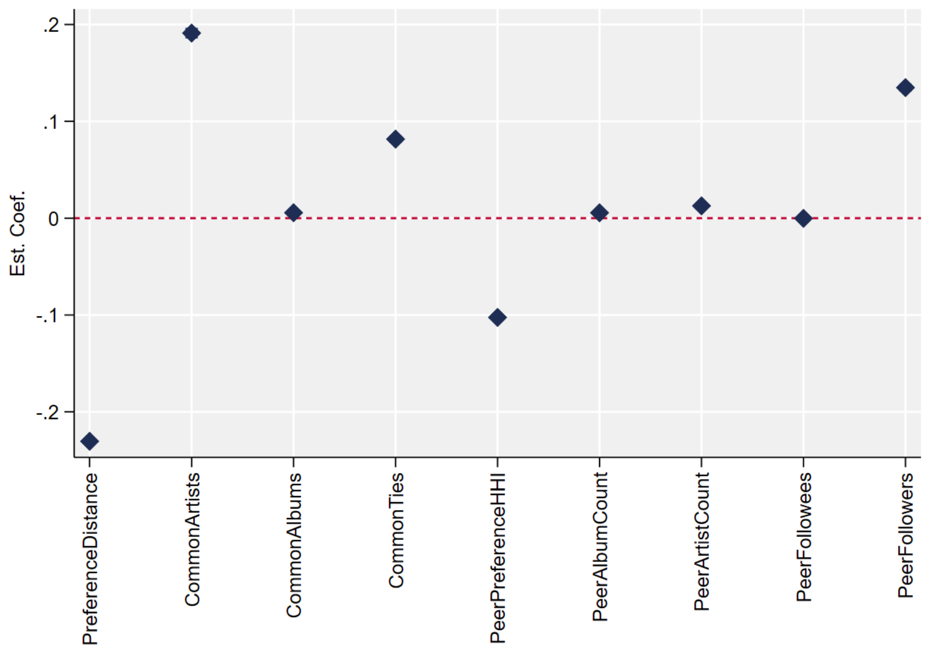
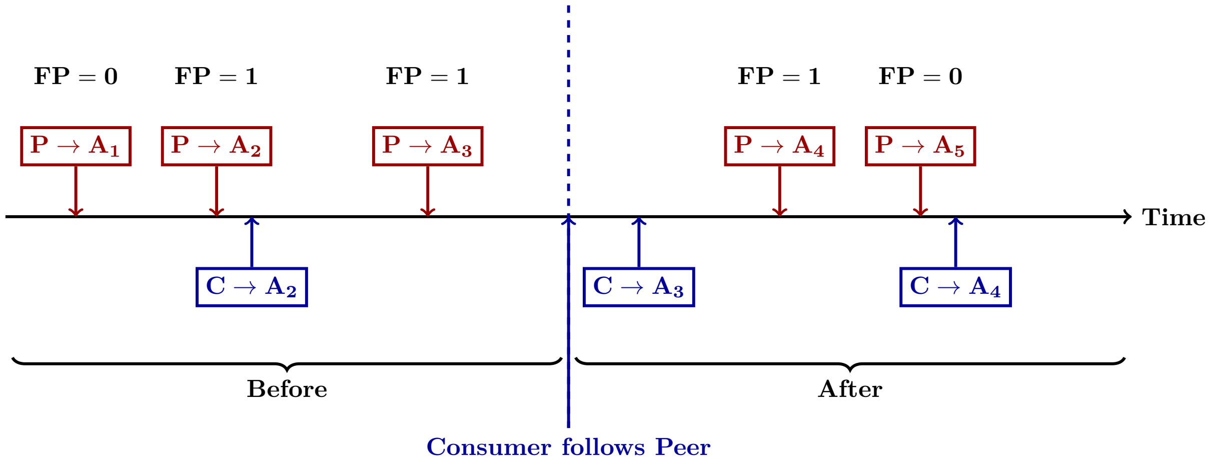


Figure 6 Identification Strategy



Notes: We show an illustrative example of our identification strategy. $\mathbf{P} \rightarrow \mathbf{A_j}$ indicate peer purchases, the unit of observation. $\mathbf{C} \rightarrow \mathbf{A_j}$ indicate consumer purchases used to determine the outcome of interest, namely whether or not she purchased an album after her peer did. In this example, this is the case for albums $\mathbf{A_2}$, $\mathbf{A_3}$, and $\mathbf{A_4}$, hence the dummy indicating n follow purchase is one. In addition, the follow purchases of $\mathbf{A_3}$ and $\mathbf{A_4}$ take place after the dyad has been established, hence when the consumer is able to observe peer purchases.

Figure 7 Distribution of Purchases from New Artists: Consumers vs. Peers

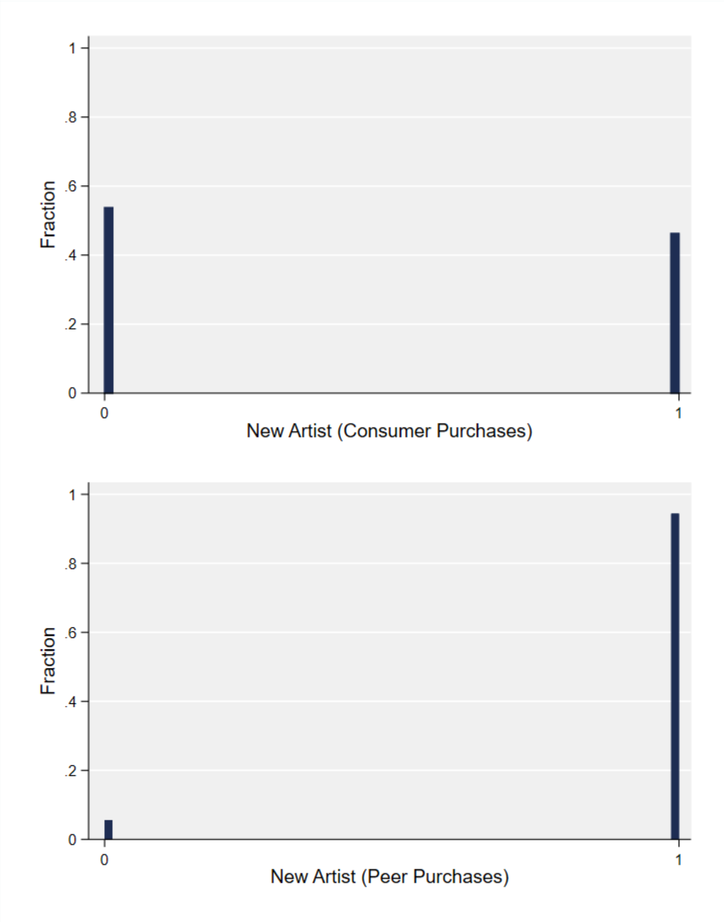


Figure 8 Distribution of Preferences Difference of Purchased Albums: Consumers vs. Peers

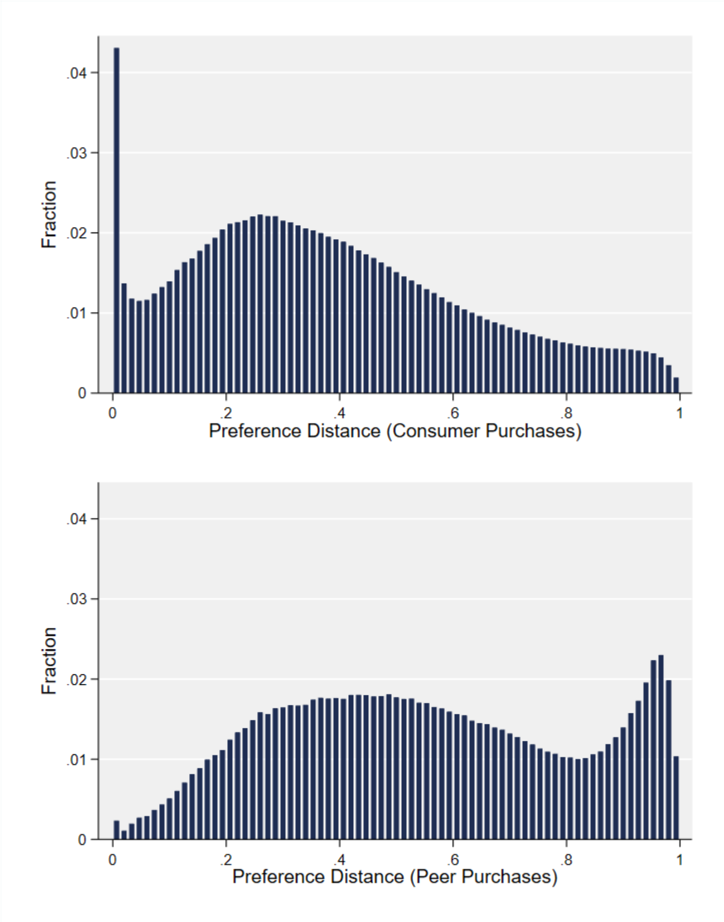


Figure 9 Peer Influence: Estimated Coefficients by Week around Follow

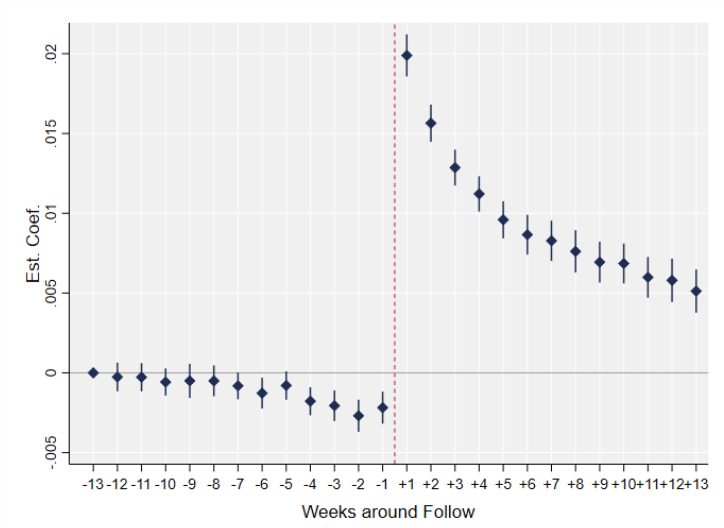
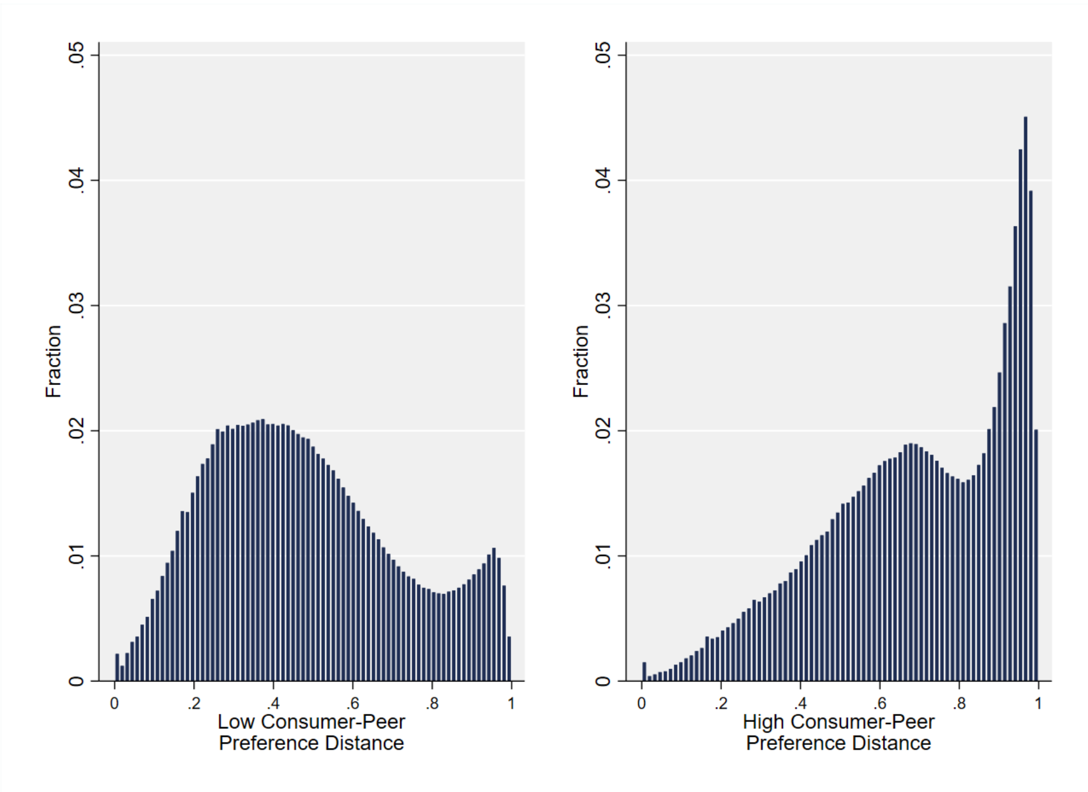


Figure 10 Preference Distance Distribution by Consumer-Peer Similarity at Dyad Formation



Tables

Table 1 Album Characteristics and Consumer Preferences

	Purchases over time		
	1	**2**	**3**
$\mathbf{C_j}$	(1 0 0)	(1 1 1)	(1 0 1)
$\mathbf{C_{it}}$	(1 0 0)	(2 1 1)	(3 1 2)
$\mathbf{P_{it}}$	(1.00 0.00 0.00)	(0.50 0.25 0.25)	(0.50 0.17 0.33)

Notes: We show an illustrative example of how we operationalize of consumer preferences over time. We use vectors with three dimensions here for the sake of presentation. The first row shows the attributes vector associated with each album the consumer purchases. The second row shows the sum of these vectors over time. The third row shows the consumer's preference vector, expressing her "taste" for a certain horizontal attribute in shares.

Table 2 Peer Selection: Summary Statistics

Variable	$\text{IsDyad}_{\text{ikt}}$	Obs	Mean	Std. dev.	Min	Max
Dependent Variable						
$\text{IsDyad}_{\text{ikt}}$		59,704,531	0.0198	0.14	0	1
Consumer Characteristics ($\mathbf{X}_{it}$)						
$\text{PreferenceHHI}_{\text{it}}$		59,704,531	0.15	0.15	0.02	1
$\text{AlbumCount}_{\text{it}}$		59,704,531	100.74	333.23	1	13988
$\text{ArtistCount}_{\text{it}}$		59,704,531	43.78	103.07	1	3015
$\text{FolloweesCount}_{\text{it}}$		59,704,531	85.38	169.43	1	999
$\text{FollowersCount}_{\text{it}}$		59,704,531	22.41	71.63	0	3531
Peer Characteristics ($\mathbf{Y}_{kt}$)						
$\text{PeerPreferenceHHI}_{\text{kt}}$		59,704,531	0.23	0.22	0.02	1
	Yes	1,184,444	0.11	0.09	0.02	1
	No	58,520,087	0.23	0.22	0.02	1
$\text{PeerAlbumCount}_{\text{kt}}$		59,704,531	17.48	84.38	1	15938
	Yes	1,184,444	190.59	429.52	1	15938
	No	58,520,087	13.98	53.96	1	15917
$\text{PeerArtisitCount}_{\text{kt}}$		59,704,531	8.65	33.95	1	3925
	Yes	1,184,444	89.37	174.72	1	3925
	No	58,520,087	7.02	20.59	1	3824
$\text{PeerFollowees}_{\text{kt}}$		59,704,531	1.46	17.87	0	999
	Yes	1,184,444	27.05	94.37	0	999
	No	58,520,087	0.94	11.49	0	999
$\text{PeerFollowers}_{\text{kt}}$		59,704,531	2.79	52.00	0	7695
	Yes	1,184,444	95.45	347.76	1	7695
	No	58,520,087	0.92	11.58	0	7601
Dyad Characteristics ($\mathbf{Z}_{ikt}$)						
$\text{PreferenceDistance}_{\text{ikt}}$		59,704,531	0.74	0.22	0	1
	Yes	1,184,444	0.31	0.23	0	1
	No	58,520,087	0.75	0.21	0	1
$\text{CommonAlbums}_{\text{ikt}}$		59,704,531	0.10	3.43	0	5323
	Yes	1,184,444	3.90	21.23	0	3494
	No	58,520,087	0.02	1.61	0	5323
$\text{CommonArtists}_{\text{ikt}}$		59,704,531	0.33	1.25	0	2324
	Yes	1,184,444	3.31	7.10	0	985
	No	58,520,087	0.27	0.62	0	2324
$\text{CommonTies}_{\text{ikt}}$		59,704,531	0.04	0.75	0	682
	Yes	1,184,444	1.62	4.65	0	264
	No	58,520,087	0.01	0.30	0	682

Table 3 Peer Selection: Conditional Logit Regression Results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\text{PreferenceDistance}_{\text{ikt}}$		-0.2634^{***}				-0.2540^{***}	-0.2307^{***}	-0.2303^{***}
		(0.0001)				(0.0002)	(0.0003)	(0.0003)
$\text{CommonAlbums}_{\text{ikt}}$			0.4584^{***}			0.0497^{***}		0.0056^{***}
			(0.0102)			(0.0024)		(0.0005)
$\text{CommonArtists}_{\text{ikt}}$				0.4580^{***}			0.1921^{***}	0.1906^{***}
				(0.0025)			(0.0023)	(0.0023)
$\text{CommonTies}_{\text{ikt}}$					0.2647^{***}	0.0887^{***}	0.0826^{***}	0.0828^{***}
					(0.0024)	(0.0013)	(0.0015)	(0.0015)
$\text{PeerPreferenceHHI}_{\text{kt}}$	-0.2891^{***}	-0.1224^{***}	-0.2919^{***}	-0.1814^{***}	-0.2878^{***}	-0.1283^{***}	-0.1015^{***}	-0.1024^{***}
	(0.0008)	(0.0006)	(0.0009)	(0.0011)	(0.0009)	(0.0007)	(0.0007)	(0.0007)
$\text{PeerAlbumCount}_{\text{kt}}$	0.0198^{***}	0.0087^{***}	0.0020^{***}	0.0104^{***}	0.0173^{***}	0.0040^{***}	0.0063^{***}	0.0056^{***}
	(0.0006)	(0.0004)	(0.0005)	(0.0005)	(0.0006)	(0.0004)	(0.0004)	(0.0004)
$\text{PeerArtisitCount}_{\text{kt}}$	0.0712^{***}	0.0297^{***}	0.0572^{***}	0.0136^{***}	0.0579^{***}	0.0291^{***}	0.0121^{***}	0.0127^{***}
	(0.0007)	(0.0004)	(0.0007)	(0.0006)	(0.0007)	(0.0005)	(0.0005)	(0.0005)
$\text{PeerFollowees}_{\text{kt}}$	0.0028^{***}	0.0072^{***}	0.0055^{***}	0.0062^{***}	-0.0170^{***}	-0.0014^{***}	-0.0002	-0.0001
	(0.0003)	(0.0002)	(0.0003)	(0.0003)	(0.0004)	(0.0002)	(0.0002)	(0.0002)
$\text{PeerFollowers}_{\text{kt}}$	0.2532^{***}	0.1236^{***}	0.2423^{***}	0.2403^{***}	0.1852^{***}	0.1104^{***}	0.1346^{***}	0.1346^{***}
	(0.0026)	(0.0014)	(0.0026)	(0.0024)	(0.0023)	(0.0014)	(0.0016)	(0.0016)
Choice FE	Y	Y	Y	Y	Y	Y	Y	Y
Observations	58,667,246	58,667,246	58,667,246	58,667,246	58,667,246	58,667,246	58,667,246	58,667,246
Psuedo-R^2	0.313	0.612	0.392	0.530	0.398	0.639	0.678	0.678
χ^2	339802	1222927	292933	218793	266508	1142472	1049785	1049402

Notes: *** p<0.001, ** p<0.01, * p<0.05, + p<0.1. Robust standard errors in parentheses are clustered at the choice level. The table reports results from conditional logit regressions. The unit of observation is a consumer's choice whether or not to follow a peer. The dependent variable is a dummy indicating that a consumer starts to follow a peer. All independent variables are measured in standard deviation units. The reported coefficients represents average marginal effects.

Table 4 Peer Influence: Summary Statistics

Variable	**After**	**Obs**	**Mean**	**Std. dev.**	**Min**	**Max**
Dependent Variable						
Follow-Purchase		27,951,139	0.0179	0.1325	0	1
	No	13,844,074	0.0117	0.1074	0	1
	Yes	14,107,065	0.0240	0.1530	0	1
Independent Variables						
After		27,951,139	0.5047	0.5000	0	1
New Artist		27,951,139	0.9441	0.2297	0	1
	No	13,844,074	0.9422	0.2333	0	1
	Yes	14,107,065	0.9459	0.2262	0	1
Preference Distance		27,951,139	0.5493	0.2566	0	1
	No	13,844,074	0.5494	0.2568	0	1
	Yes	14,107,065	0.5491	0.2564	0	1
Followees		27,951,139	74.9796	165.5528	0	977
	No	13,844,074	43.8853	104.6086	0	968
	Yes	14,107,065	105.4942	204.1704	1	977
Free Album		27,951,139	0.2903	0.4539	0	1
	No	13,844,074	0.2949	0.4560	0	1
	Yes	14,107,065	0.2858	0.4518	0	1
Album Price		27,951,139	4.3554	3.7743	0	17.9775
	No	13,844,074	4.3186	3.7725	0	17.9775
	Yes	14,107,065	4.3915	3.7758	0	17.9775
Preference HHI		27,951,139	0.1356	0.1300	0	1
	No	13,844,074	0.1383	0.1350	0	1
	Yes	14,107,065	0.1331	0.1248	0	1
Album Sales		27,951,139	152.6500	634.4990	1	16543
	No	13,844,074	165.1630	682.8617	1	16474
	Yes	14,107,065	140.3703	582.8902	1	16543

Table 5 Peer Influence: Main Effect

	(1)	(2)	(3)	(4)	(5)	(6)
After_{it}	0.0123***	0.0134***	0.0134***	0.0140***	0.0232***	0.0188***
	(0.0002)	(0.0001)	(0.0003)	(0.0002)	(0.0003)	(0.0003)
$\text{NewArtist}_{\text{ijt}}$						−0.2174***
						(0.0010)
$\text{PreferenceDistance}_{\text{ijt}}$						−0.0188***
						(0.0004)
$\text{PreferenceHHI}_{\text{it}}$						−0.0184***
						(0.0021)
$\text{Followees}_{\text{it}}$						0.0033***
						(0.0001)
$\text{AlbumSales}_{\text{jt}}$						−0.0114***
						(0.0001)
Constant	0.0117***	0.0111***	0.0111***	0.0108***	0.0425***	0.2875***
	(0.0002)	(0.0001)	(0.0002)	(0.0001)	(0.0008)	(0.0014)
Dyad FE		Y		Y	Y	Y
Album FE			Y	Y	Y	Y
Linear Time Trend					Y	Y
Observations	27,951,139	27,951,139	27,951,139	27,951,139	27,951,139	27,951,139
Adj. R^2	0.0022	0.2227	0.0227	0.2320	0.2322	0.3326
Within-R^2	0.00216	0.00252	0.00256	0.00264	0.00301	0.133
Before-Period Mean DV	0.0117	0.0117	0.0117	0.0117	0.0117	0.0117
Relative Change	+105.6%	+114.5%	+115.1%	+120.1%	+198.7%	+161.0%

Notes: *** p<0.001, ** p<0.01, * p<0.05, + p<0.1. Robust standard errors in parentheses are clustered at the dyad level. The table reports results from linear probability models. The unit of observation are the purchases of the peer in the dyad (peer purchases). The dependent variable is a dummy indicating whether or not the consumer in the dyad purchases the same album after the peer (follow purchase). The main independent variable, After_{it}, is a dummy indicating peer purchases that happen after the formation of the dyad.

Table 6 Peer Influence: Exploration

	(1)	(2) Familiar Artist	(3) New Artist	(4)	(5) Low PD_{ijt}	(6) High PD_{ijt}
$After_{it}$	0.1635***	0.1443***	0.0066***	0.0366***	0.0282***	0.0076***
	(0.0019)	(0.0034)	(0.0001)	(0.0004)	(0.0004)	(0.0002)
$After_{it}$ X $NewArtist_{ijt}$	−0.1539***					
	(0.0019)					
$After_{it}$ X $PreferenceDistance_{ijt}$				−0.0337***		
				(0.0005)		
$NewArtist_{ijt}$	−0.1396***			−0.2173***	−0.2089***	−0.2408***
	(0.0013)			(0.0010)	(0.0011)	(0.0023)
$PreferenceDistance_{ijt}$	−0.0182***	−0.1066***	−0.0158***	−0.0019***		
	(0.0004)	(0.0100)	(0.0001)	(0.0005)		
$Followees_{it}$	0.0040***	0.0622***	0.0010***	0.0039***	0.0077***	0.0007***
	(0.0001)	(0.0022)	(0.0000)	(0.0001)	(0.0002)	(0.0001)
$PreferenceHHI_{it}$	−0.0215***	−0.2901***	−0.0021***	−0.0234***	−0.0504***	−0.0083***
	(0.0020)	(0.0327)	(0.0005)	(0.0021)	(0.0042)	(0.0018)
Album $Sales_{jt}$	−0.0110***	−0.0516***	−0.0046***	−0.0114***	−0.0171***	−0.0043***
	(0.0001)	(0.0010)	(0.0000)	(0.0001)	(0.0002)	(0.0001)
Constant	0.2117***	0.5907***	0.0339***	0.2782***	0.2997***	0.2639***
	(0.0017)	(0.0116)	(0.0003)	(0.0014)	(0.0019)	(0.0023)
Dyad FE	Y	Y	Y	Y	Y	Y
Album FE	Y	Y	Y	Y	Y	Y
Linear Time Trend	Y	Y	Y	Y	Y	Y
Observations	27,951,139	1,364,899	26,373,554	27,951,139	13,805,126	13,820,073
Adj. R^2	0.3471	0.6007	0.0809	0.3335	0.3471	0.3857
Within-R^2	0.152	0.0534	0.00411	0.134	0.124	0.146
Before-Period Mean DV	0.0117	0.147	0.00282	0.0117	0.0173	0.00583
Relative Change	-	+98.0%	+234.9%	-	+162.9%	+130.0%

Notes: *** p<0.001, ** p<0.01, * p<0.05, + p<0.1. Robust standard errors in parentheses are clustered at the dyad level. The table reports results from linear probability models. The unit of observation are the purchases of the peer in the dyad (peer purchases). The dependent variable is a dummy indicating whether or not the consumer in the dyad purchases the same album after the peer (follow purchase). The main independent variable, $After_{it}$, is a dummy indicating peer purchases that happen after the formation of the dyad.

Table 7 Peer Influence: Album Price (Mechanism Test)

		High Price	Low Price		Non-Free	Free
	(1)	(2)	(3)	(4)	(5)	(6)
After_{it}	0.0207***	0.0170***	0.0202***	0.0181***	0.0178***	0.0199***
	(0.0003)	(0.0002)	(0.0005)	(0.0002)	(0.0002)	(0.0007)
After_{it} X Price_{j}	−0.0004***					
	(0.0000)					
After_{it} X Free_{j}				0.0027***		
				(0.0003)		
$\text{NewArtist}_{\text{ijt}}$	−0.2174***	−0.1977***	−0.2339***	−0.2174***	−0.2031***	−0.2455***
	(0.0010)	(0.0008)	(0.0019)	(0.0010)	(0.0008)	(0.0030)
$\text{PreferenceDistance}_{\text{ijt}}$	−0.0187***	−0.0188***	−0.0183***	−0.0187***	−0.0189***	−0.0177***
	(0.0004)	(0.0003)	(0.0008)	(0.0004)	(0.0003)	(0.0013)
$\text{PreferenceHHI}_{\text{it}}$	−0.0183***	−0.0098***	−0.0261***	−0.0229***	−0.0183***	−0.0338***
	(0.0021)	(0.0012)	(0.0039)	(0.0021)	(0.0021)	(0.0066)
$\text{Followees}_{\text{it}}$	0.0033***	0.0025***	0.0042***	0.0034***	0.0028***	0.0051***
	(0.0001)	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0004)
Album Sales_{jt}	−0.0114***	−0.0107***	−0.0116***	−0.0114***	−0.0111***	−0.0111***
	(0.0001)	(0.0001)	(0.0002)	(0.0001)	(0.0001)	(0.0003)
Constant	0.2873***	0.2619***	0.3066***	0.2874***	0.2684***	0.3197***
	(0.0014)	(0.0010)	(0.0025)	(0.0014)	(0.0010)	(0.0039)
Dyad FE	Y	Y	Y	Y	Y	Y
Album FE	Y	Y	Y	Y	Y	Y
Linear Time Trend	Y	Y	Y	Y	Y	Y
Observations	27,951,139	13,905,184	13,877,499	27,951,139	19,802,331	7,969,070
Adj. R2	0.3327	0.2474	0.4270	0.3327	0.2618	0.5062
Within-R2	0.133	0.116	0.145	0.133	0.122	0.150
Before-Period Mean DV	0.0117	0.00960	0.0135	0.0117	0.00979	0.0159
Relative Change	-	+177.4%	+149.6%	-	+181.6%	+125.2%

Notes: *** p<0.001, ** p<0.01, * p<0.05, + p<0.1. Robust standard errors in parentheses are clustered at the dyad level. The table reports results from linear probability models. The unit of observation are the purchases of the peer in the dyad (peer purchases). The dependent variable is a dummy indicating whether or not the consumer in the dyad purchases the same album after the peer (follow purchase). The main independent variable, After_{it}, is a dummy indicating peer purchases that happen after the formation of the dyad.

Table 8 Peer Influence: Low Consumer-Peer Preference Distance at Follow

	(1) Full Sample	(2) Familiar Artist	(3) New Artist	(4) Low PD_{ijt}	(5) High PD_{ijt}	(6) $Non\text{-}Free_j$	(7) $Free_j$
$After_{it}$	0.0228***	0.1447***	0.0082***	0.0285***	0.0121***	0.0215***	0.0246***
	(0.0003)	(0.0036)	(0.0001)	(0.0004)	(0.0003)	(0.0003)	(0.0009)
Observations	18,036,531	1,133,332	16,728,632	11,765,063	7,113,141	13,016,605	4,923,111
Adj. R^2	0.2945	0.0.5612	0.0738	0.3147	0.2834	0.2488	0.4422
Within-R^2	0.127	0.055	0.00533	0.121	0.119	0.119	0.137
Before-Period Mean DV	0.0134	0.134	0.00343	0.0169	0.00732	0.0117	0.0174
Relative Change	+170.0%	+107.7%	+240.2%	+169.3%	+165.9%	+182.8%	+141.5%

Notes: *** p<0.001, ** p<0.01, * p<0.05, + p<0.1. Robust standard errors in parentheses are clustered at the dyad level. The table reports results from linear probability models. The unit of observation are the purchases of the peer in the dyad (peer purchases). The dependent variable is a dummy indicating whether or not the consumer in the dyad purchases the same album after the peer (follow purchase). The main independent variable, $After_{it}$, is a dummy indicating peer purchases that happen after the formation of the dyad. All models contain the full set of controls and fixed effects.

Table 9 Peer Influence: High Consumer-Peer Preference Distance at Follow

	(1) Full Sample	(2) Familiar Artist	(3) New Artist	(4) Low PD_{ijt}	(5) High PD_{ijt}	(6) $Non\text{-}Free_j$	(7) $Free_j$
$After_{it}$	0.0095***	0.1396***	0.0032***	0.0235***	0.0058***	0.0089***	0.0103***
	(0.0004)	(0.0128)	(0.0001)	(0.0012)	(0.0003)	(0.0003)	(0.0012)
Observations	9,746,128	184,313	9,475,979	2,223,293	7,871,602	6,675,684	2,985,440
Adj. R^2	0.4680	0.7884	0.1061	0.5109	0.4645	0.3252	0.6647
Within-R^2	0.176	0.0412	0.00219	0.158	0.168	0.148	0.218
Before-Period Mean DV	0.00802	0.236	0.00160	0.0196	0.00523	0.00544	0.0132
Relative Change	+118.6%	+59.1%	+198.3%	+120.0%	+111.4%	+163.2%	+77.7%

Notes: *** p<0.001, ** p<0.01, * p<0.05, + p<0.1. Robust standard errors in parentheses are clustered at the dyad level. The table reports results from linear probability models. The unit of observation are the purchases of the peer in the dyad (peer purchases). The dependent variable is a dummy indicating whether or not the consumer in the dyad purchases the same album after the peer (follow purchase). The main independent variable, $After_{it}$, is a dummy indicating peer purchases that happen after the formation of the dyad. All models contain the full set of controls and fixed effects.

Appendix

A.1 Comparing the Periods before and after the Dyad Formation

We provide additional statistics to show that our results of the peer effects in section 5 are not driven by changes in peer behavior. First, we plot the number of peer purchases over the days before and after the dyad formation in figre A1. While we do observe a very slight increase towards that date. However, the number of observations is rather balanced, so we are not concerned that this threatens the validity of our results. Second, we may be concerned that our results pertaining to the exploration of new music may be caused by changes in peer behavior. For instance, it may be the case that he purchases more albums that are exploration opportunities for the consumer after the dyad formation than before. Our data suggests otherwise: Figure A2 shows the distribution of consumer-album preference distances between the before and after periods, and it provides clear evidence that they do not differ to the extent that would threaten the validity of our results. In addition, in table 4 we already show that the share of new artist albums does not differ substantially either. Hence, changes in the types of horizontal attributes the peer consumes do not seem to be driving our results.

Figure A1 Observations before and after the dyad formation

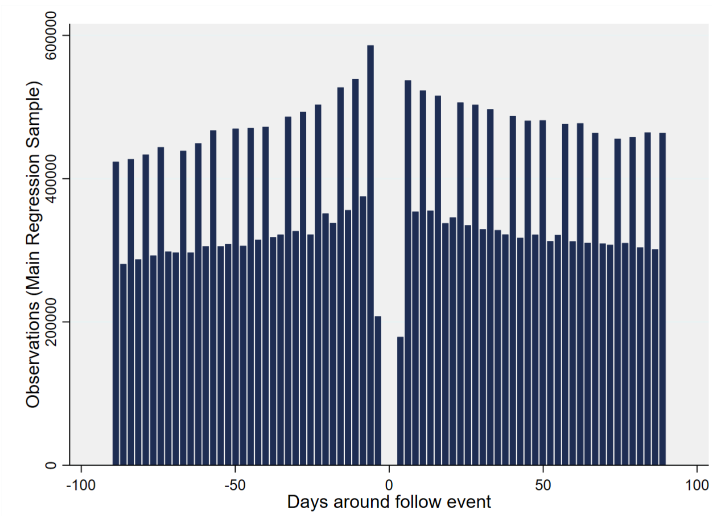
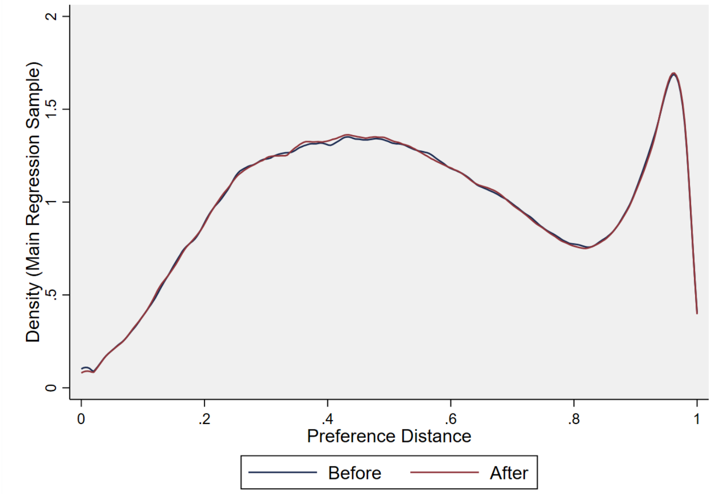


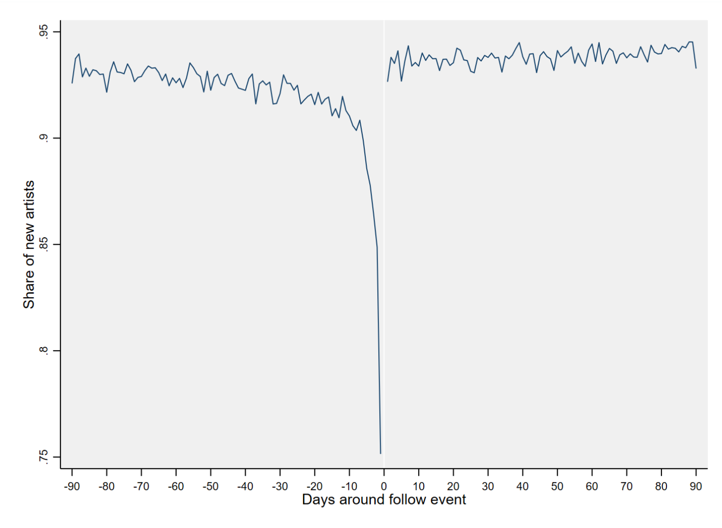
Figure A2 Kernel Density: Preference Distance, Before vs. After



Next, our identification strategy makes the (implicit) assumption that homophily between a consumer and peer is constant. However, in principle it is possible that the two develop more similar preferences over time, which could both spark the formation of the dyad and would bias our estimates. In this case, what we detect as exploration would simply reflect consumers' preference evolution that happens without peer influence. However, our data suggests otherwise. Figure A3 plots the share of new artists (for the consumer) as well as the average consumer-album preference distance for the peer purchases in our sample over time. If the consumer and peer would become more similar, we would detect a steady decrease in both. However, we only detect very slight decreases for either in the before-period. The sharp "dips" just before the dyad formation can be attributed to the mode of peer discovery: Recall that they can find them on album pages. This means, that a consumer purchases an album, then investigates others who did the same, and then may discover a peer to follow. In the period after the dyad formation, both exploration measures remain constant. Together, we are not concerned that this threatens the validity of our identification strategy. We also show the development of the share of free albums and their average price over time in Figure A4. Aside from a similar "dip" right before the dyad formation, we do not detect any trends that would cause concern.

Figure A3 Exploration Opportunities over Time

(a) Share of New Artist Peer Purchases



(b) Average Preference Distance of Peer Purchases

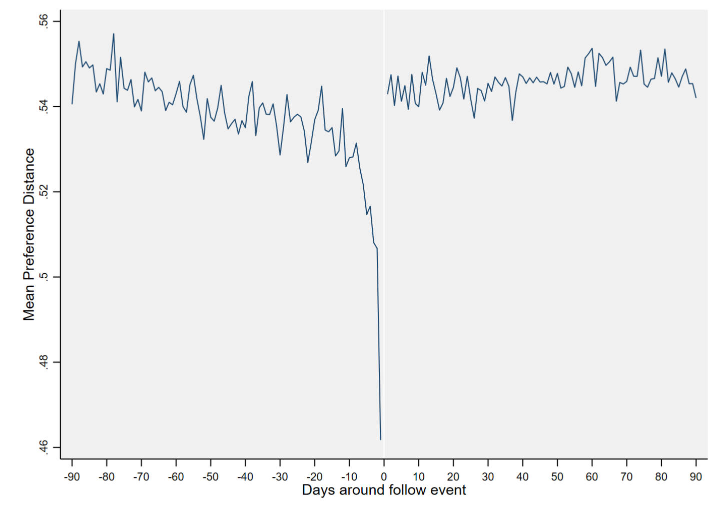
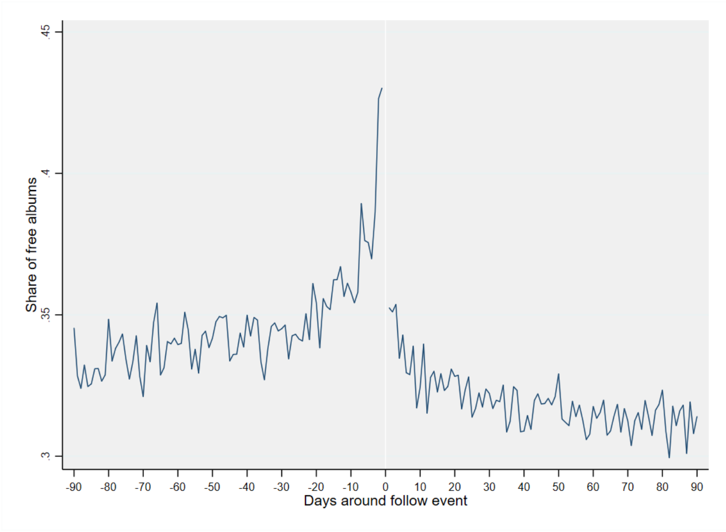
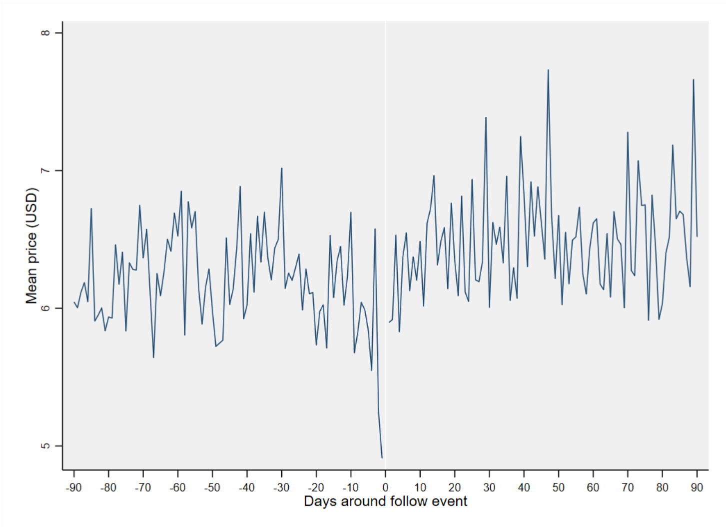


Figure A4 Peer Purchase Price over Time

(a) Share of Free Albums



(b) Average Album Price of Peer Purchases



A.2 More interactions!

Here, we investigate more heterogeneity in the peer effect. For instance, a consumer may be more willing to purchase an album from a new artist that exhibits familiar characteristics. At the same time, she may still be less willing to pay a positive price for such an exploration opportunity. Therefore, we perform further sample splits along these dimensions to explore how they interact with each other. Results are presented in Table A1. Similar as before, we find the starkest contrast in the relative increase in the probability of follow purchases between albums from new artists and others. For the former, this ranges from 211.2% to 274%, and for the latter from 47.3% to 109.6%. This adds further evidence that strongest effect of participating in the network is in the discovery of new artists. Next, contrasting the relative increases between low and high preferences distances also confirms the notion that this nudge is more pronounced for albums that exhibit relatively familiar characteristics, which is the case for both albums from new artists and others. Finally, a more nuanced picture emerges when analyzing the distinction between free and non-free products. The relative probability increase is stronger for non-free albums when they are from familiar artists. In contrast, the probability increase is stronger for free albums when they are from new artist. This suggests that consumers are more willing to explore new artists when they do not have to pay for the album. In all, we find the strongest peer effect for albums from new artists, that exhibit familiar horizontal characteristics, and are offered for free. In this case, follow purchases are 274.5% more likely when consumer are able to observe the peer purchases compared to when they are not.

=== Table A1 here ===

This also has implications for the relative importance of the awareness and signaling effects. On their own, consumers seem unable to discover these albums from new artists, speaking to a strong awareness effect from observing others purchasing from them. In contrast, the signaling effect plays a limited role. Consumers are relatively less inclined to deviate from their horizontal preferences, and they are less willing to pay a price for albums from unfamiliar artists. That indicates that their expectation of a mismatch with their preferences outweighs a potential quality premium that they may expect.

Table A1 Peer Influence: More Interactions

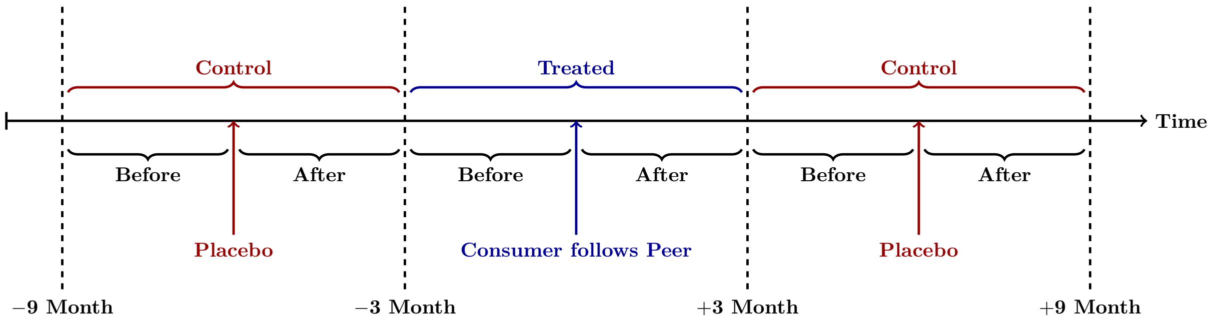
	Familiar Artist				New Artist			
	Low PD		High PD		Low PD		High PD	
	Non-Free (1)	Free (2)	Non-Free (3)	Free (4)	Non-Free (5)	Free (6)	Non-Free (7)	Free (8)
After_{it}	0.1312***	0.1758***	0.1079***	0.1347***	0.0103***	0.0100***	0.0030***	0.0027***
	(0.0029)	(0.0138)	(0.0079)	(0.0180)	(0.0002)	(0.0003)	(0.0001)	(0.0002)
$\text{PreferenceHHI}_{\text{it}}$	−0.2875***	−0.5559**	−0.2268***	−0.5147***	−0.0115***	−0.0147***	−0.0008	−0.0018+
	(0.0294)	(0.1842)	(0.0517)	(0.1115)	(0.0012)	(0.0025)	(0.0005)	(0.0010)
$\text{Followees}_{\text{it}}$	0.0588***	0.0747***	0.0644***	0.0748***	0.0026***	0.0035***	0.0002***	0.0003***
	(0.0016)	(0.0095)	(0.0046)	(0.0085)	(0.0001)	(0.0002)	(0.0000)	(0.0000)
$\text{AlbumSales}_{\text{jt}}$	−0.0588***	−0.0335***	−0.0381***	−0.0260***	−0.0071***	−0.0064***	−0.0019***	−0.0016***
	(0.0008)	(0.0029)	(0.0022)	(0.0034)	(0.0001)	(0.0002)	(0.0000)	(0.0001)
Constant	0.4960***	0.6742***	0.3367***	0.5982***	0.0383***	0.0359***	0.0108***	0.0095***
	(0.0103)	(0.0528)	(0.0272)	(0.0598)	(0.0006)	(0.0012)	(0.0003)	(0.0005)
Dyad FE	Y	Y	Y	Y	Y	Y	Y	Y
Album FE	Y	Y	Y	Y	Y	Y	Y	Y
Linear Time Trend	Y	Y	Y	Y	Y	Y	Y	Y
Observations	676,417	375,328	101,167	74,374	8,957,152	3,406,423	9,543,568	3,802,677
Adj. R^2	0.5093	0.7709	0.6699	0.9028	0.0973	0.1559	0.1066	0.1628
Within-R^2	0.0494	0.0473	0.0452	0.0565	0.00463	0.00441	0.00110	0.000918
Before-Period Mean DV	0.120	0.164	0.116	0.285	0.00431	0.00364	0.00143	0.00113
Relative Change	+109.6%	+107.4%	+92.9%	+47.3%	+239.2%	+274.5%	+211.2%	+241.7%

Notes: *** p<0.001, ** p<0.01, * p<0.05, + p<0.1. Robust standard errors in parentheses are clustered at the dyad level. The table reports results from linear probability models. The unit of observation are the purchases of the peer in the dyad (peer purchases). The dependent variable is a dummy indicating whether or not the consumer in the dyad purchases the same album after the peer (follow purchase). The main independent variable, After_{it}, is a dummy indicating peer purchases that happen after the formation of the dyad.

A.3 Time Shift Control

We lay out our empirical approach to obtain a causal estimate for the peer effect associated with observing peer purchases in section 5. We exploit the panel structure of our data to identify sample observations where consumers are able to and where they are not. This is the case after and before they follow a peer. We further control for between-dyad differences via fixed effects. One limitation of our data is that there is no natural control group. Hence, we estimate our main peer effect by comparing the probability of follow purchases before and after the formation of the dyad.

Figure A5 Time Shift Control Approach



As a robustness check, we conduct an additional analyses where we construct a set of control observations. We proceed as follows: For each dyad, we know the date of its formation. Just after, peer purchase observations are possible, and just before they are not (this is how we identify the effect in our main analysis). In total, we consider a six-month window - three before and three after. We call this entire window *treated* - the consumer's information regime changes. Next, within each dyads, we define two placebo dates for the dyad formation - one six months before the actual date, and the other six months after. For both placebo dates we again consider a time window of three month *before* and three month *after*. We call take observations with these two windows *control*. The approach is illustrated in Figure A5. Now, to estimate a causal peer effect with this control group, we compare if the change in the probability of follow purchases from the before to the after periods differs between the treated and control time windows. The idea is that this probability should not differ within the control time windows - they are defined around a placebo date. But they should differ within the treated time window, which contains the actual date of the dyad formation. Hence, we estimate the main peer effect using the following difference-in-differences model:

$$\begin{aligned}\mathrm{FP_{ij}} = {} & \beta_0 + \beta_1 \mathrm{After_{it}} + \beta_2 \mathrm{Treated_{it}} + \beta_3 (\mathrm{After_{it}} \times \mathrm{Treated_{it}}) \\ & + \beta_X \mathbf{X} + \alpha_{ik} + \alpha_j + t_{ik} + (t_{ik} \times \mathrm{Treated_{it}}) + \epsilon.\end{aligned}$$

We include the same set of time-varying controls and fixed effects as in our main analysis. In addition, we let the dyad-specific time trend vary between the control and treated time windows. Hence, we still explore within-dyad variation to account for peer selection and unobserved time trends. To evaluate effect heterogeneity, we split the sample between new and other artists, high and low consumer-album preference distances, and free and non-free albums. Table A5 contains the results from our LPM regressions. Throughout, the coefficient of the DiD term here corresponds to the one on the after dummy in the main analysis. Across the board, results are consistent with our main analysis, both qualitatively and in scope.

Table A2 Time Shift Control

	(1) Full	(2) Familiar Artist	(3) New Artist	(4) Low PD_{ijt}	(5) High PD_{ijt}	(6) Non-Free	(7) Free
$Treated_{it}$	0.0270***	0.2078***	0.0094***	0.0403***	0.0109***	0.0247***	0.0319***
	(0.0009)	(0.0091)	(0.0003)	(0.0014)	(0.0007)	(0.0006)	(0.0026)
$After_{it}$	−0.0009***	−0.0102***	−0.0004***	−0.0013***	−0.0004***	−0.0011***	−0.0003
	(0.0001)	(0.0016)	(0.0000)	(0.0002)	(0.0001)	(0.0001)	(0.0004)
$After_{it}$ X $Treated_{it}$	0.0202***	0.1723***	0.0067***	0.0310***	0.0076***	0.0190***	0.0224***
	(0.0003)	(0.0036)	(0.0001)	(0.0005)	(0.0003)	(0.0002)	(0.0009)
$NewArtist_{ijt}$	−0.2025***			−0.1973***	−0.2151***	−0.1885***	−0.2317***
	(0.0007)			(0.0007)	(0.0015)	(0.0005)	(0.0020)
$PreferenceDistance_{ijt}$	−0.0184***	−0.1264***	−0.0143***			−0.0188***	−0.0159***
	(0.0003)	(0.0054)	(0.0001)			(0.0002)	(0.0008)
$Followees_{it}$	0.0034***	0.0450***	0.0011***	0.0064***	0.0012***	0.0029***	0.0048***
	(0.0001)	(0.0009)	(0.0000)	(0.0001)	(0.0000)	(0.0000)	(0.0002)
$PreferenceHHI_{it}$	−0.0158***	−0.2610***	−0.0015***	−0.0392***	−0.0085***	−0.0086***	−0.0328***
	(0.0010)	(0.0149)	(0.0002)	(0.0020)	(0.0010)	(0.0005)	(0.0032)
Album $Sales_{jt}$	−0.0103***	−0.0531***	−0.0039***	−0.0157***	−0.0038***	−0.0099***	−0.0106***
	(0.0001)	(0.0005)	(0.0000)	(0.0001)	(0.0001)	(0.0000)	(0.0002)
Constant	0.2455***	0.4190***	0.0223***	0.2477***	0.2275***	0.2288***	0.2778***
	(0.0008)	(0.0050)	(0.0001)	(0.0010)	(0.0015)	(0.0006)	(0.0023)
Dyad FE	Y	Y	Y	Y	Y	Y	Y
Album FE	Y	Y	Y	Y	Y	Y	Y
Linear Time Trend	Y	Y	Y	Y	Y	Y	Y
Observations	74,523,242	3,695,876	70,601,953	37,110,816	37,140,504	53,403,971	20,978,744
Adj. R^2	0.2785	0.5181	0.0542	0.2953	0.3132	0.2200	0.4346
Within-R^2	0.125	0.0423	0.00342	0.117	0.132	0.113	0.145
Mean DV	0.0156	0.217	0.00408	0.0247	0.00636	0.0139	0.0196

Notes: *** p<0.001, ** p<0.01, * p<0.05, + p<0.1. Robust standard errors in parentheses are clustered at the dyad level. The table reports results from linear probability models. The unit of observation are the purchases of the peer in the dyad (peer purchases). The dependent variable is a dummy indicating whether or not the consumer in the dyad purchases the same album after the peer (follow purchase).

A.4 Additional Robustness Checks

We conducted a number of additional robustness checks for our results pertaining to the causal peer influence discussed in section 5. First, we re-ran part of our analysis using different time windows and donuts. In our main analysis we use samples consisting of peer purchases that took place within the three months before and after the consumer started following the peer. In addition, we implement a donut around this event by excluding peer purchases that took place within the three days before and after the follow event. Here, we test the robustness by using time windows of one (table A3), six (table A4), and twelve months (table A5) before and after. Our results are qualitatively consistent. However, effect sizes increase with narrower and decrease with wider time windows. This is consistent with the pattern shown in figure 9. Effects are strongest right after the follow event, but become weaker as time progresses. Hence, specifications with longer time windows include relatively more observations for which the effect has already worn off. Next, we use different donuts, namely excluding observations falling within one (table A6) and seven (table A7) days before and after the follow event. Here, effects are stronger when a narrower donut is used. Again, this is likely due to the fact that more observations just after the event are excluded with a wider donut, i.e. those for which the effect is strongest.

Second, we re-ran part of analysis excluding "hybrid" cases, where a consumer bought an album in the *after* period, but the peer purchase took place in the *before* period. In our main analysis, we attribute these cases to the after-period, hence we treat them as "under peer influence". However, here the only possible peer influence channel is the consumer browsing the peer's album collection, which she in can already do before following him. Therefore, to evaluate to what extent this drives our results, we exclude them here and present the results in table A8. Our results are qualitatively robust, but effect sizes are considerably smaller. In the one hand, this is not surprising – we do exclude instances of follow purchases from the *after* period, hence the effect can only be smaller. On the other hand, the order of magnitude is substantial: Effect sizes including hybrid cases are around six times higher compared to excluding them. Again, consistent with the pattern from figure 9, a likely explanation is that consumers browse a peer's album collection – which contains their purchases from before the follow event – quickly after following them, but do not seem to pay much attention to other peer influence channels as time progresses. As a result, the effect is rather immediate and largely driven by hybrid cases.

Table A3 Robustness: Time window of one month before and after the dyad formation

	(1) Full Sample	(2) Familiar Artist	(3) New Artist	(4) Low PD_{ijt}	(5) High PD_{ijt}	(6) Non-$Free_j$	(7) $Free_j$
$After_{it}$	0.0230***	0.1525***	0.0087***	0.0335***	0.0117***	0.0222***	0.0228***
	(0.0005)	(0.0077)	(0.0002)	(0.0008)	(0.0004)	(0.0004)	(0.0016)
Observations	9,200,532	405,764	8,628,007	4,618,516	4,797,668	6,412,245	2,613,517
Adj. R^2	0.4001	0.7226	0.1265	0.4121	0.4421	0.3169	0.5945
Within-R^2	0.0117	0.147	0.00282	0.0117	0.0173	0.00583	0.155
Before-Period Mean DV	0.0106	0.126	0.00247	0.0143	0.00643	0.00852	0.0149
Relative Change	+216.9%	+121.1%	+352.5%	+234.0%	+182.6%	+260.8%	+152.4%

Notes: *** p<0.001, ** p<0.01, * p<0.05, + p<0.1. Robust standard errors in parentheses are clustered at the dyad level. The table reports results from linear probability models. The unit of observation are the purchases of the peer in the dyad (peer purchases). The dependent variable is a dummy indicating whether or not the consumer in the dyad purchases the same album after the peer (follow purchase). The main independent variable, $After_{it}$, is a dummy indicating peer purchases that happen after the formation of the dyad. All models contain the full set of controls and fixed effects.

Table A4 Robustness: Time window of six months before and after the dyad formation

	(1) Full Sample	(2) Familiar Artist	(3) New Artist	(4) Low PD_{ijt}	(5) High PD_{ijt}	(6) Non-$Free_j$	(7) $Free_j$
$After_{it}$	0.0148***	0.1257***	0.0050***	0.0226***	0.0067***	0.0140***	0.0162***
	(0.0002)	(0.0024)	(0.0001)	(0.0003)	(0.0001)	(0.0001)	(0.0006)
Observations	52,679,593	2,620,252	49,835,808	26,552,463	29,007,839	37,600,377	14,920,962
Adj. R^2	0.2962	0.5441	0.0620	0.3114	0.3122	0.2315	0.4609
Within-R^2	0.128	0.0441	0.00359	0.119	0.125	0.116	0.146
Before-Period Mean DV	0.0125	0.162	0.00307	0.0189	0.00639	0.0107	0.0168
Relative Change	+118.8%	+77.4%	+162.4%	+119.6%	+105.0%	+131.7%	+96.3%

Notes: *** p<0.001, ** p<0.01, * p<0.05, + p<0.1. Robust standard errors in parentheses are clustered at the dyad level. The table reports results from linear probability models. The unit of observation are the purchases of the peer in the dyad (peer purchases). The dependent variable is a dummy indicating whether or not the consumer in the dyad purchases the same album after the peer (follow purchase). The main independent variable, $After_{it}$, is a dummy indicating peer purchases that happen after the formation of the dyad. All models contain the full set of controls and fixed effects.

Table A5 Robustness: Time window of twelve months before and after the dyad formation

	(1) Full Sample	(2) Familiar Artist	(3) New Artist	(4) Low PD_{ijt}	(5) High PD_{ijt}	(6) Non-$Free_j$	(7) $Free_j$
$After_{it}$	0.0107***	0.0984***	0.0034***	0.0168***	0.0047***	0.0101***	0.0120***
	(0.0002)	(0.0018)	(0.0000)	(0.0002)	(0.0001)	(0.0001)	(0.0005)
Observations	94,069,127	4,621,096	89,223,503	46,806,002	52,664,034	67,642,620	26,298,879
Adj. R^2	0.2661	0.5007	0.0494	0.2840	0.2717	0.2121	0.4162
Within-R^2	0.124	0.0405	0.00324	0.115	0.119	0.111	0.145
Before-Period Mean DV	0.0134	0.178	0.00335	0.0210	0.00661	0.0117	0.0175
Relative Change	+79.9%	+55.2%	+103.0%	+80.1%	+70.4%	+86.1%	+68.5%

Notes: *** p<0.001, ** p<0.01, * p<0.05, + p<0.1. Robust standard errors in parentheses are clustered at the dyad level. The table reports results from linear probability models. The unit of observation are the purchases of the peer in the dyad (peer purchases). The dependent variable is a dummy indicating whether or not the consumer in the dyad purchases the same album after the peer (follow purchase). The main independent variable, $After_{it}$, is a dummy indicating peer purchases that happen after the formation of the dyad. All models contain the full set of controls and fixed effects.

Table A6 Robustness: "Donut" of one day before and after the dyad formation

	(1) Full Sample	(2) Familiar Artist	(3) New Artist	(4) Low PD_{ijt}	(5) High PD_{ijt}	(6) Non-$Free_j$	(7) $Free_j$
$After_{it}$	0.0199***	0.1499***	0.0070***	0.0295***	0.0094***	0.0188***	0.0213***
	(0.0003)	(0.0032)	(0.0001)	(0.0004)	(0.0002)	(0.0002)	(0.0008)
Observations	28,807,822	1,432,003	27,159,841	14,601,882	15,604,417	20,382,433	8,245,908
Adj. R^2	0.3344	0.6028	0.0820	0.3467	0.3649	0.2634	0.5075
Within-R^2	0.135	0.0557	0.00423	0.125	0.136	0.123	0.151
Before-Period Mean DV	0.0118	0.147	0.00285	0.0172	0.00647	0.00988	0.0162
Relative Change	+168.7%	+102.3%	+244.5%	+172.2%	+145.2%	+190.5%	+131.8%

Notes: *** p<0.001, ** p<0.01, * p<0.05, + p<0.1. Robust standard errors in parentheses are clustered at the dyad level. The table reports results from linear probability models. The unit of observation are the purchases of the peer in the dyad (peer purchases). The dependent variable is a dummy indicating whether or not the consumer in the dyad purchases the same album after the peer (follow purchase). The main independent variable, $After_{it}$, is a dummy indicating peer purchases that happen after the formation of the dyad. All models contain the full set of controls and fixed effects.

Table A7 Robustness: "Donut" of seven days before and after the dyad formation

	(1) Full Sample	(2) Familiar Artist	(3) New Artist	(4) Low PD_{ijt}	(5) High PD_{ijt}	(6) Non-$Free_j$	(7) $Free_j$
$After_{it}$	0.0176***	0.1368***	0.0061***	0.0262***	0.0081***	0.0165***	0.0191***
	(0.0003)	(0.0038)	(0.0001)	(0.0005)	(0.0002)	(0.0002)	(0.0008)
Observations	26,422,901	1,266,153	24,948,898	13,340,778	14,341,979	18,741,545	7,501,154
Adj. R^2	0.3322	0.6006	0.0806	0.3473	0.3547	0.2613	0.5063
Within-R^2	0.132	0.0511	0.00394	0.123	0.131	0.120	0.149
Before-Period Mean DV	0.0117	0.150	0.00282	0.0173	0.00614	0.00979	0.0160
Relative Change	+150.4%	+91.4%	+216.9%	+151.4%	+131.7%	+169.0%	+119.8%

Notes: *** p<0.001, ** p<0.01, * p<0.05, + p<0.1. Robust standard errors in parentheses are clustered at the dyad level. The table reports results from linear probability models. The unit of observation are the purchases of the peer in the dyad (peer purchases). The dependent variable is a dummy indicating whether or not the consumer in the dyad purchases the same album after the peer (follow purchase). The main independent variable, $After_{it}$, is a dummy indicating peer purchases that happen after the formation of the dyad. All models contain the full set of controls and fixed effects.

Table A8 Robustness: Excluding "hybrid" follow purchases

	(1) Full Sample	(2) Familiar Artist	(3) New Artist	(4) Low PD_{ijt}	(5) High PD_{ijt}	(6) Non-$Free_j$	(7) $Free_j$
$After_{it}$	0.0033***	0.0341***	0.0010***	0.0057***	0.0011***	0.0034***	0.0030***
	(0.0002)	(0.0025)	(0.0001)	(0.0003)	(0.0001)	(0.0001)	(0.0005)
Observations	27,785,999	1,247,370	26,334,284	14,010,189	15,117,568	19,702,052	7,904,623
Adj. R^2	0.2757	0.5613	0.0635	0.2876	0.3099	0.2079	0.4506
Within-R^2	0.0966	0.0278	0.00213	0.0888	0.0983	0.0888	0.107
Before-Period Mean DV	0.0117	0.146	0.00282	0.0171	0.00624	0.00978	0.0159
Relative Change	+28.6%	+23.3%	+35.2%	+33.6%	+17.6%	+34.3%	+19.2%

Notes: *** p<0.001, ** p<0.01, * p<0.05, + p<0.1. Robust standard errors in parentheses are clustered at the dyad level. The table reports results from linear probability models. The unit of observation are the purchases of the peer in the dyad (peer purchases). The dependent variable is a dummy indicating whether or not the consumer in the dyad purchases the same album after the peer (follow purchase). The main independent variable, $After_{it}$, is a dummy indicating peer purchases that happen after the formation of the dyad. All models contain the full set of controls and fixed effects.

A.5 Additional Tables

Table A9 List of Tags

Tag	N	Tag	N	Tag	N	Tag	N
electronic	823752	indie-pop	60107	synthwave	31127	disco	20500
experimental	642133	post-punk	57620	trip-hop	30829	darkwave	20461
rock	557546	death-metal	57397	spoken-word	30773	r-b-soul	20037
alternative	456099	dance	55933	hiphop	30575	instrumentals	19607
ambient	384579	shoegaze	52767	doom	30350	electroacoustic	19028
punk	321466	idm	52158	chillout	30282	sludge	19026
hip-hop-rap	301406	avant-garde	48927	edm	30166	progressive-metal	18959
hip-hop	270172	progressive	48183	bass	30079	surf	18072
pop	266465	classical	47782	grindcore	30037	chillwave	17992
metal	263073	trap	47071	chill	29544	orchestral	17740
indie	227087	dubstep	46889	trance	29001	sound-art	17653
folk	218159	downtempo	46743	garage-rock	28927	stoner-rock	17613
noise	193186	lofi	46250	guitar	28832	world-music	17524
acoustic	168170	garage	44210	ambient-electronic	28220	power-pop	17377
techno	161480	underground	43682	psychedelic-rock	28107	fusion	16992
rap	160371	emo	43468	harsh-noise	27874	boom-bap	16932
instrumental	138151	dub	43079	devotional	27820	field-recordings	16909
jazz	133321	pop-punk	42955	post-hardcore	27809	jungle	16844
indie-rock	132045	deep-house	42386	glitch	26669	christian	16413
drone	121648	minimal	41558	tech-house	26493	free-jazz	16344
hardcore	121020	reggae	40869	electronic-music	26378	breakbeat	15994
house	116860	vaporwave	39667	soundscape	26168	stoner	15847
psychedelic	109919	r-b	39512	dream-pop	25562	blues-rock	15292
electronica	109806	underground-hip-hop	39394	heavy-metal	25469	ska	15052
lo-fi	96495	country	36673	atmospheric	25403	acoustic-guitar	15011
soundtrack	94071	folk-rock	36178	new-wave	24991	roots	15000
singer-songwriter	89206	improvisation	36153	noise-rock	24959	latin	14977
industrial	88137	americana	35501	experiemental	24920	abstract	14955
alternative-rock	75001	drum-bass	35284	dark	24862	live	14826
dark-ambient	74798	instrumental-hip-hop	35165	psytrance	24723	cinematic	14816
soul	73844	hardcore-punk	34456	experimental-rock	24197	experimental-hip-hop	14764
funk	71689	progressive-rock	34322	rock-roll	23991	bass-music	14644
black-metal	71521	synthpop	34210	thrash-metal	23837	math-rock	14462
beats	71485	hard-rock	33777	acid	23117	deep	14399
world	69971	synth	32770	metalcore	22505	space	14310
electro	69385	diy	32407	doom-metal	22389	bedroom-pop	14248
post-rock	65825	pop-rock	32285	music	21686		
blues	64379	indie-folk	31846	new-age	20697		
experimental-electronic	63475	piano	31823	chiptune	20567		
punk-rock	61565	grunge	31630	comedy	20513		