Learning or Herding? Understanding Social Interactions and the Distribution of Success on a Social Music Sharing Platform

James Sams *

2019–06–05

Abstract

Digital sharing platforms like YouTube and SoundCloud crowdsource the process by which users can discover high quality new products among an increasingly vast flow of new products, acting as on-going digital test markets. Social features on these platforms can accelerate the discovery process by encouraging sharing of information and facilitating learning, thereby reducing the number of people sampling poor quality products. This may more quickly concentrate platform traffic on higher quality alternatives. Social features may also include a feedback loop if people care about consuming the same products as their peers. Given previous research showing that social feedback loops can distort or even invert the relationship between product quality and product popularity, if such feedback loops exist, the discovery and filtering capabilities of crowdsourcing may be compromised, emphasizing the need to understand the nature of social interactions on such platforms. Utilizing data from SoundCloud, a music sharing and streaming site, I develop an approach to separately identify and measure these two separate endogenous social effects with and without feedback loops. Results suggest that the platform's social features do have informative effects but that the feedback loop plays a dominant role for the most successful songs.

*Graduate School of Business, Stanford University, jsams@stanford.edu. Many thanks to my advisers Harikesh Nair and Wes Hartmann, as well as the members of my committee Peter Reiss and Bart Bronnenberg, and to Navdeep Sahni for extended discussions of this paper. Additionally, thanks to the participants of the Stanford GSB Marketing Seminar and Quantitative Marketing WiP Seminar. Special thanks to Florian Stahl and SoundCloud for allowing me free access to this data.

1 Introduction

The power of social influence is increasingly recognized as a large component in many decision contexts, from the adoption of new production technologies to the choice of consumer goods and services. However, whether these social influences represent people "learning from the crowd" or simply "going with the crowd" often is less clear. This paperis a step toward understanding how an individual's response to social signals may depend on their available information, and how those responses can tell researchers whether a social network is operating to inform its participants or is instead coordinating activities among its users.

While social effects have been studied well, the test market literature has lagged in better understanding the implication of social effects on their models and practices. Current sparseness in the intersection of test markets, social effects, and the digital economy contrasts with the rich literature and carefully developed services, such as Nielsen Test Markets. This paperis a step toward bridging that gap. Some of the early literature has grappled with how to ensure that test markets results are informative for the marketing practitioner. For example, Silk and Urban 1978 describe how to pre-test products so that those products that go on to test markets return information useful to the marketing practitioner. Urban and Hauser 1993 consolidate and systematize the research on new product development, providing a roadmap from how to understand consumer preferences, how to design products with those preferences in mind, to market segmentation, product life-cycle management, and anticipating useful innovation.

Researchers then came to discover the difficulty of generalizing from early indicators of product success. Moore 2002 explores the difficulty of "crossing the chasm" between early and late adopters in models of product life cycles. This thread of research recognizes that early adopters, themselves representing a kind of temporal test market, are not necessarily representative of consumers in the full market. Thus, which features they prefer and their willingness to pay is potentially systematically different from the broader market. These differences imply that marketers may need to adjust the product itself, the product release strategy, or the product pricing schedule to reflect the heterogeneity in the wider market. However, with larger portions of the cultural economy shifting to digital services online, new challenges are presented to test markets. The integration of social features with test markets and what this may mean for test markets outcomes is the primary feature that will be explored in this paper.

I use data from SoundCloud, a streaming music platform that operates as a digital test market for independent musicians, where users can recommend songs to others in their network. It is natural to think that a person's response to a recommendation for an unfamiliar song constitutes a form of learning, where the user updates their prior beliefs about the song's quality and decides whether to sample the song based on those updated beliefs. However, if a user is already familiar with a recommended song, the scope for learning from a simple peer recommendation is substantially reduced. Thus the response to such a recommendation reflects the utility found in social interactions vis-à-vis product consumption rather than changed beliefs about the song's quality. This extra utility from being able to co-consume the same content with peers is defined as co-consumption utility.

I introduce a framework for being able to separately identify and measure these two separate endogenous social effects: social learning and co-consumption utility. The framework leverages the fact that many platforms now record the timing of signals sent to fellow users, e.g. likes on Facebook, as well as those users' response to the signals. Using the fact that some users receive recommendations before listening to a song and others receive the recommendation after having already listened to the song, the relative strength of social learning compared to co-consumption utility can be measured. Estimating the model on data from SoundCloud reveals that on their platform co-consumption utility dominates social learning for the most popular songs. Counterfactual analysis further suggests that while social learning can marginally increase the popularity of widely shared songs, co-consumption utility seems to be a great deal more potent in driving users to consume these songs. Conversely, social learning acts as a powerful force for some users to sample less popular songs that nevertheless are recommended to them.

Understanding these two types of endogenous social effects is particularly relevant in the context of a digital test market. Test markets are typically designed to reveal which products or which product features users like most. As test markets are more frequently deployed in digital spaces with integrated social features, social influence may become a more prominent force in determining outcomes in such markets. If the social effects are unique to the market, then practitioners need to respond accordingly. Insofar as social learning drives consumption in such a market, then this implies that the market outcomes can be taken at face value and the most popular products skimmed from the top and taken to broader markets. However, insofar as co-consumption utility is driving consumption on such a market, then the marketing practitioner needs to consider whether social interactions in broader markets will reflect the social interactions on the test market. If not, then it is likely that the marketing practitioner should consider investing in a broader selection of the winners in the test market to better ensure that highly valued products are brought to market.

Simultaneously handling heterogeneous consumer preferences and large, diverse product catalog while identifying and measuring multiple endogenous peer effects in a choice model with a basic learning protocol is still a challenging task for the modern researcher. First, individual level panel data on product consumption is needed because learning occurs over time at the individual-product level. Second, this panel needs to cover a wide range of users to enable the researcher to leverage correlation across users for various products where otherwise the overlap in consumption with such a large body of products would be minimal. Because social effects work through social interactions with other individuals, the specific social connections between individuals are required. Further, on a platform where most activity taking place is private in nature (listening to music), data on which information traverses the social network (peer recommendations) and which information remains private is needed to help identify the difference between homophily and genuine peer effects. Access to such granular, detailed data is rare; however, SoundCloud has provided just such a data set, which, short of experimentation, is nearly ideal for the questions at hand.

Utilizing this rich dataset, this papermakes four primary contributions. First, it outlines how different kinds of social effects induce different outcomes in test markets. Second, it proposes a method to separately identify and measure these two separate endogenous social effects, taking advantage of knowledge of the network structure and the flow of social signals through the network. An estimation framework is introduced that implements the identification strategy and estimates co-consumption utility and social learning in a choice model that can handle much larger choice sets than the standard choice models, scaling to hundreds of thousands of products. An implementation of that framework is used to estimate the model on data from SoundCloud, and I find that co-consumption utility seems to be a dominant force, relative to social learning, for song success on the platform.

The rest of this paperis organized as follows: Section 2 discusses the relevant literature; Section 3 outlines the model and identification strategy, going into detail on how to handle the fact that cultural markets like music have high degrees of heterogeneity; Section 4 details how to estimate the model on large social network data with large choice sets; Section 5 describes some of the key features of and patterns in the data; Section 6 contains the estimation results and counterfactuals; and Section 7 concludes.

2 Digital Test Markets, Social Effects, and Relevant Literature

This papercontributes to three literatures. A long literature on new product development and introduction has developed in both economics and marketing, previewed in the introduction. This literature has approached a wide variety of questions related to new product development and release strategy, including methods for organizing teams to build optimal new products (Hauser and Dahan 2008; Toubia 2006), assessing expected product demand given well defined features (Green and Srinivasan 1990; Netzer et al. 2008; Cao and Juanjuan Zhang 2017), strategizing product deployment to maximize profits (Hitsch 2006), and understanding the processes for the diffusion of product adoption (Chatterjee and Eliashberg 1990). I identify a new complication in the interpretation of realized product demand on digital test platforms, namely peer effects, and provides a method for understanding different forms of peer effects. A broad and deep literature on the importance of social effects in understanding individual level behavior has been pushed by sociologists and adopted and extended by economists and marketers (Duncan, Haller, and Portes 1968; Akerlof 1980; Crane 1991; Case and Katz 1991).

In marketing, this literature has identified the importance of understanding co-consumption experiences in estimating demand (Hartmann 2010). It has also identified how learning from peers can be influential in the spread of new innovations (Nair, Manchanda, and Bhatia 2010). What has been less recognized is that these two types of endogenous social effects may co-exist and may induce different interpretations of realized demand in the presence of peer effects. This paperexplains the salience of these two effects and shows how to separately identify and measure them.

Finally, there is a growing literature in machine learning and applied economics attempting to integrate new machine learning methods, especially those from the collaborative filtering literature (Gomez-Uribe and Hunt 2015; Koren, Bell, and Volinsky 2009; Hernando, Bobadilla, and Ortega 2016) with microfounded models for counterfactual analysis (Ruiz, Athey, and Blei 2017; Athey et al. 2018). By utilizing a latent space created by a collaborative filter to control for static unobservables related to choice and positing a stylized model of individual learning in that latent space, this paperproposes one more area where these methods can be applied to substantive questions of interest in economics and marketing.

2.1 Progress and Pitfalls in Measuring Peer Effects

Some early econometric literature on the identification of peer effects was rather pessimistic about the possibility of identifying peer effects (Manski 1993). However, Manski was prescient, realizing that information on the structural connections in the network would allow for more productive avenues for identification and called for collecting data of this nature. As more data on specific social structures among individuals in a network became available, initially through surveys, and in the past 15 years through the proliferation of digital social platforms, sociologists, econometricians, economists, and marketers have built models leveraging these sources to identify and estimate endogenous peer effects (Brock and Durlauf 2003; Bramoullé, Djebbari, and Fortin 2009; De Giorgi, Pellizzari, and Redaelli 2010; Blume et al. 2015). The models proposed in the most recent iterations of this literature are typically linear models with a static social network. The fundamental insight is that knowledge of the structure of the social graph, who is connected to whom, allows the econometrician to use the variance in the social connections to identify strength of social influence. These papers assume that social influence flows to a person only through those with whom they are connected, an assumption adopted in this work. Thus, the network itself provides an instrument: mutual friends. Friends-of-friends have influence on people only through mutual friends; thus, the behavior of friends-of-friends can be used as an instrument for the behavior of friends. While one degree of separation is typically assumed for the exclusion restriction, it can theoretically be relaxed to require more degrees of separation, though this does run into issues relating to weak instruments.

An assumption that is required by these models, however, is that the network is exogenously determined. The models do not speak to the question of endogenous network formation and why people choose to connect with whom they do. While a number of approaches have been proposed to model network formation, such models are typically geared toward analysis of the network itself and not behavior on the network (Tomasello et al. 2014). A full model of behavior on a network, including both tie formation and social interactions in choice contexts, remains unavailable. Thus, researchers are still left without authoritative guidance on how to empirically model social interactions and the network formation process. This paperthus follows the majority of the empirical literature in not addressing the question of endogenous network formation.

Despite the recent accomplishments in this literature, care still needs to be taken as model mis-specification or omitted variable bias can mislead researchers into detecting social effects when in fact none exists (Van den Bulte and Lilien 2001; Aral, Muchnik, and Sundararajan 2009; Shin, Misra, and Horsky 2011; Iyengar, Van den Bulte, and Lee 2015). Of particular concern is the fact that users who choose to connect to one another will be similar in ways the econometrician may not observe, i.e. unobserved homophily. This paperuses a technique from the recommendation systems literature to estimate user locations in a latent space and then uses those locations to control for the unobserved homophily.

2.2 Peer Effects Make Outcomes Sensitive to Initial Conditions

Experimental and simulation studies have found that peer effects have a strong influence on the song rankings in an artificial music market; while the best and worst songs were typically found near the top and bottom of the distributions, the results were otherwise highly contingent on the social effects, especially with randomly seeded initial conditions (Salganik, Dodds, and Watts 2006). Salganik et. al. implement social influence by showing users how many users have liked a song in an artificial music market. They find that this approach allows them distort the relationship between song quality and popularity in the market by changing the initial popularity rating of songs. What this paper does not show is the nature of the user response to that information. Are users inferring that the song is high quality (albeit, potentially incorrectly) due to its popularity, or are users simply interested in following the crowd? By specifying a more structured model of choice that can separate these two effects, I try to understand which phenomenon is driving this result. The counterfactual analysis will suggest, in a world where a platform may be able to influence whether social learning or co-consumption dominates the social interactions, whether one of these effects can better align underlying song quality with popularity.

Even on platforms such as SoundCloud where the nature of the social ties may not be strong because the social ties are on-line only and do not represent a "real world" connection, the existence of the ties, their non-negligible influence, and their large number suggest they are still potent enough to have a strong influence on outcomes on the platform (Bakshy et al. 2012). Thus, as the platform, artists, and labels consider moving these products to new contexts where leveraging the current state of the social effects may be difficult or impossible, knowing the role social influence played in generating success and whether the social influence is likely to be replicable in other environments can be critical.

2.3 Different Forms of Peer Effects Result in Different Outcomes

An important part of understanding the replicability of social influence is understanding the underlying mechanism that constitutes people's response to the social signals around them. Analytic models have found that different forms of social effects have different growth paths in the adoption of innovation and some may not break into an acceleration phase (Young 2009). These models make clear that different types of social effects, e.g. social learning versus social influence,[1] may generate different outcomes.

Specifically, in the context of this paper, if a product is successful primarily because of social learning, then marketers do not need to rely on the social network to replicate that success in other environments, though if they are able, there is likely still an advantage to doing so. This is simply due to the fact that products successful in such an environment are successful because of the underlying preferences consumers have for these products, and the social interactions worked to reveal the preferred products to consumers. This suggests that marketers can make use of traditional marketing tools to inform consumers and should be able to expect similar market outcomes as previously observed on the incubator platform. However, if products are successful due to co-consumption utility, success may be more difficult to replicate outside of the test market for two primary reasons. First, because products may have become popular in substantial part due to consumers' utility for the social interactions they experienced in consuming the product on the platform with their peers, the underlying private utility for these successful products may not be as consistently high. Secondly, no reliable mechanisms for translating social interactions from one context and social network to another has been identified. Thus, the marketer may need to re-establish the positive social interactions underlying the co-consumption utility that made the products successful in the first place, which may be an expensive or a futile exercise.

The importance and limitations of social learning in various product adoption and innovation scenarios has been an area of intense interest across a number fields (Ryan and

[1] "Social influence" is a term used by Young that they define as "innovations spread by a conformity motive". While a conformity motive is compatible with co-consumption utility as I model it, I remain agnostic as to the underlying psychological mechanisms that may underpin co-consumption utility.

Gross 1943; Duflo and Saez 2003; Conley and Udry 2010; Golub and Jackson 2010). The failure of social learning to properly inform people about product quality has also been well documented (Juanjuan Zhang 2009). Similarly, a number of authors have found the presence of effects similar to co-consumption utility to be a strong driver of demand or advertising effects (Schultz et al. 2007). What has been less studied are the co-existence of these two effects and how to jointly estimate them. However, related concepts have been studied, with findings indicating that different types of connections can produce different learning outcomes (Jurui Zhang, Liu, and Chen 2015) or that there appears to be different forms of social effects identifiable in data (Iyengar, Van den Bulte, and Valente 2010) and that social effects seem to be different for trial versus repeat use (Iyengar, Van den Bulte, and Lee 2015). Some recent experimental approaches have been used to separate social learning from other endogenous social effects in a setting with relatively well defined product features (finance) to better understand the nature of herding in financial markets (Bursztyn et al. 2014). Social learning has also been isolated from other peer effects in the study of the spread of women's protest driving the Temperance Movement of the 1870's (García-Jimeno, Iglesias, and Yildirim 2018). However, separately identifying two different types of endogenous social effects has not been accomplished in a setting with highly heterogeneous products, where the identification of heterogeneous user preferences is all the more important. This paperattempts to undertake that task and explores some implications for interpreting outcomes of new product introductions.

3 Model

This section describes a model of social interactions on a digital platform with a large number of products and how the model is identified. As described in the introduction, the data for this paper come from SoundCloud, a music sharing and streaming service designed with independent musicians and their followers in mind. The key features of the platform that are relevant to the question of interest and the model revolve around its embedding of social information. The platform allows users to follow other users (a unidirectional social link), whether they are an artist who uploads music or simply another user. Users can recommend their followers music they like by "reposting" a song, which is akin to a retweet on twitter. When a followed user makes a repost, their followers see that song in their news feed, including who reposted the song and information about the song, such as the title, artist and genre. The other kind of information that enters the newsfeed are songs that are uploaded by followed users. The model specification proceeds with this in mind.

3.1 Model Skeleton: Fundamentals of Separating Social Learning, Co-Consumption

Consider a user i deciding to consume song s at time t. Let $n(i,t)$ represent the set of people that i follows up to time t and $R^t_{n(i),s}$ is the log number of reposts of s in i's network in the history up to time t. P^t_{is} is the number of plays by i of s by time t. β^*_{is} is the posterior value of song quality absent any social effects, the "solo value", while β^0_{is} is the prior. m is a hyperparameter set by the econometrician indicating the number of times a user has to listen to a song before they learn their posterior value for that song, which I set to 1.

The choice of what song to listen to is the song that maximizes:

$$u_{ist} = \gamma(R^t_{n(i),s}) + \varepsilon_{ist} + \begin{cases} \beta^0_{is} + \lambda(R^t_{n(i),s}) & : P^t_{is} < m \\ \beta^*_{is} & : P^t_{is} \geq m \end{cases} \tag{1}$$

This utility specification allows others' influence to enter into the decision process in two different ways. Namely, a recommendation from a peer can be responded to as an update about beliefs about product quality, captured by λ, which become fully resolved through consumption, i.e. social learning and experiential learning. Once the product is experienced, further social signals cannot be influential through an informational effect because the product experience fully resolves the information about the product; thus, if these further signals are influential, they must be operating through another mechanism. One such mechanism is that a recommendation can be responded to as a coordination signal that a product is being consumed by one's peers and that consumption of that product will allow the person to reap the extra utility from knowingly consuming the product with others, i.e. co-consumption utility. Co-consumption utility is captured above by γ.

The model mirrors Ackerberg 2001, which separates the informative and prestige effects of advertising in a similar modeling framework. Ackerberg's model separates the informative and prestige effects by claiming that the prestige effect of advertising is the only potential response to advertising after the consumer learns their preferences via consumption, using a one-shot learning assumption. Before the consumption experience, advertising potentially has both an informative and a prestige effect, and that the variation in the effect of advertising before and after consumption of the product is what allows identification and measurement of the two effects. This model will work analogously.

The essential insight of this model is that the data can be segmented along two dimensions into 4 boxes in order to identify the parameters of interest, seen in fig. 1. For the moment, let's simplify the discussion by considering one song and a set of homogeneous users, thus dropping the song- and user- specific parameter subscripts. Running counter-clockwise in the table starting in the upper left, the identification argument is as follows:

Figure 1: Data Divisions that Identify Parameters of Interest

		Number of Reposts Seen	
		= 0	**> 0**
Number of Song Listens	$< m$	Prior on song solo value	**Social Learning** + CC + Prior
	$\geq m$	Learned solo value	**Co-Consumption** + Solo Value

- The first dataset contains all users who have never listened to song s and have not been exposed to any reposts of s. The propensity of users to sample the song in this dataset identifies their prior, $\beta^0 = \beta_i^0$.
- Next, consider users who have previously listened to the song at least m times but still have not been exposed to any reposts. These users have not been socially influenced but have learned their solo value for the song. Thus, their propensity to listen to the song identifies that solo value, $\beta^* = \beta_i^*$.
- Next, consider when all users have previously consumed song s and are also exposed to reposts of song s. In this world, any change in propensity to listen compared to β^* identifies the co-consumption parameter γ.
- Finally, consider a fourth dataset where users have not listened to the song but have been exposed to reposts of the song. Because we know β^0 and γ, any change in the propensity to listen in this scenario identifies λ, social learning.

The assumption of homogeneity required to make this argument is unpalatable in a market as complex as music. Thus, in section 3.3, a high dimensional latent space that captures the high dimensional complexity that exists in user preferences for music will be introduced. After producing the user and song locations in this space, the model can then condition on each user vector of preferences and each song vector of features. By conditioning on these vectors, the homogeneity assumption can be relaxed into a conditional homogeneity assumption.

Another concern is the non-random flow of recommendations to users as the identification argument assumes that exposure to reposts is effectively random. Given that other users will be making recommendations to their followers in a non-random fashion, this is a threat to the strategy. However, again the latent space can help develop the necessary conditions for identification. The assumption required is that reposts are made randomly, conditional on the user's preference vector, the preferences of those in their network, and the features of the song. Before discussing the latent space and how it specifically enters the model, a discussion of the potential for strategic interactions and how those may affect the model follows.

3.2 Strategic and Non-Strategic Behavior in Models of Social Interactions

The purpose of this model is to enable large scale micro-founded choice estimation of highly heterogeneous units, the consumption of which is substantially influenced by peers. Particularly in economics, social effects are frequently modeled as a game, with the action of some agent i entering into the decision of another agent j and vice versa. In many scenarios, the necessity of such a structure is clear. For example, if two people wanted to play soccer together, they need to coordinate their schedules with one another to show up at the same time and place. If they want to play in a full scale game against a new team each week over the summer, they likely need to create a mechanism, some semi-formal league, that enables participants to schedule the games and make sure there is space for the game to be played. These are both examples of social interactions with coordination mechanisms. If the players were to show up at the field at the time of their convenience without taking other players' behavior into account, it is unlikely they would be able to play the game.

In a simultaneous decision structure, a feedback loop, or social multiplier, is created when the effect of the other agent's action on the ego's action is in the same direction for both agents, with the strength of the multiplier depending on how strongly the other's actions enter the decision. However, in a sequential scenario, this logic breaks down unless agents are assumed to be forward looking in their decisions.[2] Adopting such an assumption in this context is problematic from both substantive and methodological perspectives.

Substantively, in many scenarios, forward-looking behavior is to be expected; however, in music listening and other hedonic scenarios, particularly when budget constraints are essentially limited only by the time it takes to consume something, it becomes difficult to justify forward-looking behavior. That is, there is a cognitive cost associated with the anticipatory behavior with unclear payoffs. Perhaps the strongest argument for considering forward-looking behavior in social interactions on such a platform is that there are a number of people who are looking to self-promote, i.e. they are musicians looking to grow their audience, and to achieve that goal would require forward-looking, strategically thinking

[2] An accessible discussion of these modeling details can be found in Hartmann et al. 2008.

behavior about when and what to recommend to others.[3] Two facts help alleviate this concern, however. One is that self promotion is not possible on the platform through the social features. SoundCloud disallows users to repost their own material, and further, SoundCloud has made efforts to ban accounts that appear to exist for the purpose of promotion. Secondly, by the beginning of the data under examination, the vast majority of people on the platform are simply listeners. In addition to this, given the enormous number of listeners on the platform, it would be unrealistic for a user to believe that they can systematically change the equilibrium outcomes with the single recommendation given to their followers.

But even if forward-looking behavior was a reasonable assumption to make, there remains the methodological problem alluded to above: our tools for modeling decisions with a large number of states and a large number of strategic actors in a dynamic context are still lacking. Substantive progress has been made in handling a large number of actors through the application of the oblivious equilibrium concept; however, this approach still requires a relatively small number of states to be used (Weintraub, Benkard, and Van Roy 2010). In this context, the states to consider would be song or artist specific, e.g. the level of observed sharing per song, implying an explosion in the number of states for all but the most restrictive product assortments. Recalling that the curse of dimensionality tends to limit the number of computable states to half a dozen or so, trying to model people's choices across hundreds of thousands of products with a state or two per product is not a realistic approach. Given that the research question and its motivation is directly concerned with whether observed outcomes in the top of the distribution reflect underlying private valuations, attempting to bypass the large number of states by selecting a small sample of songs to use in a more traditional approach is untenable as there is not a good basis on which to select such a small sample of the songs. A small random sample is exceedingly unlikely to produce a set that represents any successful songs that benefit from social multiplier effects. Since there is an expectation of variance at the song level in the nature of the social effects, a reasonably large sample that is representative of the platform but also captures songs with such effects is required.

While still tracking a large number of states, my approach makes reasonable assumptions about the dynamics of social interactions while avoiding the need to solve the Bellman equation or other non-closed form functions in the space of those states, thereby avoiding an exponentially exploding computational cost.

[3] Forward-looking behavior would also be justifiable by considering the nature of variety seeking by users on the platform. As a platform for experimental music production with an enormous product catalog, a high degree of variety is nearly guaranteed for any given sequence of products consumed, making detailed consideration of such behavior likely a second order concern.

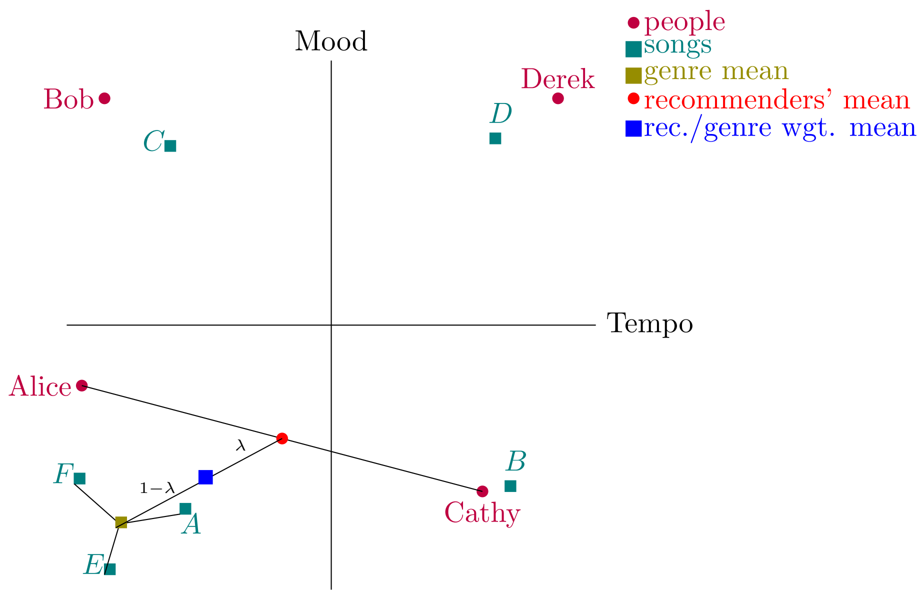
3.3 Latent Space to Capture Heterogeneity and Represent Learning

The argument made above for separating social learning from co-consumption rests on the assumption of homogeneity, which is hard to accept in this context. While a separate user and song fixed effect approach may be feasible in some contexts, it is rather unsatisfactory when discussing music and many other creative contexts. An essential feature of music preferences is that some songs appeal very differently to some users compared to others. Absent the existence of well-defined product features that relate strongly to consumer choice, our usual models of heterogeneity, random or fixed effects in users and separately in products, do not allow for flexibility in the preferences users have for songs at the user-song level. Unfortunately, relaxing this assumption to allow a high degree of flexibility where each user is allowed to have a preference for each song independent of other users would cause parameter growth on the order of the number of users times the number of songs. This is an unreasonable number of parameters to try to estimate, and in any case it is difficult to imagine a dataset where all those parameters would be sufficiently informed by the variation found in the data. The typical platform dataset is characterized by users sampling a small fraction of the number of products available. For example in this dataset, the average user listens to 1000 songs out of the 426,000 songs listened to by other users in the sampled network, not to mention the millions of other songs on the platform that no one in this network consumes.

The model skeleton also forces recommendations to enter the model in a purely additive manner for both learning and co-consumption. That is, each recommendation adds λ and/or γ to the expected utility of consuming the song. However, it is natural to assume that users trust some recommendations more than others. If we think to our own social network, song recommendations from different colleagues and friends would carry different weight depending on what we know about their music taste and how it aligns with our own. This is an important feature of social learning that should be captured where possible and is approximated in the model by taking into account the correlation in preferences between users, as explained further below.

Heterogeneity of user preferences for song features is modeled using a latent space. Specifics about the construction of the latent space that is implemented here will be discussed in section 4, but that describes one particular implementation that could be replaced with other structures. The model requires only that we have a way to construct for each user a vector representing their latent preferences, for each song a vector representing their latent characteristics, and a function, in the model the scalar product, to compute the match value between the user and song vectors. This latent space introduces a flexible pattern of heterogeneity to the model. Heterogeneity is captured by the interaction of the song and user vectors rather than looking at users divorced from songs or vice versa as would be the case in a feasible fixed or random effects approach to heterogeneity. The latent space lends itself to a richer story about how users may form and adjust their priors

Figure 2: Learning Process: Weighing Genre Prior versus Recommendation



The axis labels here are not generally possible to generate but are included to simplify exposition and aid the reader in thinking about how a latent space may look.

based on recommendations. The diagram in fig. 2 represents the behavioral assumption the model makes about how users resolve information about the genre of a song and user recommendations to generate a prior for the location of an unconsumed song.

This picture is analogous to ideal point models such as Goettler and Shachar 2001, which also utilize latent spaces to control for heterogeneity. Section 4.1 describes the difference between those ideal point models and this latent space. In fig. 2, people are represented by dots and songs by squares. A latent vector for each user and each song places both users and songs in the same space, i.e. they are interchangeable.

How might listeners operating in such a space construct priors about songs they have not had prior exposure to? Users have a common location belief determined by the mean characteristics of the top songs in the genre, with the motivation that these top songs are most likely to represent what an average user might know about the genre without having delved more deeply into the broad collection of songs that constitute the genre. This is represented by the point at the center of the A, E, F cluster in the lower left quadrant. Note that this is a shared assumption, across users, of the genre location in the space, while individual specific preferences for the genre come out of their own location in the space: Alice has a relatively high match quality for the genre represented by the AEF cluster compared to Derek, who has a low match quality with that location compared to whatever cluster of songs might exist near him in space. However, they have the same belief in the location of the genre.

When a user recommends a song, it suggests that the song lies near them in the space; without further information, the behavioral assumption is that users perceive the recommendation to suggest that the song is located at the recommending user's location. Normally, a user would not know where other users lie in space; however, when a user decides to follow another user and therefore get recommendations from that user, this suggests that they are sufficiently familiar with the user's tastes, i.e. location, to want to follow them in order to get recommendations from that space of tastes. However, when two users recommend a song to another user, that user has to adjudicate where in space that song is located. The same assumption made for songs in a genre above is used: the song is centrally located in characteristic space between the recommending users. In the figure, this is represented by the point halfway between Alice and Cathy, who are making recommendations for a song.

Then, users have to adjudicate where in space between their genre prior and the information from their peer recommendations suggests a song lies. Users do this by choosing some point on the segment between the two points parameterized by λ, and their propensity to sample a song given the genre prior versus. In spirit, this is the same λ as in the model skeleton above; however, in the model skeleton, recommendations adjust utility up and down based on the number of recommendations directly. Here, $\lambda \in (0, 1)$ allows the econometrician to rationalize whether the user is behaving as if they trust recommendations more than their prior. This value is what determines the strength of social learning on the platform.

If a user has listened to an artist before, when they listen to a song never before listened to by the same artist, it is likely that their beliefs about that song are formed more by their experience with the artist rather than the genre associated with the song/artist. Genres are typically large, diverse categories. While artists' music can change over time, it typically does so slowly and the time span of my data is sufficiently low (only two years) that it seems unlikely that most artists are making dramatic changes in the type of music they produce.

Thus, each user's play history is used to construct artist specific priors by averaging the characteristics of the songs by an artist in that user's play history. This looks exactly like the genre prior in fig. 2, but is user specific. It only comes in to play when the user has previously listened to the artist before and changes over time as the user listens to more songs from the artist.

3.4 Modeling the Above

With the purpose of the latent space explained, let's re-build the model with the latent space in mind. To do this, some more notation is necessary. Let's call a user's feature set from the above space ψ_i and a song's feature set ϕ_s.[4] First, match values will be represented by various m terms:

- A user's realized posterior match value after listening to a song is: $m_{is} \coloneqq \psi_i \cdot \phi_s$,
- The prior on the match value is: $m_{i,g(s),t}(\lambda) \coloneqq \psi_i \cdot \bar{\phi}_{(i,g(s),t)}(\lambda)$,

where $g(s)$ represents the genre of song s and may be replaced by $a(s)$, the author of a song. The λ here is a parameter that responds to how intensely a user prefers the genre location information versus the reposters' location information, precisely as specified in fig. 2. The λ is dropped when there are no reposts on which to update.

The individual user and song vectors are used to construct the following terms:

- The prior on genre is based on the top 50 songs in the genre: $\bar{\phi}_{g(s)} \coloneqq \frac{1}{50} \sum_{s' \in g(s)} \phi_{s'}$,
- The user-artist specific prior based on each user's play history. The prior on an artist, conditional on having listened to at least one of the artist's other tracks before is the average over the previously listened songs' characteristics: $\bar{\phi}_{a(s)} \coloneqq \frac{1}{|P^t_{i,a(s)}|} \sum_{s' \in P^t_{i,a(s)}} \phi_{s'}$,
- The same average can be constructed from the set of users who have reposted a song to user i as of t: $\bar{\psi}_{r(i,s,t)} \coloneqq \frac{1}{|r(i,s,t)|} \sum_{i' \in r(i,s,t)} \psi_{i'}$,
- Finally, the genre or artist prior is updated with reposts, parameterized by λ: $\bar{\phi}_{i,g(s),t}(\lambda) \coloneqq \bar{\phi}_{g(s)}(1 - \mathbb{I}^R \lambda) + \mathbb{I}^R \lambda \bar{\psi}_{r(i,s,t)}$.

This leaves the full utility specification as:

$$
\begin{aligned}
u_{ist} = {} & \gamma \log(1 + \text{Repost Count}_{is,t}) + \alpha \text{Track Age}_{st} + \theta \mathbb{I}(\text{Top Track}_{st}) \\
& + \tau log(1 + \text{Last Played}_{ist}) + \varepsilon_{ist} \\
& + \beta \begin{cases} m_{is} & : \text{played track before} \\ \delta_a m_{i,a(s),t} & : \text{have seen artist's other songs, no reposts} \\ \delta_a m_{i,a(s),t}(\lambda_a) & : \text{seen artist, track reposted} \\ \delta_g m_{i,g(s)}(\lambda_g) & : \text{seen neither track/artist, seen reposts} \\ \delta_g m_{i,g(s)} & : \text{seen neither track/artist, no reposts} \end{cases}
\end{aligned}
\tag{2}
$$

[4] This diverges slightly from much of the literature on matrix factorization, where these matrices will typically be called W, H or U, V. However, my notation, which is similar to Wan et al. 2017, makes clear that these are parameters that are to be estimated.

What has been referred to as "the latent space" is the set of vectors ψ_i for users and ϕ_s for songs. Their dimension and specific values are estimated based on their out-of-sample prediction properties as explained in section 4. They are considered latent because they cannot, in principle, be related to specific features of the users or songs, though of course various forms of cluster analysis can used to make further estimates of which latent dimension relates to some specific song or user feature. However, if data is available to do such an analysis, it likely makes more sense to include that data directly in the model where practical. These vectors are considered "latent" in the sense that the researcher cannot give a specific definition of what those factors relate to.

β captures the importance of match value in utility space in determining a user's propensity to choose a song. The $\delta_a, \delta_g \in [0,1]$ allows β to be scaled down to account for the fact that users may put less weight on their prior compared to the learned posterior. This captures the strength of belief in the prior and is allowed to differ for genre versus artist priors. Note that this is not a per-artist or per-genre δ but rather a δ for all genres. There is another δ for all artists. These parameters capture the average strength of a genre prior and the average strength of an artist prior; they do not capture the heterogeneity across genres or the heterogeneity across artists in the strength of their priors on users. $\lambda_a, \lambda_g \in [0,1]$ captures how much weight is placed on the genre or artist prior compared to the information provided by their peer recommenders, when the prior information is artist or only genre specific. The closer λ is to 1, the more weight is placed on reposter information compared to the other source for the prior.

In addition to the latent space, a few terms here that did not show up in the model lacking heterogeneity include: track age, whether the track is a top track, and the log of the amount of time since the last play of the song. There is probably some state dependence in song listening choice such that a song that was just listened to a second ago is unlikely to be listened to again immediately afterwards, despite having a high match value.

Track age is an attempt to capture that there might be more off-platform discussion of newer tracks that might direct users to listen to the track for reasons that the econometrician cannot otherwise to observe. While SoundCloud does not have a recommendation engine, they do have a weekly top 50 chart that presumably users check to see if they have missed interesting music in their preferred genre(s). Thus the top track indicator is included that tries to reconstruct this top 50 list on weekly basis using the full database of billions of individual plays for several million users.

The advantage of casting everything in terms of m instead of the underlying scalar products, besides being a bit easier to read, is that it makes explicit the notion that the scalar product between user and song features is an implementation detail. While one method for learning the space in this paperis used, the machine learning and recommendation systems literature is full of other approaches. Matrix factorization has been oft-used due to its computational efficiency, usual ease-of-implementation, and generally high performance in out-of-sample tests.

3.5 Identification: The Latent Space, Heterogeneity, and Homophily

The problem of unobserved heterogeneity in biasing the estimation of choice models and models of social interactions is well understood. Without appropriately accounting for such heterogeneity, systematic differences and similarities between users not taken into account by the econometrician end up being endogenous factors that bias estimation. The idea behind using a latent space is that the econometrician backs out latent factors that are otherwise unobserved that predict user choice.

With the high-dimensional "observed" latent space captured for both users and songs, the arguments above regarding the identification of the parameters continues to hold. The latent factors give us "data" on which we can condition to implement the conditional homogeneity assumption. Because these factors are estimated at the individual level, they further account for correlation in preferences across users. Thus, bias due to correlated preferences among users, homophily, is being captured. This is further bolstered by the fact that the model does not rely on correlated outcomes, that is plays, between socially connected users to judge whether social effects are at play. Instead, the model is only trying to adjudicate whether a response to a recommendation causes a change in users' beliefs or acts as a coordination signal that users respond to.

However, other threats to identification remain. Specifically, unobserved, time varying, network specific shocks, such as when a group of users go to a concert and a new band plays may contaminate my results. The effect of such time varying shocks on the estimated parameters is ambiguous. Recall that what is being judged here is the relative strength of social learning versus co-consumption in the decision to adopt and continue listening to a song. While these time varying shocks are likely to cause an overestimate of the parameters summarizing social learning and co-consumption utility, whether co-consumption utility or social learning are more affected depends on the timing and sequence of plays and reposts that occur in response to that shock. If the shock causes a large wave of plays, then reposts, then more plays, this will increase the estimate on co-consumption utility. However, if the shock causes a smaller wave of plays but still enough reposts to reach a reasonably sized audience who then make plays, this will push up the importance of social learning relative to co-consumption.

If behavior around recommendations is more strategic and forward-looking than is modeled here, this implies a varying but likely consistently positive continuation value that is shared across all states. Thus, parameters that enter into utility across all states, such as β are likely to be inflated whereas the state-specific influence is likely to be deflated, specifically the parameters associated with social learning and co-consumption utility will be underestimated. Again, whether social learning or co-consumption utility will be more affected is not clear.

4 Estimation Methodology

The model is estimated in two stages: (1) the latent space and (2) using the learned latent space as an input in to the choice model.

4.1 Learning the Latent Space

A number of approaches have been taken to generate latent spaces to place users and products with varying degrees of success. Indeed, latent spaces are not new to the marketing literature, for example, ideal point models such as those mentioned above. Two primary differences exist between that model and what is explicated below: the use of (sometimes weighted) Euclidian norms and an effort to integrate out unobserved heterogeneity that remains due to the lack of underlying variation in their data along the dimensions assumed by the model. The integration is an expensive operation that the machine learning community has typically eschewed. Instead, they have favored higher dimensional spaces estimated using out-of-sample prediction accuracy to estimate the vectors of interest. While this approach is not always theoretically motivated, it often performs well in prediction exercises and has become widely adopted.

In high dimensions, it is critically important not to use standard notions of distance, such as the Euclidian norm. The intuitions that are typically used to build models based on our usual notion of distance do not tend to hold in higher dimensions. As an example, for many distance metrics in high dimensions, the nearest and farthest point have approximately the same distance to a target point, rendering comparisons using the distance metric essentially an exercise in random draws (Beyer et al. 1999; Aggarwal, Hinneburg, and Keim 2001; Domingos 2012). Indeed, one corner of research in machine learning is exploring ways to generate better performing distance metrics (Yang 2006; Kulis 2013; Bellet, Habrard, and Sebban 2013). However, the mainstream practice is to use the scalar product, or transformations of that scalar product such as with the logistic sigmoid, $\sigma(x) = \frac{1}{1+e^{-x}}$, to produce match values.

One of the approaches that has found a great deal of success in generating solidly performing rankings is the method adopted here: Bayesian Personalized Ranking (BPR) by Rendle et al. 2009. It uses revealed preference data to generate a user specific ranking of the items in the dataset. The input to the algorithm is the (extremely sparse) matrix of who played what song, and is thus an $N \times S$ matrix. The goal of the algorithm is to learn a set of k features (a hyperparameter decided on by the researcher) for each user and each song. In the matrix factorization implementation of the algorithm used here, the scalar product of the user features, ψ_i with song features ϕ_s gives the relative likelihood of that song being chosen for that user compared to other songs. Thus, it is primarily a ranking algorithm, with the ranking being specific to each user. Multiplying the learned user vector times the matrix of song features creates another vector with the relative match values for that user with the indexed song. The ordering generated by these match values generates the personal ranking for that user of all the songs in the catalog.

At a high level, BPR suggests to repeatedly update a set of user and song parameters according to a binary logit model for a single user, a song they listened to and a song they did not listen to. The update steps are randomized across different users and songs so that each user and song gets updated sufficiently to achieve a good result, namely out-of-sample prediction.

More technically, the features are produced by maximizing a logit-like likelihood function with a normal prior producing a Ridge regularization:

$$\mathcal{L} = \sum_{i,s \in S_i^+, s' \in S_i^-} \log \sigma \left(\psi_i \cdot \phi_s - \psi_i \cdot \phi_{s'} \right) - \lambda_\Theta \left\| \Theta \right\|^2$$

where i is a user, s is a song consumed by that user, and s' is a song not consumed by that user. For each user, the song catalog, S, can be split into songs consumed by that user S_i^+ and songs not consumed by that user S_i^-. Each user is characterized by a vector ψ_i and each song a vector ϕ_s, which are randomly initialized. The set of all parameters is $\Theta = \{\{\psi_i\}_{i=1}^N, \{\phi_s\}_{s=1}^S\}$, and λ_Θ represents the regularization terms.

The likelihood function is estimated with stochastic gradient descent (SGD). The algorithm is to repeatedly draw an (i, s, s') triplet at random, and to update the parameters ψ_i, ϕ_s, and $\phi_{s'}$ with three separate regularizers λ_i, λ_s, and $\lambda_{s'}$. [5]

The three separate regularizers are used because the amount of information the algorithm learns about the user, the song they listened to, and the song they did not listen to will be

[5]The hyperparameters, α, the three regularizers λ_i, λ_s, $\lambda_{s'}$, and k, were tuned on a sample drawn from the larger dataset independently of the sample used to estimate the model, with the model fit evaluated on a hold-out set from that independently drawn sample. This should prevent overfitting. The algorithm is run for 50 million iterations. Note that algorithms estimated with SGD do not converge per se. Due to the noise inherent in using a stochastic process, any given draw will deviate from the mean established by the algorithm. Nevertheless, after the iterations, the algorithm is allowed to continue to run until the average change in 4096 evaluations of the objective function (i.e., the likelihood) after the gradient updates is less than 0.00001, helping to ensure that the parameters are in a good state.

Algorithm performance is adjudicated by whether the learned features properly predict that a user prefers the song they listened to over one they did not on a hold out sample of one song from S_i^+ and one song from S_i^- for each user: $AUC = \frac{1}{N} \sum_i \mathbb{I} \left(\psi_i \cdot \phi_{s_i^+} > \psi_i \cdot \phi_{s_i^-} \right)$. The AUC on my holdout sample was 0.93 for $k = 75$.

different, and thus the parameter updates should be regularized differently.[6] With large product catalogs, the fact that a user did not listen to a song is a weak signal about their taste. It is possible that they saw that the song existed and did not listen to it. However, it is more likely that the user was never exposed to the song just because of the sheer size of the product catalog. Importantly, this is not purely random noise that the econometrician can ignore but is generated by the user's behavior on the platform and what they decide to sample. What this implies is that $\lambda_{s'}$ will be larger in magnitude thus forcing $\phi_{s'}$ toward 0 to a greater extent in that update iteration. Thus, to inform the algorithm about the unconsumed song effectively, an update step needs to be performed for someone who actually consumes the song. Therefore, another random draw is required. Eventually, with enough iterations, each song will have its parameters updated multiple times in the case that it was consumed. The algorithm, then, uses information from across users to update the song parameters and across songs to update user parameters.[7]

4.2 Model Estimation: Avoiding the Full Softmax

Using the standard assumption that the ε in the utility specification are distributed as EVT-1 implies a multinomial logit specification. However, despite the reduction in the parameter space by using the latent factors, the enormous choice space still poses a number of problems. The standard logit formulation requires constructing a $S \times P$ vector for the P inputs in the utility function for each song. This implies, for just the 215 million plays under consideration, several petabytes of data. Thus, stochastic methods are required to avoid that memory consumption.

However, stochastic optimization methods don't get me all the way to being able to estimate the model. There remains two facets of the same problem: the large number of items. The first issue here is the fact that the lookup of all the social data for each item is an

[6]The update step is as follows:

$$\begin{aligned}
\psi_i' &= \psi_i + \alpha \left[\sigma(\psi_i \cdot \phi_s - \psi_i \cdot \phi_{s'}) \cdot (\phi_s - \phi_{s'}) + \lambda_i \psi_i\right] \\
\phi_s' &= \phi_s + \alpha \left[\sigma(\psi_i \cdot \phi_s - \psi_i \cdot \phi_{s'}) \cdot \psi_i + \lambda_s \phi_s\right] \\
\phi_{s'}' &= \phi_{s'} + \alpha \left[\sigma(\psi_i \cdot \phi_s - \psi_i \cdot \phi_{s'}) \cdot (-\psi_i) + \lambda_{s'} \phi_{s'}\right]
\end{aligned}$$

As you can see, depending on the relative size of the various λ regularizers, the algorithm will take smaller or larger steps for that set of parameters. Specifically, $\lambda_s < \lambda_{s'}$, set by out-of-sample CV on a separate data sample, implies that small steps are taken for songs not listened to and larger steps are taken for songs listened to.

[7]In practice, the most striking drawback of BPR is that the sampling space, which needs to be thoroughly explored, grows in an exponential fashion with the number of products because the algorithm samples not just songs listened to but pairs of songs, one listened to and one not listened to. Unlike some algorithms whose complexity is linear with number of users and the products they consume, it grows in the number of users and combinatorically with the number of products in the catalog. Thus, larger product catalogs become more difficult to estimate compared to larger user bases.

expensive operation on its own. This requires traversing, joining, and aggregating several tables multiple times for each data point, i.e. multiple non-trivial SQL queries in the inner loop of the estimation procedure. And again, this data also can't be precomputed due to an unreasonable growth in memory required. The second facet of the problem is the computation of the softmax function in the multinomial logit likelihood.[8] The large sum is expensive to compute, and the need to use more accurate summation functions due to the large number of terms increases the computational cost further.[9]

Fortunately, especially due to the extensive use of the softmax in modern machine learning applications such as neural nets, there has been active research in speeding up the softmax computation.[10] Estimation is implemented by adopting one promising approach, the one-versus-each method of Titsias 2016. This method establishes a lower bound on the softmax probabilities implied by the multinomial logit that can be subsampled without bias. While the original paper has an in-depth discussion of the proof of the bounds and the quality of the simulation results in using that bound, the basic result follows below.

The fundamental result is the following lower bound, where the approximation represents an unbiased approximation derived from rewriting the usual sum as the "one-vs-each" sum (line 2) and then applying the fact that for non-negative numbers $1 + \alpha_1 + \alpha_2 \leq (1 + \alpha_1)(1 + \alpha_2)$ and more generally $1 + \sum_i \alpha_i \leq \prod_i (1 + \alpha_i)$:

$$\begin{aligned} p(y = j | x, \beta) &= \frac{e^{x_j \beta}}{\sum_k e^{x_k \beta}} \\ &= \frac{1}{1 + \sum_{k \neq j} e^{-(x_j \beta - x_k \beta)}} \\ &\geq \prod_{k \neq j} \frac{1}{1 + e^{-(x_j \beta - x_k \beta)}} \end{aligned} \tag{3}$$

Applying the log to get the log likelihood yields:

$$\log \prod_{k \neq j} \frac{1}{1 + e^{-(x_j \beta - x_k \beta)}} = \sum_{k \neq j} \log \frac{1}{1 + e^{-(x_j \beta - x_k \beta)}}$$

which can then be estimated by taking a sample of items and scaling the resulting sum according to the number of items sampled compared to the total number of items. Sim-

[8] $\text{softmax}(x_j, \{x_k\}_1^J) = \frac{e^{x_j}}{\sum_{k=1}^J e^{x_k}}$

[9] A naïve summation algorithm has a worst case error that grows with n terms and an average error that grows with $\sqrt{n}$. For the small number of terms in our usual logit models, this is easy to ignore. With a large number of terms, however this can introduce non-trivial inaccuracies.

[10] Although other activation functions are becoming more dominant in that space.

ulation results by Titsias indicate that the relative probabilities are representative of the full logit probabilities, which suggests that we can use this lower bound as if it were the true probability, giving us a quasi-maximum likelihood estimator, replacing $x_j\beta$ with the u_{ist} as described in the choice model, eq. (2).[11]

This subsampling approach solves both problems: the social structure needs to be queried for only a relatively small handful of songs per iteration, and the computation of the probability requires several orders of magnitude fewer operations. Parallelizing the data production can provide essentially linear speedups in the computation of the objective function for a relatively small number of parallel processes.

4.2.1 Weighted Sampling and Consideration Sets

Unfortunately, application of the subsampling technique is not as direct as one might hope. Consider the cases found in eq. (2). In a random sample of tracks, 99%+ of tracks fall in the last case: some song that a user has never listened to from an artist they haven't heard of and that has no associated reposts or otherwise any observed mechanism of exposure to the user. This poses both conceptual and estimation problems.

With these giant product catalogs, it does not make sense to think of users as having been exposed to all the songs and choosing their favorite song subject to their current state. Certainly, some notion of a consideration set should be incorporated into the decision process. Rather than modeling the consideration sets directly, the subsampling mechanism is used in the estimation process to provide an estimation analogue of what would be a reasonable consideration set.

To construct such a consideration set, I try to reconstruct what might be on the user's newsfeed or a small number of clicks away. Using the full universe of data at my disposal, I recompute the daily top charts for each genre to include in the consideration set. Further, previously played songs, other songs from previously played artists including the artist of the song that was in fact played, and songs that have been reposted are sampled to better inform all the cases. A sufficient statistic for a song that a user has no exposure to is the genre vector for the genre of that song. Thus, rather than sampling songs that the user has no exposure to at random, each genre vector is included once to represent the body of all songs in that genre that the user is not exposed to.

While this approach does not fully model the consideration set, it forces the model to

[11] Say that 10 items were sampled out of a total of 1000, then the resulting formula with scaling factor would be $\frac{1000-1}{10}\sum_{k\neq j}\log\frac{1}{1+e^{-(x_j\beta - x_k\beta)}}$.

Note that this equation is the same as the BPR objective up to the scaling factor. In BPR, the algorithm samples one consumed item and one unconsumed item for comparison. By using BPR over other methods of generating a latent space brings the first step in the estimation that much closer to the second.

reconcile why the song that was chosen was chosen versus some other song in a set of songs that might reasonably constitute the consideration set without further data about the user's browsing habit. It has the further advantage that each iteration should create variation in the data that is informative to the parameters in the model.

The transformed likelihood that user i listens to a song s at time t can be written:

$$\begin{aligned}\mathcal{L}_{ist} = &\sum_{s'} \log\left(\frac{1}{1+\exp(-(u_{ist}-u_{i,s',t}))}\right)\\ &+\sum_{g} \log\left(\frac{1}{1+\exp(-(u_{ist}-u_{ig}))}\right)\end{aligned} \tag{4}$$

Estimation is implemented using ADAM as described in appendix A.

5 Empirical Context and Data

To estimate this model, this paperuses data from SoundCloud, a platform developed by independent musicians to enable other musicians to share their work with each other and to try to gain larger audiences and long term fans. The platform, founded in 2007 and launched in 2008, evolved from a niche place for artists to communicate and collaborate with their music, to a large platform with hundreds of millions of unique monthly visitors and approximately 40 million registered, "regular" users that has helped to launch the careers of a handful of successful, mainstream artists. Users only get 30 MB of free space but can pay for more space with a nominal fee. Users do not need to pay for the bandwidth used by themselves or their listeners. Thus, the cost structure is for the platform to bear the cost of hosting content in exchange for opportunities to monetize. However, a monetization strategy came later in the platform's evolution, with ads and sponsorships entering the picture only in late 2014.

The most relevant change to the platform for the empirical exercise, however, is the introduction of the "repost" feature on 5 December 2012. For long before this, a major feature was the ability to follow an artist. This is a unidirectional connection, thus artists can be followed without the artist necessarily following the user back. Note that, as with many other platforms, there is no clear delineation between a generic listener on the platform compared to an artist. The definition of an artist is operationalized as simply a user who uploads content.[12] Before the release of the repost feature, the primary role of following

[12]This approach certainly overcounts the number of users who are artists. However, in the analysis, the artist-ness of a user is only in the context of another user listening to the content of the artist. Thus, users

was to enable users to get updates when another user uploaded new content. Notifications for these updates came in the form of a news feed and, optionally, email updates.

However, the release of the repost feature brought with it the opportunity for a great deal more sharing. With reposts, updates were not only for content uploaded by people a user followed but also included songs that other users liked. Thus, broadcasted peer recommendations were officially brought to the platform. At the time, SoundCloud didn't have anything resembling a reasonable recommendation engine nor advertising, so the introduction of reposts had the potential to substantially increase the number of potentially high value songs a user would be exposed to, if they were to follow the people with the right taste in music.

Unfortunately for the analysis, on 9 May 2012, an invite-only beta of the new site design with this repost feature was publicly announced on the SoundCloud blog. Thus, any attempt to leverage what might otherwise be relatively simple event study designs will be muddied with the presence of contaminated but unidentified users.

The data contain four main tables with which to work: (1) the track list with uploader, date of upload, and user-reported genre, (2) the history of user plays, including user id, track id, and time of play, (3) history of affiliations, indicating who followed whom at what time. If a user removed an affiliation, this was not recorded as a removal but rather the relationship was not shipped. (4) And the history of what each user reposted when.

Absent experimentation, this is a nearly ideal dataset. Specifically, it meets the requirements of allowing an econometrician to observe individual level consumer learning of a large number of products over time and the specific flow of information through a social network.

This latter piece of information is rarely available to researchers. Indeed, many of the estimation routines that have been developed leverage each person's individual behavior with the presumption that behavior is observed by connected users under the assumption such information would not be observed by the econometrician (Bramoullé, Djebbari, and Fortin 2009). In this case, this would mean that the model would have to attempt to identify the social effects through the plays of other users. Instead, the available data on the specific timing of the reposts better identifies the flow of social signals through the network.

---

who are simply uploading content to play with the upload feature but who otherwise have no other listeners will not enter into the data as artists.

Table 1: Summary of Platform Usage

	N	Mean	Median	SD
Plays per Person	122420	1743.93	751	2851.66
Unique Tracks per Person	122420	1025.83	506	1477.53
Plays per Track	426257	500.85	299	755.14
Unique Listeners per Track	426257	294.62	184	362.10
Reposts per Person	62323	26.13	3	156.48
Reposts per Track	567329	2.87	1	6.89
Reposts Seen	120206	1635.18	697.50	2534.31
Tracks Seen in Reposts	120206	1539.70	676	2298.22
Reposted Tracks Played	120206	77.55	20	173.39
Tracks not Reposted	242157			

5.1 Data Description

Due to the hefty computational requirements described in the estimation procedure, it was necessary to engage in subsampling. However, because sampling a network is easy to get wrong, appendix C provides details on the data collection and sampling methodology used to ensure that the fidelity of the information contained in the network is maximized given the constraints. As a constraint of the subsampling process, the data that was retained has the property that at least 100 users listen to each track and each user listens to at least 10 tracks. These data inform the level of participation on the platform and indicate how reposts changed the observed behavior on the platform.

A play is any time a user is exposed to a song's audio.[13]

The key features that are observed in table 1 are that users sample widely from songs

[13]The data lacks the duration of most songs, thus inferring the fraction of a song listened to is a non-trivial affair. Further, the data do not include the ending times of plays. For those songs where duration data is available, a large proportion of plays are spaced substantially further apart than the duration, e.g. with anywhere from seconds or minutes extra to the next play. Whether this represents song search, repeat listening by moving the song's slider bar back to the beginning, or simply pausing the music player is not something that can be determined from the data. The point being that inferring duration of the play is itself a non-trivial task subject to a wide range of possible errors, given the available data. Therefore the analysis is simplified by assuming that the play observed is a sufficient level of information for the user to form an opinion on whether they liked the song or not, i.e. the one-shot learning assumption. Indeed, if a user immediately clicks through a song after a second, this may be a mistake, but it is more likely that they were able to quickly determine the song was not for them and moved on. Furthermore, the "origin" of a play is not available, i.e. whether the play was due to playing through a playlist, was clicked on from another website, was clicked on in a newsfeed, was specifically searched for, or was found through some other browsing mechanism. Only the fact that the song was played is observed. For platforms that record this information, it would be valuable in differentiating between different types of choice (default, deliberate, system recommended, etc.).

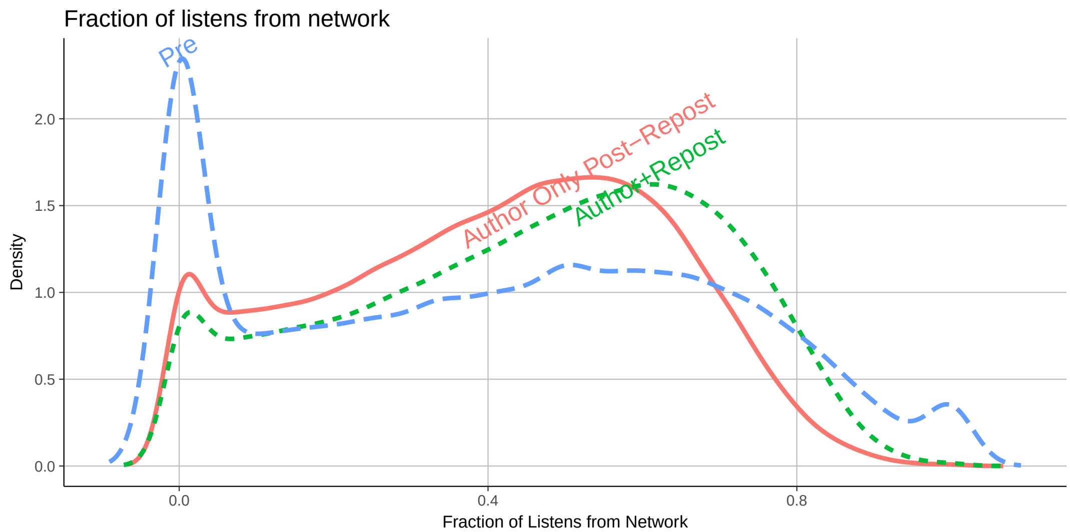
on the platform: of 1743 plays per user, 1000 unique tracks are represented. So, repeat listening of content is not the norm on this platform. However, the reader should not interpret this as users not enjoying using the platform because of the low quality of the content available. 1750 plays is a lot of listening, even if only a few seconds are sampled. A more reasonable interpretation of this data feature is that users are interested in exploring new and innovative content not found on mainstream platforms and that this non-repeat listening is an artifact of that desire to explore more than dissatisfaction with the content consumed.

In terms of reposts, a few features of the data are relevant to keep in mind for the forthcoming analysis. About half of the users repost a song; so this is a widely used platform feature and not a minimal feature only a handful of users adopted. Conditional on reposting at all, users repost on average 36 songs over the 56 weeks I observe reposts in the data; so, users are on average reposting a song about every other week or so. However, users are exposed to 1540 tracks through reposts, only a few of which are reposted by multiple users. So, it isn't immediately obvious that the most popular tracks on the platform are the ones being reposted. Rather, it appears that the use of the repost feature fits in with the exploration story told above and that the reposts are the other side of the coin of that exploration is the announcement of the otherwise undiscovered gems that are found. Finally, the average user plays 78 songs in the 56 weeks from their repost feed. Some of these may also have been uploaded by authors they follow, but the rate of listening to a reposted song is 25 times higher compared to the background 500,000 other tracks in this sample (not to mention the hundreds of millions of tracks that were on the platform by this time). Finally, of the 426,000 songs played by the users, 242,000 are not reposted. Thus, reposts have not replaced user exploration of music on the platform.

To see this latter point in more detail, let's look at fig. 3. This plot contains three lines each representing what fraction of plays of a given user comes "from their network". "From their network" means one of two things, which are not mutually exclusive. First, a listen from an author or from an artist is when the user is observed to be following the author of the song before they played it. Note, this does not imply that the user was following the author before the song was uploaded; they may have been digging through the author's back catalog and discovered the song that way or through some other mechanism. The second way in which a listen may originate from a user's network is that the song was reposted to a user before they listened to it. Specifically, this means that the user followed the reposter before the song was played and the the play came after the song was reposted.

The three lines plot two measures at two points in time: Network/author listens in the pre-repost regime (these are of course synonymous at this time), author-only listens in the post-repost regime, and author-or-repost listens in the post-repost regime. Probably the most obvious feature of this plot is that the network listens in the pre-repost period (the "Pre" line) have a lot of mass near 0. That is, a large number of users listen to most of their

Figure 3: Change in Plays' Origin before and after reposts feature



music from artists they don't follow. One might surmise that this represents an exploration phase where users have to acclimate themselves to the platform and discover worthwhile people to follow before their fraction of network listens can travel upward. However, after limiting the pre-period to 30 or 60 days before the introduction of the repost feature, the character of the plot remains the same, suggesting that this is not a valid explanation for this data pattern.

Moving to the next line, "Author Only", after the introduction of reposts, users have a much larger fraction of their listens from the authors they follow. Finally, in addition to the authors they follow, an extra approximate 10% bump is observed once listens sourced from either authors or reposts are included. Thus, in a direct since, it seems that following an author is a stronger influence on what a user listens to than their social connections. However, there is some evidence that reposts influence what authors to follow. Therefore, it seems that reposts are inducing a change in the sourcing of plays. However, conclusive causal evidence is lacking, but the data patterns suggest that the claim that social influence is playing a substantial role in observed outcomes on the platform is on reasonable footing.

5.2 Descriptive Indicators of Herding

One area where we might expect the effect of reposts to play a role is in herding behavior. This may show up in the degree to which users seek out the most popular songs.

This possibility is explored with two dependent variables: the average market share of the song they listen to and the number of unique songs they listen to. The latter captures whether activity on the platform is significantly increasing (or decreasing) in response to reposts. The former gets at the question of whether users are herding toward more popular songs.

These variables are aggregated at the user-month level (i, t), include a dummy for whether the repost feature is active D_t^{repost}, count the number of other users user i follows as a proxy for participation in the social network, the total number of plays they make as a measure of their overall activity on the platform, and include a time trend to absorb how the platform may be evolving over time. This regression is run on full balanced panel before the sub-sampling and thus has more individuals in the regressions compared to the rest of the analyses.

$$\begin{aligned}
\log(\text{uniq songs})_{it} = \beta_0 + \beta_1 D_t^{repost} + \beta_2 \log(\text{follows})_{it} + \beta_3 D_t^{repost} \log(\text{follows})_{it} + \\
\beta_4 \log(\text{plays})_{it} + \beta_5 D_t^{repost} \log(\text{plays})_{it} + \tau \text{ time_trend}_t + \epsilon_i + \varepsilon_{it} \\
\log(\text{song share})_{it} = \text{same as above.}
\end{aligned}$$

While statistically significant, there is not a strong indication that reposts are substantially increasing the number of songs users are listening to, perhaps about a song per month. In contrast, the average market share of the songs that users listen to rises substantially. This increase in song share is a convenient if imperfect way to capture herding that could be due to herding towards higher vertical quality songs vis-a-vis learning. Or herding could be due to co-consumption utility. Nevertheless, this is reasonable descriptive evidence that suggests reposts generate herding behavior.

Table 2: Does Reposting Generate Herding

	Unique Songs	Unique Songs	Unique Songs	Song Share	Song Share	Song Share
reposts		0.012***	0.015***		1.212***	1.153***
		(0.0004)	(0.0004)		(0.003)	(0.003)
log_num_follows	0.018***	0.026***	0.030***	0.366***	0.148***	0.082***
	(0.0001)	(0.0001)	(0.0001)	(0.001)	(0.001)	(0.001)
log_num_plays	0.897***	0.906***	0.906***	0.394***	0.446***	0.445***
	(0.0001)	(0.0001)	(0.0001)	(0.0003)	(0.001)	(0.001)
time_trend			−0.001***			0.016***
			(0.00001)			(0.0001)
repostsTRUE:log_num_follows		0.0001	0.002***		−0.044***	−0.085***
		(0.0001)	(0.0001)		(0.001)	(0.001)
repostsTRUE:log_num_plays		−0.012***	−0.012***		−0.067***	−0.062***
		(0.0001)	(0.0001)		(0.001)	(0.001)
Ind. FE	✓	✓	✓	✓	✓	✓
Num. Ind	2221059	2221059	2220913	2221059	2221059	2220913
Observations	47,185,948	47,185,948	47,151,445	47,185,948	47,185,948	47,151,445
R^2	0.980	0.980	0.980	0.578	0.593	0.595

Note: *p<0.01; **p<0.001; ***p<1e-04

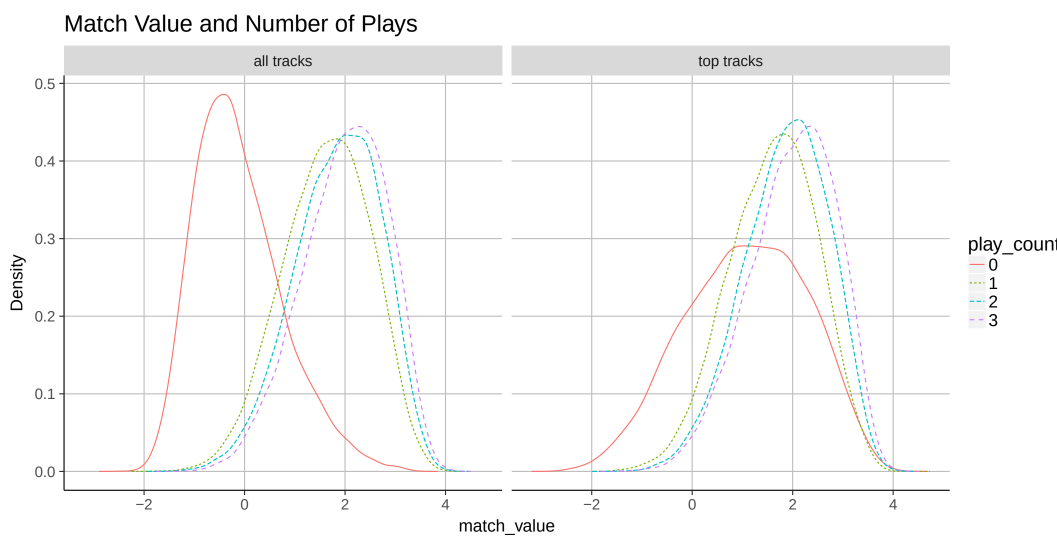
Errors are clustered at the user level.

6 Results

6.1 Latent Space Performance

Before moving on to estimates of the parameters of interest, let me take a moment to convince you that the latent space has the capacity to work in the way I claim. The two novel ways in which I use latent space is to take into account correlated preferences when estimating a network model and in this learning process. The high level of performance of the algorithm already suggests it does well in predicting people's listening preferences as evidenced by the AUC[14] in footnote 5 in Chapter 2, and even distinguishes repeated from single listens as seen in fig. 4. What remains to be seen is how well it does for its other uses in the model: capturing correlated preferences and creating sensible priors and adjustments to those priors.

Figure 4: Machine Learning Effectiveness



The density of the match value for a random sample of songs at the individual level as to whether that individual listened to the song 0, 1, 2, or 3+ times is plotted. Note that the input to the machine learning algorithm is only *whether* a song was listened to by the given user or not, *not* how many times the user listened to the song. Nevertheless, the machinery is able to separate the distribution of single listens from multiple listens. The separation works even when the zero play songs are sampled from the 1000 most popular songs in the sample (right panel). All songs and users are sampled randomly across the panels, thus the densities will not match exactly.

[14] AUC is the "area under the receiver operating characteristics curve". The curve itself plots the true positive rate against the false positive rate, while the resulting AUC measurements gives the probability of a correct relative ranking of a randomly chosen pair.

What comes out of BPR are a set of features for users and songs, but comparing things in this high dimensional space is difficult. The natural starting point, then, is to use the features to construct the measures that are used in the model to judge whether those measures perform in expected ways. Specifically, the inner product of two items gives some notion of the match value between those items; the higher the match value between two terms, the more similar they are to one another. This is because for two random vectors with the same expected norm, the highest match value between the vectors will be achieved when the vectors are equal to one another. In this specific sense, the similarity can be defined between people and between songs in the same manner that the match value of a song with a person.[15]

Recall that the input into BPR is only who plays what song, with no direct information about the social network nor the number of times nor duration of time a song was played. Nevertheless, as seen in fig. 5, the fact that users have correlated preferences is effectively captured. To construct this graph, 100 unaffiliated users are sampled, then up to 20 of the people each one follows is also sampled. For all pairwise combinations in the resulting dataset, each pair is classified as having 2, 1, or 0 connections between them depending on whether the two users follow each other, one follows and the other doesn't, or neither follows the other. The match value for each pair is computed, normalized as described above, and the density of match values is plotted for each of the three conditions. As can be seen, unaffiliated users have an approximately 0 mean match value, while users with one way follows have a much higher match value, and the second follow providing yet another bump. Thus there is substantive evidence that the latent space is capturing a great deal of the correlated preferences among socially connected users.

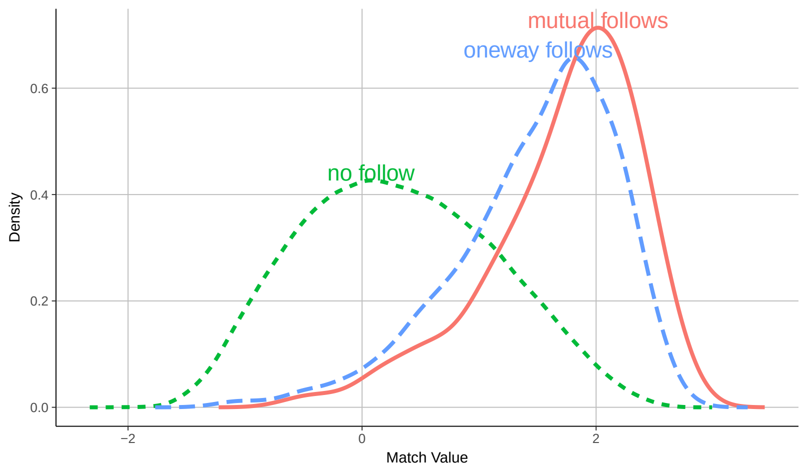
Next, I want to be sure that the learning construct is sensible in terms of the observed match values. Here, in fig. 6, I sample one upload from 500 authors. I generate the match value of these uploads to 4 different sets of comparisons: 1) some random tracks, 2) the genre average as computed above, 3) users who repost the track (with their feature vector rescaled as noted above), and 4) the author's other uploads. I then plot the densities of the match values for each of these comparisons.

The performance here is reasonable. Random tracks have very little similarity with the

[15] The optimal regularization terms for people and songs were different, leading to the average user norm being larger than the average song norm. When comparing people with people versus comparing people with songs, this produces different expected ranges of match values depending on the categories of comparisons. Therefore, when comparing a user in place of a track for comparison purposes (for example, in the artist prior or in the reposter adjustment to the artist or genre priors), the comparison user vectors are scaled by the average ratio between the norm of a user vector with a track vector:

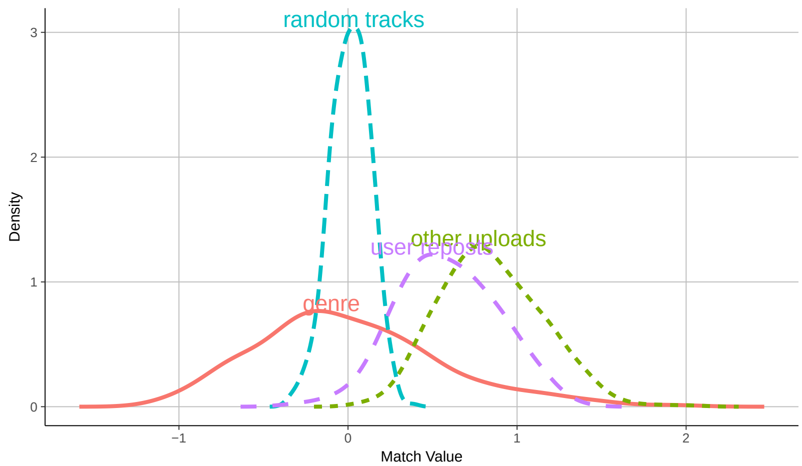
$$\psi_j^{(\text{as track})} := \psi_j \frac{\frac{1}{|S|}\sum_{s \in S} \|\phi_s\|}{\frac{1}{|I|}\sum_{i \in I} \|\psi_i\|}$$

Figure 5: Connected People Have More Correlated Preferences



100 unaffiliated users are sampled, then for each user, sample up to 20 of their connections. For all pairwise connections in this sample, classify the pair as unconnected if neither user follows the other, having a one way connection if one user follows the other but not vice versa, or a mutual follow if both users follow each other. Socially connected users clearly have more similar tastes to unconnected users. Users with a mutual connection have even more similar tastes. Thus, it appears that the latent factor analysis produces an effective method for capturing correlated preferences.

Figure 6: Potential Sources of a Prior for a User's New Upload



One song from each of 500 authors who are also users is sampled. For each author, the focal track with the average characteristics of: the author's other uploaded tracks, the genre of the track, a random selection of tracks, and the user characteristics of the people who reposted the focal track is compared. Genre is a high variance, low mean estimator that sometimes outperforms random tracks, but is in general, no better. Thus, we should not expect genre to be a great predictor of match quality for a new upload. Other uploaded tracks is the next best estimator. Finally, using the characteristics of users who reposted the song has a rather high match value as well, though not as high as the author's other uploads.

author's uploaded track, which is an encouraging sanity check. An author's other uploads appear to be similar to the track under examination, which is consistent with our belief that artists should be producing self-similar music over time. Users who repost the track are the next most similar metric.[16]

The more difficult issue to judge is the performance of the genre prior. On the one hand, its mean match value is approximately 0, which we might believe indicates a poor performing prior. On the other hand, there is a lot of mass in that distribution to the right. I believe this is a reasonable outcome. While we might in general want a simple category such as genre to do a good job summarizing the horizontal qualities of a single piece of music, we should not be surprised that it does not work that way. Indeed, if one thinks about ones own preferences in music, it is likely that you have tastes for certain kinds of music within a specific genre rather than for the genre itself. Thus, the fact that some music in the genre is well summarized by the top songs in the genre while others are not should be expected.

What is more difficult to explain is the mass of the distribution for the genre comparisons to the left even of the random tracks. One possibility, unconfirmed, is that certain genres load on certain features but not others, but within a genre, the loadings on those features may be highly varied. For example, a set of features may capture something about the lyrical content of the music. Consider that hip-hop or rock can be quite lyrical, but the nature of the lyrics may be fast versus slow, political versus personal, positive versus negative, etc. So, the interaction in that feature is strong, but it can potentially be inverted when the song diverges from the genre prototype. In comparison, a random track from, say, electronic music, has no lyrics and thus loads not at all and generates zero match value versus strongly negative. If this were true, this would imply a near-zero match value between songs from different genres because of a lack of interaction between the high value components of the respective song vectors. But songs within a genre may generate high match values when their correlated loadings agree, but negative match values when their correlated loadings disagree.[17]

[16]This is somewhat sensitive to the rescaling I use. Without the rescaling, user reposts still have a similarly high level of similarity, but the variance is much higher, with more mass both to the right and left of the current distribution.

[17]To explain this hypothesis with a simple numerical exercise, consider songs from two genres and the latent space has $k = 2$ with the first dimension relevant only to one genre and the second dimension relevant only to the other genre. Take song A represented by vector $(0, 1)$ and B from the other genre which loads on the first dimension $(1, 0)$. The match value between A and B is 0. But consider song C from the same genre as A, thus loading on the second dimension as well, but in a different way $(0, -1)$. Then the match value between A and C is -1 despite the fact that they are from the same genre.

6.2 Model Estimates

The full set of results can be found in table 3. Focusing on the first column, I estimate the model without considering peer effects, which gives us a baseline against which to understand how the model captures peer effects. The match value (β) appears to consistently capture a great deal of the variation in regard to why users choose to listen to the songs they do. Recall that δ_a, δ_g capture uncertainty and risk avoidance associated with the prior on the artist and the genre, with the parameter values transformed to be in $[0, 1]$ by the logistic sigmoid. What we see is that δ_a is larger than δ_g, that is, users seem to place more weight on their prior beliefs regarding artists compared to genres. This is consistent with the notion that genres are not particularly informative categories, either for listeners deciding what to listen to or for the econometrician to use to control for heterogeneous preferences. The social learning parameters λ_a, λ_g capture similar logic. Given that a user has a high degree of trust in their artist-specific prior, we should expect the learning effect from peers to be weaker when recommending a song from an artist a user is familiar with. Indeed, this is what is seen in the table. While users do learn from peers when considering their artist prior, the contrast with the learning with a generic song from a genre is striking. In this case, the user places no trust in their genre prior and put all faith in the peer recommendation.

Another interesting shift in parameter estimates is the impact of including co-consumption utility on θ, the parameter indicating how people respond to track popularity. Before considering the impact of social interactions, it appears that users have a slight preference (ignoring the standard errors, see section 6.5) for more popular tracks. However, after including co-consumption utility, it appears that people have a disutility for popularity. This is indicative of an in-group mentality among users on SoundCloud, which is consistent with what researchers have found in other consumption contexts (Kim and Chintagunta 2012). Including the effect of social learning may slightly moderate the estimated relative in-group preference, but the estimates are too noisy for further comparison.

Table 3: Model Estimates, with Peer Effects

	Match Value		
β	1.65^{***}	1.55^{***}	1.58^{***}
	(0.0305)	(0.0346)	(0.0357)
	Co-Consumption Utility		
γ		2.75^{***}	2.63^{***}
		(0.101)	(0.0975)
	Learning Uncertainty		
δ_a	1.6^{***}	1.83^{***}	1.86^{***}
	(0.1)	(0.191)	(0.177)
δ_g	-0.273	-0.493	-0.445
	(0.0533)	(0.064)	(0.0619)
	Reposter Priority		
λ_a			0.634
			(0.529)
λ_g			9.37
			(18.7)
	Habit		
α	0.736^{***}	0.796^{***}	0.765^{***}
	(0.0538)	(0.0574)	(0.0583)
τ	-0.273	-0.229	-0.234
	(0.0128)	(0.0144)	(0.0144)
	Top Track		
θ	0.162	-0.435	-0.341
	(0.176)	(0.18)	(0.123)
batch size	256	256	256
niters	8000	8000	8000
$\mathcal{L}$	-20.0	-22.0	-23.0
AICc	53.0	58.0	65.0

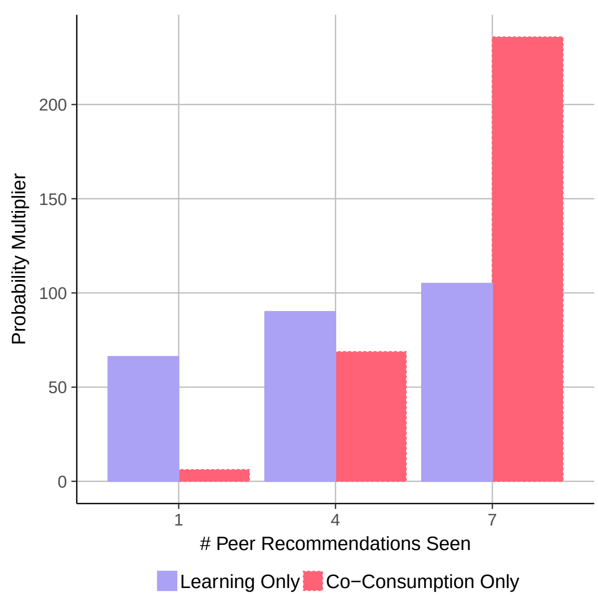
Note: ***p<0.001; **p<0.01; *p<0.05

The first model is the estimated without any regard to social effects. The second column adds the model with co-consumption but no learning; this might be considered equivalent to how a simpler model with only one social effect might be estimated. In the third model, the learning parameters are applied in characteristic space. I use the batch size as the n for the small sample size adjustment in AICc and for the degrees of freedom for the t-stat in the p-value calculation.

6.3 Counterfactual: Social Learning Versus Co-Consumption in Song Adoption

This counterfactual is used to better understand what is more influential in terms of song success. What is estimated is the probability of listening to a song that a user is reasonably amenable to (with a match value of 0.5) in one of two worlds: one with only social learning and one with only co-consumption utility. The probability of adoption is computed using the above estimates, but with either the social learning parameter, λ_g, set to 0 to simulate a world where learning does not take place or the co-consumption parameter, γ, set to 0 to simulate a world with no co-consumption. I then expose users to a number of recommendations and under the social learning condition, allow the true solo value (1.5) to be learned after 7 recommendations and otherwise allow co-consumption utility to have the effect implied by taking the log number of recommendations and having that enter the model directly. What results is the probabilities estimated in fig. 7.

Figure 7: Counterfactual Probabilities of Adoption Due to Social Learning versus Co-Consumption



The probability of adoption of a song in a world with only social learning ($\gamma = 0$) or only co-consumption utility ($\lambda_g = 0$) with increasing number of exposures of a user to recommendations from their peers.

What is seen here is that social learning can be a potent force for song adoption, but it has its limits. Specifically, after the location is more precisely revealed by more recommendations, the marginal effect on adoption of yet another decreases as there is no more precision in location to be gained. Co-consumption utility, however, has a relatively small effect for just one recommendation. But as the number of recommendations increases, it becomes a more and more potent force for encouraging song adoption, such that, once there is between 4 to 5 recommendations, it becomes more influential than social learning.

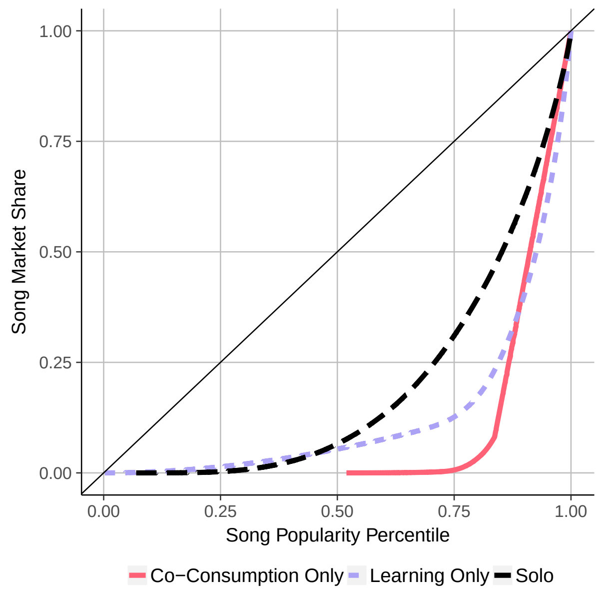
Considering that most songs on the platforms, if they are reposted at all, only get one exposure per user, social learning is a common force on the platform and likely leads to quite a lot of song adoption that might not otherwise have occurred. However, for the most popular songs that get many reposts and therefore several exposures per user, co-consumption utility dominates social learning in terms of why they may be adopted. Thus, when looking at the top of the distribution of popularity, it is likely there are many songs there due more to people listening because of seeing repeated reposts than their underlying private valuation of those songs. Thus marketers need to beware that as they try to translate success from SoundCloud to other contexts, that they may be facing challenges besides the purely informational: they will need to overcome the chicken-and-egg problem that adoption of these songs may be driven by the fact that others were already listening to the song.

6.4 Counterfactual: Social Effects, Superstars, and the Long Tail

A secondary issue of interest is the distribution of success under regimes with high levels of learning versus high levels of co-consumption utility. An issue that has been of on-going interest with the rise of internet commerce has been understanding when and where long tail products become more significant in the product landscape compared to ecosystems where superstar products dominate. The effects of bundling, the marginal costs, search, and recommendation engines have all been explored in a variety of papers on the subject (e.g. Fleder and Hosanagar 2009). What seems to have been less studied is the role of social effects, much less different types of social effects in producing super star versus long tail phenomenon. For example, social learning may help niche, high quality products to gain more market share when resources for marketing are limited or otherwise under-utilized compared to realized product quality.

To provide some insight on this question, I run a similar exercise compared to the previous conterfactual. Instead of looking at the probability of adoption of a song, though, I look at which songs out of a catalog are adopted under the above regimes and a perfect information world with no social effects. I simulate the utilities in each of these three scenarios for the top 2000 most popular songs in my dataset, and allow my 100,000 users to each adopt their 50 favorite songs implied by the model under study. I use the number of times each song is reposted to represent the learning process from the genre average to the true value. I then produce a Lorenz plot in fig. 8, showing how cummulative market share changes under the different scenarios.

Figure 8: Social Effects, Superstars, and the Long Tail



A Lorenz plot of market shares in a perfect information world with no social effects ("Solo"), a world with only social learning, and a world with only co-consumption utility. Market shares represent number of listeners adopting a song out of the total number of listeners and is summed across songs to produce the cummulative market share numbers.

Using the perfect information, no social effect world as a baseline, the comparison between social learning and co-consumption utility is striking. In a world with only social learning, we see users sampling throughout the entire distribution. However, as we get to the top of the distribution, it seems as if the songs with the most reposts become more liked as users learn that they like the true value represented by those songs due to the large number of reposts. This imposes a superstar effect relative to the perfect information, no social effect world; however this effect is rather small compared to a world with co-consumption utility only. In this co-consumption utility only world, no user listens to a song in the bottom half of the distribution. The attractiveness of the potential to participate in the gains from co-consumption outweighs all private value that these songs may have for these users. Thus, in this sense, social learning is a potent force for informing users about and encouraging them to adopt more niche songs compared to the impact of people's desire to co-consume, causing the consumption of twice as many unique songs.

6.5 Batch Gradient Descent and Statistical Inference

A caveat to these results is that the standard errors and model summary statistics are based on a single batch of 256 data points. The exact procedure for handling the small sample in the computation of the information matrix but the large sample in the estimation procedure does not seem to have been worked out for producing reliable test statistics. However, some early results have recently been publicly released, but not yet reviewed, that suggest that allowing the batch size to grow or the learning rate to shrink can produce, under some assumptions, valid inference. While I allow my learning rate to shrink, I currently do so under a different schedule than is suggested by these papers. Whether the current standard errors under- or over- state the precision is not clear as there are forces working in both directions. The fact that I could have used a larger batch size to estimate the errors would imply that my current errors imply a higher variance than if I had used a larger batch. However, the fact that I used many other batches to arrive at my estimates but don't include a correction for this used data, nor the data used from the ML estimate, suggests that I am overstating my precision. Thus, how the new approaches to inference with stochastic estimation methods will effect statistical inference in the model is ambiguous.

7 Conclusion

This papermakes four contributions. I delineate how two different endogenous social effects may be at play in many decision contexts and the application to digital test markets, specifically how those social effects may produce different interpretations for the observed outcomes in such markets. I present an identification strategy for separately identifying and measuring these two social effects. This identification strategy is embedded in a choice framework that can be estimated at scale, using machine learning of a latent space to better capture the high dimensional heterogeneity we expect in creative and cultural markets such as music. Counterfactual simulations suggest that on SoundCloud, co-consumption utility is twice as influential in song adoption for the most popularly shared songs despite the importance of learning effects. However, those learning effects are key in introducing users to new songs they may enjoy. Comparing worlds with only co-consumption utility versus only social learning, social learning helps introduce users to twice as many new songs as they would otherwise sample.

The approach taken in the paper of modeling complex social interactions in a relatively simple manner was enabled by a rich dataset with detailed information on user behavior on the platform. Similar approaches may be taken in other contexts where directly modeling systems behavior is difficult or impractical. One area of concern is the potentially self-fulfilling nature of recommendation engines (Knijnenburg, Sivakumar, and Wilkinson 2016). If users select the top recommendation of a recommendation engine, and the recommendation engine recommends the most popular choices, then the long term validity of that system becomes precarious without further intervention. If platforms log choice contexts and specifically which choices were made at the behest of the recommendation engine, then platforms may be able to better back out which items in their catalog are being over-represented by their recommendation system and which may be under-represented following an approach similar to that which has been taken here.

Appendices

A Mini-Batch Gradient Descent with ADAM

I estimate this using mini-batch gradient descent, specifically the ADAM variant, with a batch size of 256 random plays. This implies a batch likelihood:

$$\mathcal{L}_{batch} = \sum_{(i,s,t)\in \text{plays}}^{256} \mathcal{L}_{ist} \tag{5}$$

where I randomly select i, s, t from the data, then randomly select s' based on i's play and repost exposure history. I compute the gradient using exact automatic differentiation, NOT numerical differentiation (nor symbolic differentiation) (Revels, Lubin, and Papamarkou 2016). Automatic differentiation, in an over-simplified summary, differentiates the source code of your objective function, thus giving a gradient that is exact up to the numerical error inherent to the computation of your objective function. See appendix B for more details on what automatic differentiation is, and some motivation for why you should adopt it.

The mini-batch gradient is computed as the mean of each of the individual gradients rather than the sum to ensure that the effect of the choice of the batch size does not affect the SGD learning rate parameter:

$$\nabla \tilde{\mathcal{L}}_{batch} \coloneqq \frac{1}{256} \sum_{(i,s,t)\in plays}^{256} \nabla \mathcal{L}_{ist}$$

Using the mean rather than the sum here is a common technique as otherwise changing the batch size results in changes to the effective learning rate.

I randomly initialize the parameters $\Theta = \{\alpha, \beta, \delta, \gamma, \lambda, \tau, \theta\}$, and update using the ADAM (adaptive moment estimation) variant of SGD (Kingma and Ba 2014). The idea is to carry forward past movement to keep noisy deviations from the path to the optimal point to a minimum. While typically SGD is updated as:

$$\Theta^{(1)} = \Theta^{(0)} - \alpha \nabla_{\Theta^{(0)}} f(\Theta^{(0)})$$

ADAM combines momentum and RMSProp variants of SGD regularization with bias correction into a single well behaved optimization routine (Kiefer and Wolfowitz 1952; Tieleman and Hinton 2012). The most common pitfall of SGD is that it can get "stuck" in relatively flat areas of an objective function such as inflection points or valleys, where the "way out" is by traversing the relatively flat space of the valley and ignoring the slopes. By combining the current gradient with the decayed average of previous moves and scaling by the decayed average of the variance (second moments), the algorithm has proven effective in various incarnations of this pitfall:

$$
\begin{aligned}
\beta_1 &= 0.9 \\
\beta_2 &= 0.999 \\
\epsilon &= 10^{-8} \\
m_\theta^{(0)} &= 0 \\
v_\theta^{(0)} &= 0 \\
g_\theta^{(t)} &= \nabla_{\theta^{(t)}} f \\
m_\theta^{(t+1)} &= \beta_1 m_\theta^{(t)} + (1-\beta_1) \cdot g_\theta^{(t)} \\
v_\theta^{(t+1)} &= \beta_2 v_\theta^{(t)} + (1-\beta_2) \cdot g_\theta^{(t)^2} \\
\theta^{(t+1)} &= \theta^{(t)} - \alpha \frac{m_\theta^{(t+1)}}{1-\beta_1^t} \frac{1}{\sqrt{\frac{v_\theta^{(t+1)}}{1-\beta_2^t}} + \epsilon}
\end{aligned}
$$

Iteration of ADAM proceeds exactly as it would with SGD, except now the vectors m_θ and v_θ need to be retained to compute the update. The extra cost here is minimal compared to the cost of computing the gradient itself and only requires storing a couple extra copies of vectors the same size as the input parameters. Strictly speaking, while the β here are hyperparameters that can be tuned, practitioners have not found this to be a useful exercise, and these values seem to work well in a wide variety of applications.

In the below results, I use 8000 iterations of these mini-batches. Parameters appear to converge in under 1000 iterations; thus, I can use the remaining iterations to lower the learning rate to produce a more stable final result, less subject to the variance inherent to stochastic methods. I therefore allow the learning rate to exponentially decay from 1.0 to 0.001 over the 8000 iterations:

$$\beta_{\text{decay}} = \frac{0.001}{1}^{\frac{1}{8000}} \approx 0.9991369$$

and then on each iteration, the α used for the update on the ith iteration is:

$$\hat{\alpha} = 1 \cdot \beta_{\text{decay}}^{i}$$

B Automatic Differentiation

The more one forays into the world of stochastic methods, the more useful precise computation of gradients is. The reason for this is that there is already quite a lot of noise in the computation of the gradient simply by virtue of the *stochastic* method at hand. If it can be avoided, not adding to that error with inaccurate derivatives can improve performance. Furthermore, using well tuned automatic differentiation libraries results in the computation of these gradients faster than finite differences can accomplish because fewer operations may need to be computed and those operations can be better vectorized for better cache performance.

If you are not familiar with automatic differentiation, it is a useful result out of computer science and mathematics that has been rediscovered several times over in slightly different forms over the last 150 or so years.[18] Back propagation, a popular algorithm used in the training of neural networks, is a special case of reverse mode automatic differentiation. Perhaps the earliest form of automatic differentiation was in construction of the algebra of derivatives, called derivations. With dual numbers, the space is created $x+x'\varepsilon$ where $\varepsilon^2 := 0$ and the rules otherwise follow much as they would in the complex plane. However, instead of generating the complex plane, computing on this dual space generates derivatives. As a quick example, consider computing on pa^2 but in the dual space:

$$\begin{aligned} p \cdot (a + b\varepsilon)^2 &= p \cdot \left(a^2 + 2ab\varepsilon + b^2\varepsilon^2\right) \\ &= pa^2 + b\varepsilon \cdot \underbrace{2pa}_{=\frac{d}{da}pa^2} + \underbrace{b^2\varepsilon^2}_{\varepsilon^2=0} \end{aligned}$$

The first term here is the original polynomial and the second term is our sentinel b, the dual value ε, and the derivative $2pa$; and since this required only one forward pass through

[18] You may be wondering, if this has been around for 150 years, why this method is not a great deal more common. A couple of answers to this is that this was a narrow niche of mathematics whose application to computing was not always recognized. Secondly, the vast majority of programming languages do not allow re-definition of base types such as "integer" or "double floats". This technology has been available in some C++ libraries; however, use there requires rewriting substantial portions of numerical code to accommodate the custom data types. In Julia, a modern language designed with numerical and scientific computing in mind, users can define their own types that operate at the same level as basic machine types (Bezanson et al. 2017). Thus, generic Julia code can operate nearly seamlessly with the automatic differentiation libraries without any loss of performance or the need to write custom code.

the computation is called "forward mode" automatic differentiation. The automatic differentiation library can then tag its sentinel value(s) so that they are not inappropriately computed with non-sentinel values, allowing cheap computation of the original objective and the gradient. Depending on the size of the parameter space, other forms of automatic differentiation can be faster, though the exact crossing point will be application specific and is changing over time due to further engineering advancements of these libraries. It is up to the researcher to try different methods for best performance on their specific objective function.

C Social Sampling

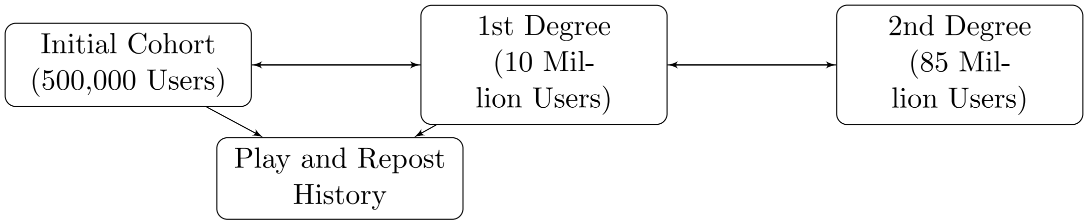
While random sampling is usually taken for granted as a simple procedure, in the context of networks, sampling appropriately can be far from trivial (Papagelis, Das, and Koudas 2013). The literature on identification has found that researchers must maximize the retention of social influences in their data (Blume et al. 2015). A major concern is that one might overstate social effects by ascribing the influences of several people, not sampled, to one person who is.

Thus, let me take a few moments to convince you that I have taken reasonable steps to preserve a sufficient quantity of information about the connections in the network to conduct a convincing network analysis. Sampling was undertaken in several stages to construct the data examined below. First, SoundCloud provided a large sample of data using the follow process. Sampling was done using the snowball strategy: Two separate cohorts of users (that is, users who registered in a given period) were selected; then all of the users' connections to and from other users were selected; and then all those users' connections were selected. Only the affiliations of this last group of "second degree" users with the other users were recorded while the data described above was provided for the first two levels of the network. Figure C.1 is a visual display of this process. This strategy allows me to identify to what degree I capture the complete network for each user in the sample.

This data included 85 million users and their 2 billion connections, and billions of plays from around 10 million users over the 4 year period from 2012-2016. Repost data was generally available from 2012-2016 as well; however, for a substantial subset of users, reposts were only made available through 2013. Thus, in this analysis, I include data only for 2012 and 2013.

I then removed any user who did not have at least 10 plays in both 2012 and 2013 and at least one play after the first half of 2014. This latter requirement was instituted because a substantial anti-spam policy took effect approximately in May 2014. By requiring users to have activity after this policy was instantiated, I can be relatively assured that these users

Figure C.1: Data Sampling from SoundCloud



are not bots or other kinds of fraudulent accounts at any point in the data dating back to the beginning of the data set.

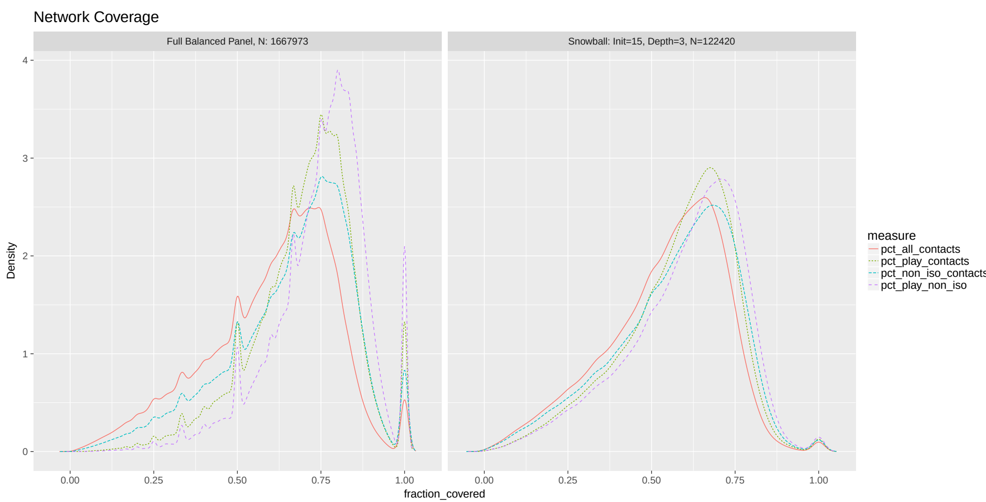
However, for the present analysis, the amount of data remaining was still too large to deal with feasibly with available resources. I then tried several different sampling strategies with the objective of minimizing the number users in my sample while still retaining a large amount of the network information. The networks literature suggests several possible measures to identify whether a network's structure was preserved after sampling. Here, I am using the fraction of each individual's total network that is retained in the data. I use this measure because the literature on identifying network effects emphasizes the importance of maximizing the amount of information the econometrician should have regarding each user's social influences. Thus, measuring the network information at the level of the user and to what degree I observe who they follow is most relevant to properly identify the flow of information through my sample to my users.

I found that using the snowball strategy again was the most productive to maximize retained network information while minimizing the number of users retained in the data set. I seeded the snowball with 15 mutually unaffiliated users and grabbed all social influences (users that they follow) for 3 iterations. I start with mutually unaffiliated users to try to pull in 15 separate networks so that my network identification does not rest on the properties of any one given network. Of course, due to the low degree of separation exhibited in social networks, these are not mutually exclusive networks, but it is a reasonable approach to the problem, short of attempting to select 15 users with a maximum network distance from one another. I then trimmed away all users who did not follow anyone else in the network until every user was influenced by at least 1 other user in the sample. This resulted in 139,790 users and all their plays and reposts.

Finally, I require each track to have 100 unique listeners and each user to have listened to 10 separate tracks. I do this to maximize the efficiency and reduce the run time of a machine learning technology I use below, though in principle I can relax this last filter when the final set of analyses is settled to check that my results are robust.

This 100 listener per track and 10 tracks per user sample contains approximately 122,000

Figure C.2: Network Coverage



These density plots plot to what degree in my two data samples I have coverage of the social network with the denominator being a) the number of contacts a user has, b) all non-isolated contacts, c) all contacts I have plays data on, and d) the intersection of b + c. The numerator is the number of contacts that I retain in that sample, which by the sampling strategy are defined to not be isolated (i.e., to follow at least one other person) and for whom I have plays data. Silverman's rule of thumb is used for the bandwidth (and all density plots that follow) with no bias correction for the bounds.

users, 500,000 tracks, and 215 million plays. Figure C.2 shows to what degree I was able to recover the network information that was in the initial panel before further subsampling was undertaken. What's measured here is, for each user, the number of the people they follow that are retained in my sample. I represent this as a fraction with four different bases for comparison: 1) All users as represented in fig. C.1 are the denominator for the red line, 2) All users with play data in fig. C.1 for the green line, 3) Going back to 1) but removing users who do not follow anyone, i.e. are "isolated" in the context of the wider sample, and 4) the intersection of 2) and 3), i.e., all users for whom I have play and repost data and who follow at least one person in the network.

The left panel contains all users who meet the 2012-2014 filter discussed above. The right panel contains the discussed sub-sample of the users in the left panel. The left panel should be viewed as a basis for the upper limit of what I can achieve in terms of network retention with the balanced panel approach. A distribution substantially to the right of what's seen in the left panel would suggest that the sampling is not preserving the structure of the network. Similarly, a distribution substantially to the left would indicate that I had lost the ability to capture the flow of information in the network.

While there is a non-trivial loss of fidelity in network content, most strikingly a substantial mass of users for whom I no longer have nearly 100% of their network, the shape of the distribution otherwise mimics the balanced sample rather closely. Thus, it appears that this sampling strategy worked to effectively preserve a high level of the network structure I care about for my analysis for these users.

D References

Ackerberg, Daniel A. (2001). "Empirically Distinguishing Informative and Prestige Effects of Advertising". In: *The RAND Journal of Economics* 32.2, pp. 316–333. ISSN: 0741-6261. DOI: 10.2307/2696412. JSTOR: 2696412.

Aggarwal, Charu C., Alexander Hinneburg, and Daniel A. Keim (2001-01-04). "On the Surprising Behavior of Distance Metrics in High Dimensional Space". In: *Database Theory — ICDT 2001*. International Conference on Database Theory. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 420–434. ISBN: 978-3-540-41456-8 978-3-540-44503-6. DOI: 10.1007/3-540-44503-X_27. URL: https://link.springer.com/chapter/10.1007/3-540-44503-X_27 (visited on 2018-06-18).

Akerlof, George A. (1980-06-01). "A Theory of Social Custom, of Which Unemployment May Be One Consequence". In: *The Quarterly Journal of Economics* 94.4, pp. 749–775. ISSN: 0033-5533. DOI: 10.2307/1885667. URL: https://academic.oup.com/qje/article/94/4/749/1858931 (visited on 2018-06-25).

Aral, Sinan, Lev Muchnik, and Arun Sundararajan (2009-12-22). "Distinguishing Influence-Based Contagion from Homophily-Driven Diffusion in Dynamic Networks". In: *Proceedings of the National Academy of Sciences* 106.51, pp. 21544–21549. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0908800106. pmid: 20007780. URL: http://www.pnas.org/content/106/51/21544 (visited on 2018-06-19).

Athey, Susan et al. (2018-01-22). "Estimating Heterogeneous Consumer Preferences for Restaurants and Travel Time Using Mobile Location Data". In: arXiv: 1801.07826 [cs, econ, stat]. URL: http://arxiv.org/abs/1801.07826 (visited on 2018-06-08).

Bakshy, Eytan et al. (2012). "The Role of Social Networks in Information Diffusion". In: *Proceedings of the 21st International Conference on World Wide Web*. WWW '12. New York, NY, USA: ACM, pp. 519–528. ISBN: 978-1-4503-1229-5. DOI: 10.1145/2187836.2187907. URL: http://doi.acm.org/10.1145/2187836.2187907 (visited on 2018-06-19).

Bellet, Aurélien, Amaury Habrard, and Marc Sebban (2013-06-27). "A Survey on Metric Learning for Feature Vectors and Structured Data". In: arXiv: 1306.6709 [cs, stat]. URL: http://arxiv.org/abs/1306.6709 (visited on 2018-06-18).

Beyer, Kevin et al. (1999-01-10). "When Is "Nearest Neighbor" Meaningful?" In: *Database Theory — ICDT'99*. International Conference on Database Theory. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 217–235. ISBN: 978-3-540-65452-0 978-3-540-49257-3. DOI: 10.1007/3-540-49257-7_15. URL: https://link.springer.com/chapter/10.1007/3-540-49257-7_15 (visited on 2018-06-18).

Bezanson, J. et al. (2017-01-01). "Julia: A Fresh Approach to Numerical Computing". In: *SIAM Review* 59.1, pp. 65–98. ISSN: 0036-1445. DOI: 10.1137/141000671. URL: https://epubs.siam.org/doi/10.1137/141000671 (visited on 2018-06-14).

Blume, Lawrence E. et al. (2015-04). “Linear Social Interactions Models”. In: *Journal of Political Economy* 123.2, pp. 444–496. ISSN: 0022-3808, 1537-534X. DOI: 10.1086/679496. URL: https://www.journals.uchicago.edu/doi/10.1086/679496 (visited on 2018-06-08).

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin (2009-05). “Identification of Peer Effects through Social Networks”. In: *Journal of Econometrics* 150.1, pp. 41–55. ISSN: 03044076. DOI: 10.1016/j.jeconom.2008.12.021. URL: http://linkinghub.elsevier.com/retrieve/pii/S0304407609000335 (visited on 2017-05-03).

Brock, William A. and Steven N. Durlauf (2003-03). *Multinomial Choice with Social Interactions*. Working Paper 288. National Bureau of Economic Research. DOI: 10.3386/t0288. URL: http://www.nber.org/papers/t0288 (visited on 2018-06-08).

Bursztyn, Leonardo et al. (2014-07-21). “Understanding Mechanisms Underlying Peer Effects: Evidence From a Field Experiment on Financial Decisions”. In: *Econometrica* 82.4, pp. 1273–1301. ISSN: 1468-0262. DOI: 10.3982/ECTA11991. URL: https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA11991 (visited on 2018-06-20).

Cao, Xinyu and Juanjuan Zhang (2017). “Prelaunch Demand Estimation”. In: p. 43.

Case, Anne C. and Lawrence F. Katz (1991-05). *The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youths*. Working Paper 3705. National Bureau of Economic Research. DOI: 10.3386/w3705. URL: http://www.nber.org/papers/w3705 (visited on 2018-06-25).

Chatterjee, Rabik Ar and Jehoshua Eliashberg (1990-09-01). “The Innovation Diffusion Process in a Heterogeneous Population: A Micromodeling Approach”. In: *Management Science* 36.9, pp. 1057–1079. ISSN: 0025-1909. DOI: 10.1287/mnsc.36.9.1057. URL: https://pubsonline.informs.org/doi/abs/10.1287/mnsc.36.9.1057 (visited on 2018-06-25).

Conley, Timothy G. and Christopher R. Udry (2010-03). “Learning about a New Technology: Pineapple in Ghana”. In: *American Economic Review* 100.1, pp. 35–69. ISSN: 0002-8282. DOI: 10.1257/aer.100.1.35. URL: https://www.aeaweb.org/articles?id=10.1257/aer.100.1.35 (visited on 2018-06-19).

Crane, Jonathan (1991-03-01). “The Epidemic Theory of Ghettos and Neighborhood Effects on Dropping Out and Teenage Childbearing”. In: *American Journal of Sociology* 96.5, pp. 1226–1259. ISSN: 0002-9602. DOI: 10.1086/229654. URL: https://www.journals.uchicago.edu/doi/10.1086/229654 (visited on 2018-06-25).

De Giorgi, Giacomo, Michele Pellizzari, and Silvia Redaelli (2010-04). “Identification of Social Interactions through Partially Overlapping Peer Groups”. In: *American Economic Journal: Applied Economics* 2.2, pp. 241–275. ISSN: 1945-7782, 1945-7790. DOI: 10.1257/app.2.2.241. URL: http://pubs.aeaweb.org/doi/10.1257/app.2.2.241 (visited on 2018-06-08).

Domingos, Pedro (2012-10). “A Few Useful Things to Know About Machine Learning”. In: *Commun. ACM* 55.10, pp. 78–87. ISSN: 0001-0782. DOI: 10.1145/2347736.2347755. URL: http://doi.acm.org/10.1145/2347736.2347755 (visited on 2018-06-18).

Duflo, Esther and Emmanuel Saez (2003-08-01). "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment". In: *The Quarterly Journal of Economics* 118.3, pp. 815–842. ISSN: 0033-5533. DOI: 10.1162/00335530360698432. URL: https://academic.oup.com/qje/article/118/3/815/1943005 (visited on 2018-06-19).

Duncan, Otis Dudley, Archibald O. Haller, and Alejandro Portes (1968). "Peer Influences on Aspirations: A Reinterpretation". In: *American Journal of Sociology* 74.2, pp. 119–137. ISSN: 0002-9602. JSTOR: 2775916.

Fleder, Daniel and Kartik Hosanagar (2009-05). "Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity". In: *Management Science* 55.5, pp. 697–712. ISSN: 0025-1909, 1526-5501. DOI: 10.1287/mnsc.1080.0974. URL: http://pubsonline.informs.org/doi/abs/10.1287/mnsc.1080.0974 (visited on 2016-03-17).

García-Jimeno, Camilo, Angel Iglesias, and Pinar Yildirim (2018-04). *Women, Rails and Telegraphs: An Empirical Study of Information Diffusion and Collective Action*. Working Paper 24495. National Bureau of Economic Research. DOI: 10.3386/w24495. URL: http://www.nber.org/papers/w24495 (visited on 2018-06-15).

Goettler, Ronald L. and Ron Shachar (2001). "Spatial Competition in the Network Television Industry". In: *The RAND Journal of Economics* 32.4, pp. 624–656. ISSN: 0741-6261. DOI: 10.2307/2696385. JSTOR: 2696385.

Golub, Benjamin and Matthew O. Jackson (2010-02). "Naïve Learning in Social Networks and the Wisdom of Crowds". In: *American Economic Journal: Microeconomics* 2.1, pp. 112–149. ISSN: 1945-7669. DOI: 10.1257/mic.2.1.112. URL: https://www.aeaweb.org/articles?id=10.1257/mic.2.1.112 (visited on 2018-06-19).

Gomez-Uribe, Carlos A. and Neil Hunt (2015-12). "The Netflix Recommender System: Algorithms, Business Value, and Innovation". In: *ACM Trans. Manage. Inf. Syst.* 6.4, 13:1–13:19. ISSN: 2158-656X. DOI: 10.1145/2843948. URL: http://doi.acm.org/10.1145/2843948 (visited on 2018-06-15).

Green, Paul E. and V. Srinivasan (1990). "Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice". In: *Journal of Marketing* 54.4, pp. 3–19. ISSN: 0022-2429. DOI: 10.2307/1251756. JSTOR: 1251756.

Hartmann, Wesley R. (2010-07). "Demand Estimation with Social Interactions and the Implications for Targeted Marketing". In: *Marketing Science* 29.4, pp. 585–601. ISSN: 0732-2399, 1526-548X. DOI: 10.1287/mksc.1100.0559. URL: http://pubsonline.informs.org/doi/abs/10.1287/mksc.1100.0559 (visited on 2018-06-15).

Hartmann, Wesley R. et al. (2008-12). "Modeling Social Interactions: Identification, Empirical Methods and Policy Implications". In: *Marketing Letters* 19.3-4, pp. 287–304. ISSN: 0923-0645, 1573-059X. DOI: 10.1007/s11002-008-9048-z. URL: http://link.springer.com/10.1007/s11002-008-9048-z (visited on 2017-05-03).

Hauser, John R. and Ely Dahan (2008). "Chapter 8: New Product Development". In: *Marketing Management: Essential Marketing Knowledge and Practice*. URL: http://

www.mit.edu/~hauser/Papers/Chapter%208%20Hauser_Dahan%20Book%20Chapter%20on%20New%20Products.pdf (visited on 2018-06-25).

Hernando, Antonio, Jesús Bobadilla, and Fernando Ortega (2016-04-01). “A Non Negative Matrix Factorization for Collaborative Filtering Recommender Systems Based on a Bayesian Probabilistic Model”. In: *Knowledge-Based Systems* 97, pp. 188–202. ISSN: 0950-7051. DOI: 10.1016/j.knosys.2015.12.018. URL: http://www.sciencedirect.com/science/article/pii/S0950705115005006 (visited on 2018-06-15).

Hitsch, Günter J. (2006-01-01). “An Empirical Model of Optimal Dynamic Product Launch and Exit Under Demand Uncertainty”. In: *Marketing Science* 25.1, pp. 25–50. ISSN: 0732-2399. DOI: 10.1287/mksc.1050.0140. URL: https://pubsonline.informs.org/doi/abs/10.1287/mksc.1050.0140 (visited on 2018-06-15).

Iyengar, Raghuram, Christophe Van den Bulte, and Jae Young Lee (2015-01-19). “Social Contagion in New Product Trial and Repeat”. In: *Marketing Science* 34.3, pp. 408–429. ISSN: 0732-2399. DOI: 10.1287/mksc.2014.0888. URL: https://pubsonline.informs.org/doi/abs/10.1287/mksc.2014.0888 (visited on 2018-06-25).

Iyengar, Raghuram, Christophe Van den Bulte, and Thomas W. Valente (2010-12-10). “Opinion Leadership and Social Contagion in New Product Diffusion”. In: *Marketing Science* 30.2, pp. 195–212. ISSN: 0732-2399. DOI: 10.1287/mksc.1100.0566. URL: https://pubsonline.informs.org/doi/abs/10.1287/mksc.1100.0566 (visited on 2018-06-25).

Kiefer, J. and J. Wolfowitz (1952-09). “Stochastic Estimation of the Maximum of a Regression Function”. In: *The Annals of Mathematical Statistics* 23.3, pp. 462–466. ISSN: 0003-4851, 2168-8990. DOI: 10.1214/aoms/1177729392. URL: https://projecteuclid.org/euclid.aoms/1177729392 (visited on 2018-06-22).

Kim, Minki and Pradeep K. Chintagunta (2012-09-01). “Investigating Brand Preferences across Social Groups and Consumption Contexts”. In: *Quantitative Marketing and Economics* 10.3, pp. 305–333. ISSN: 1573-711X. DOI: 10.1007/s11129-011-9117-0. URL: https://doi.org/10.1007/s11129-011-9117-0 (visited on 2019-06-04).

Kingma, Diederik P. and Jimmy Ba (2014-12-22). “Adam: A Method for Stochastic Optimization”. In: arXiv: 1412.6980 [cs]. URL: http://arxiv.org/abs/1412.6980 (visited on 2018-06-22).

Knijnenburg, Bart P., Saadhika Sivakumar, and Daricia Wilkinson (2016). “Recommender Systems for Self-Actualization”. In: *Proceedings of the 10th ACM Conference on Recommender Systems*. RecSys ’16. New York, NY, USA: ACM, pp. 11–14. ISBN: 978-1-4503-4035-9. DOI: 10.1145/2959100.2959189. URL: http://doi.acm.org/10.1145/2959100.2959189 (visited on 2019-06-04).

Koren, Yehuda, Robert Bell, and Chris Volinsky (2009-08). “Matrix Factorization Techniques for Recommender Systems”. In: *Computer* 42.8, pp. 30–37. ISSN: 0018-9162. DOI: 10.1109/MC.2009.263. URL: http://ieeexplore.ieee.org/document/5197422/ (visited on 2018-06-08).

Kulis, Brian (2013-07-31). "Metric Learning: A Survey". In: *Foundations and Trends® in Machine Learning* 5.4, pp. 287–364. ISSN: 1935-8237, 1935-8245. DOI: 10.1561/2200000019. URL: https://www.nowpublishers.com/article/Details/MAL-019 (visited on 2018-06-18).

Manski, Charles F. (1993). "Identification of Endogenous Social Effects: The Reflection Problem". In: *The Review of Economic Studies* 60.3, pp. 531–542. ISSN: 0034-6527. DOI: 10.2307/2298123. JSTOR: 2298123.

Moore, Geoffrey A. (2002-08-20). *Crossing the Chasm: Marketing and Selling Disruptive Products to Mainstream Customers*. Harper Collins. 254 pp. ISBN: 978-0-06-051712-0.

Nair, Harikesh, Puneet Manchanda, and Tulikaa Bhatia (2010). "Asymmetric Social Interactions in Physician Prescription Behavior: The Role of Opinion Leaders". In: *Journal of Marketing Research* 47.5, pp. 883–895. ISSN: 0022-2437. JSTOR: 20751550.

Netzer, Oded et al. (2008-12-01). "Beyond Conjoint Analysis: Advances in Preference Measurement". In: *Marketing Letters* 19.3-4, p. 337. ISSN: 0923-0645, 1573-059X. DOI: 10.1007/s11002-008-9046-1. URL: https://link.springer.com/article/10.1007/s11002-008-9046-1 (visited on 2018-06-25).

Papagelis, M., G. Das, and N. Koudas (2013-03). "Sampling Online Social Networks". In: *IEEE Transactions on Knowledge and Data Engineering* 25.3, pp. 662–676. ISSN: 1041-4347. DOI: 10.1109/TKDE.2011.254.

Rendle, Steffen et al. (2009). "BPR: Bayesian Personalized Ranking from Implicit Feedback". In: *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*. UAI '09. Arlington, Virginia, United States: AUAI Press, pp. 452–461. ISBN: 978-0-9749039-5-8. URL: http://dl.acm.org/citation.cfm?id=1795114.1795167 (visited on 2018-06-08).

Revels, Jarrett, Miles Lubin, and Theodore Papamarkou (2016-07-26). "Forward-Mode Automatic Differentiation in Julia". In: arXiv: 1607.07892 [cs]. URL: http://arxiv.org/abs/1607.07892 (visited on 2018-06-14).

Ruiz, Francisco J. R., Susan Athey, and David M. Blei (2017-11-09). "SHOPPER: A Probabilistic Model of Consumer Choice with Substitutes and Complements". In: arXiv: 1711.03560 [cs, econ, stat]. URL: http://arxiv.org/abs/1711.03560 (visited on 2018-06-08).

Ryan, Bryce and Neal C. Gross (1943-03-01). "The Diffusion of Hybrid Seed Corn in Two Iowa Communities." In: *Rural Sociology; College Station, Tex., etc.* 8.1, pp. 15–24. ISSN: 0036-0112. URL: https://search.proquest.com/docview/1291026197/citation/6E6880DA0B7D4DCAPQ/1 (visited on 2018-06-20).

Salganik, Matthew J., Peter Sheridan Dodds, and Duncan J. Watts (2006-02-10). "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market". In: *Science* 311.5762, p. 854. DOI: 10.1126/science.1121066. URL: http://science.sciencemag.org/content/311/5762/854.abstract.

Schultz, P. Wesley et al. (2007-05). "The Constructive, Destructive, and Reconstructive Power of Social Norms". In: *Psychological Science* 18.5, pp. 429–434. ISSN: 0956-7976, 1467-9280. DOI: 10.1111/j.1467-9280.2007.01917.x. URL: http://journals.sagepub.com/doi/10.1111/j.1467-9280.2007.01917.x (visited on 2018-06-25).

Shin, Sangwoo, Sanjog Misra, and Dan Horsky (2011-06-16). *Disentangling Preferences and Learning in Brand Choice Models*. SSRN Scholarly Paper ID 1865866. Rochester, NY: Social Science Research Network. URL: https://papers.ssrn.com/abstract=1865866 (visited on 2018-06-08).

Silk, Alvin J. and Glen L. Urban (1978-05). "Pre-Test-Market Evaluation of New Packaged Goods: A Model and Measurement Methodology". In: *Journal of Marketing Research* 15.2, p. 171. ISSN: 00222437. DOI: 10.2307/3151248. URL: https://www.jstor.org/stable/3151248?origin=crossref (visited on 2018-06-25).

Tieleman, Tijmen and Geoff Hinton (2012). "Lecture 6.5 - RMSProp". Technical report, 2012 (COURSERA: Neural Networks for Machine Learning). URL: http://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf (visited on 2018-06-22).

Titsias, Michalis K. (2016). "One-vs-Each Approximation to Softmax for Scalable Estimation of Probabilities". In: p. 9.

Tomasello, Mario V. et al. (2014-07-14). "The Role of Endogenous and Exogenous Mechanisms in the Formation of R&D Networks". In: *Scientific Reports* 4, p. 5679. ISSN: 2045-2322. DOI: 10.1038/srep05679. URL: https://www.nature.com/articles/srep05679 (visited on 2019-06-04).

Toubia, Olivier (2006-09-01). "Idea Generation, Creativity, and Incentives". In: *Marketing Science* 25.5, pp. 411–425. ISSN: 0732-2399. DOI: 10.1287/mksc.1050.0166. URL: https://pubsonline.informs.org/doi/abs/10.1287/mksc.1050.0166 (visited on 2018-06-25).

Urban, Glen L. and John R. Hauser (1993). *Design and Marketing of New Products*. Prentice Hall. 744 pp. ISBN: 978-0-13-201567-7.

Van den Bulte, Christophe and Gary L. Lilien (2001-03-01). "Medical Innovation Revisited: Social Contagion versus Marketing Effort". In: *American Journal of Sociology* 106.5, pp. 1409–1435. ISSN: 0002-9602. DOI: 10.1086/320819. URL: https://www.journals.uchicago.edu/doi/abs/10.1086/320819 (visited on 2018-06-25).

Wan, Mengting et al. (2017). "Modeling Consumer Preferences and Price Sensitivities from Large-Scale Grocery Shopping Transaction Logs". In: ACM Press, pp. 1103–1112. ISBN: 978-1-4503-4913-0. DOI: 10.1145/3038912.3052568. URL: http://dl.acm.org/citation.cfm?doid=3038912.3052568 (visited on 2018-06-08).

Weintraub, Gabriel Y., C. Lanier Benkard, and Benjamin Van Roy (2010-06-03). "Computational Methods for Oblivious Equilibrium". In: *Operations Research* 58 (4-part-2), pp. 1247–1265. ISSN: 0030-364X. DOI: 10.1287/opre.1090.0790. URL: https://pubsonline.informs.org/doi/abs/10.1287/opre.1090.0790 (visited on 2018-09-21).

Yang, Liu (2006). "Distance Metric Learning: A Comprehensive Survey". In: p. 51.

Young, H. Peyton (2009-12). "Innovation Diffusion in Heterogeneous Populations: Contagion, Social Influence, and Social Learning". In: *American Economic Review* 99.5, pp. 1899–1924. ISSN: 0002-8282. DOI: 10.1257/aer.99.5.1899. URL: https://www.aeaweb.org/articles?id=10.1257/aer.99.5.1899 (visited on 2018-06-19).

Zhang, Juanjuan (2009-07-23). "The Sound of Silence: Observational Learning in the U.S. Kidney Market". In: *Marketing Science* 29.2, pp. 315–335. ISSN: 0732-2399. DOI: 10.1287/mksc.1090.0500. URL: https://pubsonline.informs.org/doi/abs/10.1287/mksc.1090.0500 (visited on 2018-06-25).

Zhang, Jurui, Yong Liu, and Yubo Chen (2015-04-07). "Social Learning in Networks of Friends versus Strangers". In: *Marketing Science* 34.4, pp. 573–589. ISSN: 0732-2399. DOI: 10.1287/mksc.2015.0902. URL: https://pubsonline.informs.org/doi/abs/10.1287/mksc.2015.0902 (visited on 2018-06-25).