Quantifying Social Influence in an Online Music Community

Research-in-Progress

Sanjeev Dewan
University of California, Irvine
California, United States
sdewan@uci.edu

Yi-Jen Ian Ho
University of California, Irvine
California, United States
hoy1@uci.edu

Jui Ramaprasad1
McGill University
Quebec, Canada
jui.ramaprasad@mcgill.ca

Abstract

This paper studies two types of social influence in an online music community: observational learning influence based on aggregate consumption data, and social network influence based on music consumption by friends in social proximity. The analysis uses a variety of empirical methods, applied to highly granular user listening and “favoriting” behavior on the largest music blog aggregator site. Our analysis finds positive evidence for observational learning effects, but no evidence for social network influence. Thus, any social influence in this music context is channeled through popularity cues offered by aggregate consumption statistics, rather than contact and communication with friends in close social proximity. We discuss implications of these results for research and practice.

Keywords: Blogging community, music, social media, social influences, observational learning

1 Author names in alphabetical order.
Introduction

Social or peer influence has long been recognized as a driver of adoption decisions (e.g., Bass 1969, Brown and Reingen 1987), but its importance has only been heightened recently with the proliferation of online social media and social networks. In the music industry, the context we study here, social media have made sharing of tastes and preferences easier than ever before, and in a recent survey 54% of subjects indicated that they base their music purchasing decisions on positive recommendations from friends (The Nielsen Company 2012a). Per Nielsen Global Trust’s survey (The Nielsen Company 2012b), 92% of consumers say that recommendations from people they know are the most trusted sources of information when making consumption decisions, followed by 70% of consumers who say that they trust consumer opinions posted online. Despite this anecdotal evidence, we do not yet have a good understanding of the true extent, and mechanisms by which, social influence shapes online music consumption—a gap that we hope to begin to fill with this research.

Recently, understanding the role of social influence on consumer choices has been examined in a variety of contexts, such as movie sales (Moretti 2011), Facebook applications (Aral and Walker 2011), adoption of the iPhone 3G (de Matos et al 2012), music subscription services (Bapna and Umyarov 2012), among others. In this paper, we study the role of peer influence on consumption in an online music community, where we investigate peer influence at two levels. First, we examine observational learning effects, defined by Bikhchandani et al (2005) as “influence resulting from rational processing of information gained by observing others.” Second, we study social network influence, resulting from contact and communication with “friends” in one’s social network (see, e.g., Ma et al. 2010 and Egebark and Ekstrom 2011).

Music is a rich context in which to study IT-enabled social influence on consumer behavior, for a number of reasons. First, music is an experience good, so that consumers value the opinions and actions of other consumers as signals of whether or not they would like the music themselves. Second, music is an information good, where discovery and consumption are increasingly done online, and in our case, at the very same website (as we explain below). Finally, the music industry has been transformed by technology and social networks in profound ways, so that understanding social influence in this context will foreshadow what we can expect for other information and experience goods, such as movies, software, digital media, etc.

Thus, drawing from the current literature on social influence and the music industry, we ask the following questions: (i) What is the extent of observational learning influence in online music consumption? Are these effects stronger for newly posted or previously posted music? (ii) What role does a user’s local social network play in influencing music consumption decisions?

To answer the first question, we exploit a natural experiment enabled by a newly implemented feature in an online music community. This feature allowed users to observe all other users’ music favoriting behavior in the aggregate, albeit anonymously. We deploy a difference-in-difference (DD) methodology to measure the impact of aggregate peer consumption information on other users’ consumption decisions. More precisely, we examine whether there exists observational learning (OL) influence due to the revealed past popularity of the music among other consumers. In the second part of the paper we focus on social network influence, corresponding to the second research question above. We adopt instrumental variables (IV) analysis to identify and measure social network influence. The IV analysis attempts to measure the impact of music consumption of social network “friends” on the music listening choices of the focal user, in a linear regression framework. Here, music consumption by friends of friends who are not friends (i.e., intransitive triads) is used as the instrumental variable to separate social network influence of friends from homophily and other endogenous effects, building on the recent work of de Matos et al. (2012).

To summarize our results, we find positive evidence for observational learning effects. Our empirical results confirm that being able to observe aggregate peer consumption decisions does have an impact on subsequent consumption. In addition, we find that this impact attenuates over time, with statistically significant effects lasting about three days after the disclosure of popularity information. On the other

---

2 Homophily refers to social correlation in actions due to the fact that people tend to befriend others who have similar tastes and preferences (i.e., birds of a feather flock together).
hand, we find no significant evidence of social network influence, using either the instrumental variables or the randomization approach. Thus, peer influence in our context is channeled through observational learning of aggregate consumption information, and not through social proximity in one’s social network. We discuss implications of these findings in the conclusions section.

**Literature Review**

This paper draws from two main streams of work: literature around observational learning effects and social network influence. The first stream consists of studies that look at how individuals make decisions based on aggregate adoption decisions of previous customers. The second stream of literature examines the role that social network ties play on individual consumption decisions.

**Observational Learning effects**

Observational learning is the process by which consumers make decisions based on aggregate consumption statistics of prior users. In economics, this process is often described as herd behavior. Bikchandani et al. (1998) described herding behavior, as a type of observational learning process where “reports of the actions or endorsements of one set of economic decision makers often influence the reactions and purchases of others.” Since music is an experience good, where a consumers must first experience a good before being able to value it (Nelson 1970), signals from others, in the form of their decision to adopt, are often used as a means to filter high quality goods from low quality goods.

Whether or not knowledge of aggregate consumption decisions has an effect on subsequent individual consumption has been examined in prior work in the context of books (Sorenson 2007), software adoption (Duan et al. 2006) and online music (Salganik et al. 2006). More recently, Chen et al. (2011) look at the effect of observational learning (OL) in the presence of word of mouth (WOM) effects, based on Amazon.com data, and find that not only do OL and WOM individually drive purchase decisions, but the interaction between the two processes is significant as well. Finally, Moretti (2011) looks for social learning in movie sales, finding that sales of movies with positive surprise and negative surprise in opening weekend sales diverge over time; i.e., successful movies become more successful while unsuccessful movies become more unsuccessful. Further, the effect of surprise (relative to ex ante priors as reflected in the number of screens dedicated to a movie) is larger for audiences with larger social networks.

**Social Network Influence**

Social network influence is due to social proximity (or contact and communication) between social network “friends.” A key challenge in identifying social influence is to be able to separate it from homophily (Manski 1993), where the latter refers to social correlation in actions due to the fact that people tend to befriend others who have similar tastes and preferences (i.e., “birds of a feather flock together”). There have been various new methods applied to find evidence of social network influence, solving the reflection problem. Aral et al. (2009) develop a dynamic matched sample estimation framework (propensity-score matching) to conclude that homophily accounts for both correlations and temporal interdependence of behaviors among individuals. Ma et al. (2010) construct a hierarchical Bayesian model to study the effects of peer influence and homophily on both the timing and choice of consumer purchases within a social network. Aral and Walker (2011) design a randomized experiment on Facebook to quantifying social network influence. De Matos et al. (2012) apply the intransitive triads instrumental variable approach (Tucker 2008) to separate social influence from homophily.

**Social Influence in Music Consumption**

For the reasons mentioned in the Introduction, there is great emerging interest on the role of IT-enabled social influence in the music industry. New music arrives to the marketplace at a growing pace and the growing Long Tail nature of the music market is increasing the importance of social media in the process of music discovery and consumption. Accordingly, a number of studies have recently studied the impact of social media on music consumption. For example, Dewan and Ramaprasad (2012) study the impact of
music blogging on online sampling, and find that observational learning effects are stronger in the tail relative to the body of music sales distribution. Dhar and Chang (2009) find that the volume of user-generated content is predictive of music sales. Dewan and Ramaprasad (2013) study the interaction among social media (blog buzz), traditional media (radio play) and music sales and find that while blog buzz is positively related to album sales, it is negatively related to song sales, possibly due to the sales displacement effect of online free sampling.

In a study closest in objective to ours, Salganik et al. (2006) looked at the impact of aggregate prior consumption decisions on the ultimate inequality and unpredictability in an artificial music market. They found that social influence in the form of observation of prior aggregate consumption decisions “contributes both to inequality and unpredictability in cultural markets,” providing evidence that “collective behavior” plays a part in consumption decisions. The main difference between Salganik et al. (2006) and this study is that while the prior work created an artificial music market, ours is based on real observational data, and designed to exploit a natural experiment with data at a high level of granularity and richness. Further, we examine both observational influence and social network influence, whereas the prior study was restricted to just the former effect.

Data and Methodology

We have a unique dataset that allows us to discern any observational learning effects of social influence on an individual’s consumption decision by using an exogenous feature implementation within the online music community we are studying, while controlling for song-level characteristics. In addition, after constructing a social network of users within this community we are able to understand the local social network influence that the neighbors in a social network have on an individual’s choice of music consumption.

We use data from an online music community, The Hype Machine (THM), the leading music blog aggregator. THM aggregates mp3s that are posted in their entirety on thousands of music blogs, which it tracks, allowing for users to stream the song (but not download it) from the site. Since its inception, THM has allowed users to “listen” to a song that is posted (by clicking on the “listen” link). In 2007 and 2008, THM implemented two innovations that we use to conduct this study. First, in October of 2007, THM implemented a social networking feature, where visitors can create accounts and become members. Within this social network community, each member can create their own personal dashboard, highlighting the three main “social network” type activities that one can participate in: the individual can add favorite tracks, blogs or other users to their dashboard. Our dataset includes information on all of the members’ favorite tracks, favorite blogs, and favorite users. In this study, we focus on creating a social network of users from the “favorite” data. This act of “favoriting” a person is akin to following another user on Twitter in that it creates a unidirectional tie, which is not necessarily reciprocated. In addition to the dashboard data, THM has provided the daily logs of the music that each of the users have listened to.

THM implemented a second innovation on October 1, 2008 — they added a number next to each track posted on their site, which indicated how many people had “favorited” the song. While individuals could favorite a song starting in October 2007, the number of favorites of the song was not viewable by other visitors to the site (even registered users) until the implementation of this special feature on October 1, 2008. We have data on user behavior on THM from before and after THM implemented this visible “popularity” measure, providing an opportunity for a natural experiment.

Difference-In-Difference Model of Observational Learning Influence

To conduct the analysis of observational learning, we employ a difference-in-difference (DD) methodology, exploiting both the “before” and “after” time periods as well as the variation in the number of “favorites” of a song. Given that the implementation of this feature is exogenous, the DD model allows us to see the relative change in listening patterns before and after the intervention by comparing the songs that experienced the intervention (the treatment group) to a set of songs that did not (the control group). Specifically, on the basis of our DD model, we can quantify how much the appearance of the number of favorites impacts listening by comparing songs between two time periods.
In our model, we have one day in the “after” time period, so we estimate coefficients for “after” time periods and estimate the interaction effect with both of these time periods as well. The model specification is as follows, for Song $j$ at Time $t$:

$$
\log(\text{listen}_{jt}) = \beta_0 + \beta_1 \text{Treatment} + \beta_2 \text{Time}_t + \beta_3 \log(\text{favorites}_{jt}) + \beta_4 \text{Treatment} \times (\text{Time}_t) + \epsilon_{jt}, \quad (1)
$$

where $\text{Time}_t$ and $\text{Treatment}$ are dummy variables to identify the day following the feature implementation and the treatment group, respectively. The regression covers the time period running from one day before the feature implementation to one day after; i.e., $t = -1, 0, 1$, where $t = 0$ is the day the feature was implemented. The variables $\text{listen}_{jt}$ and $\text{favorites}_{jt}$ denote the total number listens and favorites of Song $j$ on Day $t$. We use OLS to estimate the coefficient of the interaction term, which captures the observational learning effects.

We restrict both the treatment and control groups to songs posted to THM on the same day, in order to control the duration for listening and favoriting by all users. Because the number of song favorites was made visible for all songs on October 1, 2008, we cannot construct treatment and control groups from a single set of songs. Instead, we take the treatment group to be songs posted on September 29, and two alternate control groups, as the sets of songs that were posted one and two weeks before the feature implementation, respectively. These control groups are similar to our treatment group, with the exception of posting date. Given that we analyze a time window of three days, the songs in our control group do not experience the feature implementation. Thus, by looking at parallel time periods of three days, we can successfully measure the impact from the visibility of song favoriting information on subsequent music consumption.

**Linear Regression Model of Social Network Influence**

To examine the influence of consumption behavior of friends in the social network, we create a separate dataset that allows us to examine listening and favoriting behavior at the individual level. This dataset consists of unique user-song observations, for songs released on September 1, 2008, where at least one registered user has listened to the song at least once in either the first or second week. We compile this data for the two weeks following September 1, where week one data consists of activity during September 2 through September 8, 2008 and week two data consists of activity during the week of September 9 through September 15, 2008.

We apply a linear regression model to estimate the effect of a user’s friends listening and/or favoriting a song in Week 1 on the User $i$ listening or favoriting Song $j$ in Week 2. The model specification is as follows:

$$
\log(w1\text{listen}_{ij}) = \beta_0 + \beta_1 \log(w1\text{listen}_{ij}) + \beta_2 \log(w1\text{totallisten}_{ij}) + \beta_3 \log(w1\text{totalfavorite}_{ij}) + \beta_4 \log(w1\text{friendlisten}_{ij}) + \beta_5 \log(w1\text{friendfavorite}_{ij}) + \epsilon_{ij}, \quad (2)
$$

where the variables $w1\text{listen}_{ij}$ and $w2\text{listen}_{ij}$ measure the number of times User $i$ has listened to Song $j$ in Week 1 and Week 2, respectively, $w1\text{totallisten}_{ij}$ and $w1\text{totalfavorite}_{ij}$ are the number of times Song $j$ has been listened to and favorited, respectively, in Week 1. The key social network influence variables are $w1\text{friendlisten}_{ij}$ and $w1\text{friendfavorite}_{ij}$, which count the number of times the focal user $i$’s friends have listened to and favorited Song $j$, respectively, in Week 1. Thus, $\beta_4$ and $\beta_5$ are the key coefficients of interest representing the magnitude of social network influences.

Identifying social influence is a challenge because of the reflection problem (Manski 1993), where the social network influence is difficult to separate from the influence of homophily, user heterogeneity, and unobservables that are correlated with listening and favoriting. Following the approach of de Matos et al. (2012) to solve this issue, we use instrumental variables based on intransitive triads, also known as friends of a friend, who are not directly friends, to distill the effect of influence from that of unobserved user heterogeneity and endogeneity.
Preliminary Results and Discussion

Results of Observational learning Effects

We start with discussion of observational learning effects. Table 1 presents the results that correspond to the OLS estimation of equation (1). With the same model specification, we examine three scenarios. In Scenario 1, the treatment group is the set of songs posted on September 29, 2008 while the control group is the set posted one week earlier (i.e. on September 22). The coefficients of interest are the interaction terms, which indicate the effect that the favorites have on consumption after the intervention, incremental to their impact prior to the intervention. Indeed, the coefficient is positive and significant ($p<0.01$), indicating that showing the number of favorites does influence consumer choice and is consistent with the predictions of the observational learning theories discussed earlier.

The coefficient of the treatment dummy captures the difference between the treatment and control groups in terms of total number of listens. We can infer that the treatment group intrinsically attracts less listens than the control group, but the difference is not a concern for our observational learning study. The Time1 dummy intuitively shows that the total number of listens of a song declines with time (this is a general pattern for all songs). The number of favorites of a song is a control variable to show that popular songs have higher consumption.

For robustness, we examine Scenario 2 where the control group is replaced by the set posted two weeks before September 29. The results are consistent those for Scenario 1, showing support for the existence of observational learning effects. In addition, we compare the two control groups, where we would expect to see no effect. Indeed, the results of Scenario 3 show the coefficient of interaction term is insignificant, which strengthens our result that it is the visibility of the number of favorites on a song that increases the number of listens.

Analogous to the case where a song loses popularity and is consumed less over time, the influence of the observational learning effects also decreases. To understand how long the observational learning effects last, we run Scenario 4, where we look at songs posted on 9/28, three days before the intervention instead of two days before. Interestingly, the results show that the observational learning effects only last for three days in the THM community. This suggests that the users take the number of cumulative favorites as a signal to sample newly posted songs. Once they decide initially whether to sample or not, aggregate consumption information is no longer relevant to them.

<table>
<thead>
<tr>
<th>Table 1. Results of Observational Learning Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario 1 09/29 vs. 09/22</td>
</tr>
<tr>
<td>Intercept</td>
</tr>
<tr>
<td>Treatment</td>
</tr>
<tr>
<td>Time1</td>
</tr>
<tr>
<td>log(favoritejt)</td>
</tr>
<tr>
<td>Treatment*Time1</td>
</tr>
<tr>
<td>Adj. R-squared</td>
</tr>
<tr>
<td># of Observations</td>
</tr>
</tbody>
</table>

Notes: *** and ** denote significance at 1% and 5%, respectively.
Results for Social Network Influence

The second part of our analysis focuses on social network influence, and the results are detailed in Table 2. In estimating equation (2), we used both OLS and 2SLS estimation. Model 1 is based on OLS estimation which incurs the issues of endogeneity and reverse causality while Model 2 employs 2SLS, where $\log(w1friend\text{listen}_{ij})$ and $\log(w1friend\text{favorite}_{ij})$ are treated as endogenous variables in a hierarchical manner. The related instrumental variables are $\log(w1ffn\text{listen}_{ij})$, which is the number of times the friends of friends who are not friends of user $i$ listened to song $j$ in week one, and $\log(w1ffn\text{favorite}_{ij})$, which is the number of times the friends of friends who are not directly friends of User $i$ favorited Song $j$ in week one, respectively.

First, note that $\log(w1\text{listen}_{ij})$ has a negative and significant coefficient for both samples, using either OLS or 2SLS. This shows that users typically listen to songs at a decreasing rate over time, and they seldom repeatedly listen to the same songs after initially sampling them. The negative coefficient on $\log(w1\text{total\text{listen}}_{ij})$ captures the natural decline in the number of times a song is listened to over time. Turning to the key social network influence variables, OLS estimates indicate that a user’s music consumption significantly correlates to her friends’ sampling behavior, but not to their favoriting behaviors. In the 2SLS model (Model 2), however, neither the listening nor favoriting behavior of friends is significant; i.e., there is no evidence of social network influence. The disparity between the OLS and 2SLS models is likely due to the presence of homophily-driven endogeneity effects.

<table>
<thead>
<tr>
<th>Table 2. Results for Social Network Influence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Intercept</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>$\log(w1\text{listen}_{ij})$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>$\log(w1\text{total\text{listen}}_{ij})$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>$\log(w1\text{total\text{favorite}}_{ij})$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>$\log(w1\text{friend\text{listen}}_{ij})$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>$\log(w1\text{friend\text{favorite}}_{ij})$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Adj. R-squared</td>
</tr>
<tr>
<td>Wald-test</td>
</tr>
<tr>
<td># of Observations</td>
</tr>
</tbody>
</table>

Notes: *** and * denote significance at 1% and 10%, respectively.
Conclusion and Work-in-Progress

In this study, we have examined both observational learning and social network influence in an online music community. We did this using data captured at a very granular level, which allows us to observe actual user listening and favoriting behavior on The Hype Machine. Our results suggest positive evidence for observational learning or bandwagon effects, but no evidence for social network influence. Thus there is social influence, but it is only channeled through observational learning from aggregate consumption decisions of peers, and not through social proximity (contact and communication) with social network friends.

In thinking about these results, we believe they reflect the nature of the music blogging context that we study. The positive evidence of observational learning effects suggests that users find popularity cues to be relevant in deciding what songs to listen to. The increasingly “long tail” nature of the music industry — large number of new releases are constantly being added to an already immense library of songs — means that users have very little private or independent information to guide music discovery and consumption. Therefore, they rely on overall consumption statistics of songs in deciding which songs to sample. Such statistics seem to be useful only to discover newly released songs, and hence the significance of popularity information attenuates over time, and goes away within three days of release. Based on the observational learning literature we can expect that initial conditions matter, so that the success of songs depends heavily on initial conditions, leading to inequality in consumption (popular songs will get more popular, while unpopular songs will get more unpopular) and to unpredictability of outcomes (“good” songs may not become popular, while “bad” songs may become viral hits), consistent with the findings of Salganik et al. (2006).

The lack of social network influence could be simply due to the fact that social networking features were relatively new at the time of our analysis, having been introduced only a year back. Therefore, the correlation between any focal user’s music consumption and that of their friends is relatively very sparse during the time frame of our data set. This sparseness is probably also due to the fact that THM is likely to be a fringe social network — as compared to say Facebook or LinkedIn — for most users. In any case, the long tail nature of music also works against finding social network influence. Because of the very large variety of newly released and older songs, it is difficult for there to be a critical mass of consumption in one’s social network neighborhood, so that social network influence forces are likely to be weak. By contrast, in the movie industry, far fewer titles are released every week, so that there is a greater likelihood of significant social network effects.

Our results have implications for research and practice. First of all, reading industry reports, and coming across statistics such as 92% of consumers say positive recommendations from people they know are the most trusted sources of information (The Nielsen Company 2012b), it is easy to take social network influence for granted. However, our results suggest that such anecdotal evidence does not hold up to scientific evidence based on highly granular data — at least not in the music context. At the same time, when surveys indicate that 70% of consumers trust consumer opinions posted online (The Nielsen Company 2012b), our results suggest that what might be driving the implied social influence might not be direct contact and communication between consumers, but rather distant observation of aggregate consumption statistics. Finally, the observational learning results indicate that promotion of music (and similar long tail products) might benefit more from dissemination of popularity information, rather than trying to mobilize adoption in online social networks.

This research is currently a work-in-progress. While we currently examine network influence on a weekly level, a more granular analysis, at a daily or even hourly level, would provide more insight into the nature of the influence. At the same time, we currently examine these two effects separately. In our further research we intend to examine the interaction between these two types of social influence. This will allow us to understand whether, for example, the opinions of the user’s local network is heightened or reduced by the opinions of the whole community.
References


