FRIENDS, CROWDS, AND THE LONG TAIL: AN EMPIRICAL INVESTIGATION ON ONLINE MUSIC LISTENING

Meng Zhao
Beijing University of Posts and Telecommunications, xyfzhm@bupt.edu.cn

Jiayin Qi
Beijing University of Posts and Telecommunications, qijiaoyin@139.com

Follow this and additional works at: http://aisel.aisnet.org/pacis2014

Recommended Citation
http://aisel.aisnet.org/pacis2014/3

This material is brought to you by the Pacific Asia Conference on Information Systems (PACIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in PACIS 2014 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
FRIENDS, CROWDS, AND THE LONG TAIL: AN EMPIRICAL INVESTIGATION ON ONLINE MUSIC LISTENING

Meng Zhao, School of Economics and Management, Beijing University of Posts and Telecommunications, Beijing, xyfzhm@bupt.edu.cn

Jiayin Qi, School of Economics and Management, Beijing University of Posts and Telecommunications, Beijing, qijiayin@139.com

Abstract

In the area of Web 2.0, many studies have proved online feedback has an important impact on online users’ buying behaviour, considering the crowds influence mostly. Based on the data from Last.fm website, our study examines the impact of both friends and crowds online feedback, compare the different influences of these two factors on users’ behaviour, and find out how these effects change in the tail as well. Our results suggest that friends and crowds online feedback both have the positive influence on users’ music listening behavior and the influence of friends’ online feedback is much stronger than crowds’. And compared to mainstream music, the influence of both friends and crowds online feedback will become stronger to niche music. Put together, the findings of our research contribute to get the new value-intention framework in Web 2.0 and can provide novel managerial implications for enterprises.

Keywords: friends online feedback, crowds online feedback, user behaviour, long tail.
1 INTRODUCTION

The development of the Internet has brought a great impact on the music industry. For example, traditional music listening pattern has gone through a transfer from offline listening to the combination of both offline and online (Peitz and Waelbroeck 2006). And online music is dominating the music industry gradually. Reflecting this trend, in 2011, online music sales climbed past physical sales to take a 50.3% market share of all music purchases (Time, 2012) and the number is expanding. Online music has become an important part of our lives.

Online music is a very special good with the following three features. Firstly, online music comforts the diversity of demands of most people, regardless of whether it is rock or popular, mainstream or niche. The long tail theory can easily explain this phenomenon of music industry where consumers’ preferences have far greater depth than what one could find in a typical brick-and-mortar storefront. Secondly, online music updates quickly and it is very easy to get. Thirdly, online music is an experience good so it seems that friends’ and crowds’ information will be more helpful.

Since music products have such a large rich database, long tail is an inevitable factor we have to consider. The mainstream music has more information: the more listeners, the more shouts and consumers can reduce their uncertainty about these mainstream products, while the niche music has less information that will increase the uncertainty about the products.

Consumers purchase decision theory contains five progresses: problems cognition, information searching, alternatives evaluation, purchase decision and post-purchase evaluation. Researchers have found four main sources of searching information——friends, crowds, advertisement and experience, proving the influences of them on consumers’ purchase behaviors. However, they don’t focus on the different influences of information sources on mainstream and niche product. Hence, it is worthwhile to extend the consumers purchase decision theory. In music industry, will the impacts of friends and crowds information on consumers’ behaviour change when they listen to niche music, comparing to the mainstream music?

Therefore, to solve these gaps and concerns, we consider friends and crowds influences separately, analyzing the changes between mainstream and niche sample as well. The rest of this research is organized as follows. Firstly, the literature review is presented. Then, the conceptual framework and hypotheses development are described. Thirdly, the methodology is presented, including the data collection method, research models, measurement of variables and data analysis. Finally, this research is concluded.

2 Literature Review: The long tail

2.1 Friends and crowds influence on users’ behaviours

Current researches have focused on the friends and crowds influence on users’ music listening behaviours, while most of them offer little guidance on how friends and crowds affect online music purchasing, studying the influence of crowds on outcome variables. These studies suggests that the wisdom of crowds can be a good indicator showing whether customers decide to buy or not to buy a specific item in online environment since user generated content such as customer reviews and ratings
provide diagnostic value to customer’s decision-making (Jiang and Benbasat, 2004; Kempf and Smith, 1998; Pavlou and Fygenson, 2006). However, the research on online customer reviews and ratings show equivocal results. For example, a couple of studies demonstrated that customer reviews and ratings can have a positive influence on performance outcomes (e.g., Clemons et al., 2006; Delarocas et al., 2007; Ghose and Ipeirotis, 2006), yet several other researchers indicated that user generated content such as customer reviews and ratings can produce biased decisions (e.g., Chevalier and Mayzlin, 2006; Duan, Gu, and Whinston, 2009) due to social influence bias (Muchnik, Aral, Taylor, 2013). The mixed results call into question the reliability of “wisdom of crowds”. Therefore, researchers are turning to the wisdom of both crowds and friends and assessing their impact on outcomes variables. Prior studies have generated insightful observations regarding how crowds and friends impact customers decision making. For example, Lee, Tan and Hosanagar (2011) found that higher crowds ratings tend to increase the rating of followers ratings for products, but a friend’s ratings have less impact on his or her friends rating for products. Abbassi, Aperjis, and Huberman (2012) found that negative opinions from friends are more influential than positive opinions in customers’ decision making, and people exhibit “more random” behavior in their choices when the decision involves less cost and risk. Despite the contributions from prior research, it is still unclear how crowds and friends differentially affect performance outcomes such as sales, consumers’ purchasing considering the limited evidence (Abbassi, Aperjis, and Huberman, 2012; Lee, Tan and Hosanagar, 2011).

2.2 The long tail

Long tail means distributions of numbers is the portion of the distribution having a large number of occurrences far from the "head" or central part of the distribution. Since the long tail phenomenon in the distribution of product sales was first observed through the comparison between offline and online sales (Anderson 2006, Brynjolfsson et al. 2006, Tucker and Zhang 2007), many studies have focused on this phenomenon both in offline and online area, especially in music world (e.g., Gaffney, 2009; Dewan et al, 2012) and have found many interesting conclusions: Chen et al. (2004) took Amazon.com as empirical data source and studied that the influence of recommendation on non-popular books is greater than the influence on popular books; Zhou and Duan (2012) examines the impact of online user reviews and product variety on the long tail and superstar phenomena in the context of online software down-loading, finding that the overall impact of the increased product variety helps niche products to get more downloads.

Especially, some researchers have proved the different influence of online feedback (friends and crowds) on user behavior between mainstream and long tail product. For example, Zhao et al. (2008) found that positive user reviews have a stronger impact on the mainstream than on the niches and that negative user reviews hurt niche products more. Dewan and Ramaprasad (2012) analyzed the relationship between music blogging and full-track sampling and found that the association of sampling with blog popularity and music popularity are both stronger in the tail compared with the body of the sales distribution. However, the researchers above always focused on the influence of crowds on the long tail, scarcely considered both crowds and friends impact. Therefore, our study pays attention to both crowds and friends online feedback and studies the changes of these two factors between mainstream and long tail.
The long tail measurement has raised disagreements among researchers, which will influence the different results between mainstream and the niches. Some researchers, such as Brynjolfsson et al. (2011) identified the long tail by comparing the demand shares accounted for by the mainstream and the niches differentiated by the amount of sales. Others examined the long tail phenomenon using the relative measures such as percentage of sales. What’s more, different researchers have got lots of evidence to prove the effectiveness of each method. Tan and Netessine (2009) considered that the relative measure is more appropriate because it controls for the significant increase in number of products over time or across channels, while Brynjolfsson et al. (2010) pointed out that the two measures have their strengths and weaknesses, so the study should be depended on actual situation. Our track sample was extracted from the list of Top 1000 Tracks in the billboard and users rarely notice the tracks in the long tail, so we set the cut-off of bottom 50% tracks as niche music and others as mainstream music.

3 Conceptual Framework and Hypotheses Development

3.1 Conceptual Framework

Consumer’s perceived value of online music is a significant factor in predicting the purchase intention of buying online music (Chu and Lu, 2007). In web2.0 era, consumer’s perceived customer value for online music comes from the satisfied needs, the sound quality and the driven powers from online feedback, which represents user, product and online feedback factors. In our studies, we mainly focused on the different influence of online feedback, which produces a great impact on product evaluations and purchase decisions (Tong et al. 2007). In previous researches, several researchers divide peers into two categories of group: the crowds or friends (Lee et al. 2011, Abbassi et al. 2012). Accordingly, two sources of online feedback information are summarized. One is the reviews or listening behavior information from the group of crowds, the other one is the reviews or listening behavior information from the group of friends. Hence, the online feedback factors can be divided into crowd feedback factors and friend feedback factors. The relations among online feedback, user and product factors are shown in Figure 1.

![Figure 1. Factors Influencing Music Listening Behavior of User](image)

Online music is a typical long tail good, which means that only a few tracks are very popular with huge scale of listeners while vast majority of tracks only have a small amount of listeners. Therefore, we call the few very popular tracks as mainstream tracks, and the huge unpopular tracks as niche
tracks, exploring the different effects of online feedback from crowds and friends on user’s listening on different types of music, mainstream music and niche music. What’s more, the friends also have different types, which can reflect the degree of friends and users. Hence, we introduced the taste match between friends and users to study the different effects of friends with different tastes and friends with similar tastes on user’s listening respectively. The conceptual model of our study is presented as follows in Figure 2.

Figure 2. Conceptual Framework

3.2 Hypotheses Development

3.2.1 Online Feedback from Friends

In the study of social network analysis, there is a common network with a dense, cohesive core and a sparse, unconnected periphery, which is called core/periphery structure (Borgatti 1999). On the basis of social differentiation, Bourgeois and Friedkin (2001) studied the interpersonal ties and they found the distribution of actors in core/periphery structure is generally in multidimensional social space, in which the likelihood and strength of an interpersonal tie are negatively associated with the distance that separates the positions of actors in the social space of a group. Based upon the subjective “type” of social relation, Granovetter (1983) categorized interpersonal ties into strong ties and weak ties, where friendship and family relationship are the typical examples of strong ties.

Being in a circle of strong ties, friends are easier to obtain a user’s trust and have an impact on the user. In a research involving 7,000 users in seven European countries, Kotler (2000) stated that 60% of these users’ purchase for a new brand is affected by their family members and friends. Bansal and Voyer (2000) pointed out the professional degree of information source had positive impacts on the changes in recipients’ attitudes but users tended to believe information with reliable information sources such as information from friends. Previous work on social impact theory (Latané 1981) has demonstrated that the strength of social impact is associated with the distance on the space. Therefore, it will be different between the impact of friends on user’s behavior and that of crowd since friends have closer distance to a user. Thus we have:

H1: Friend’s online feedback is positively associated with a user’s online listening behavior.

3.2.2 Different Friend’s Effects on a user’s online Music Listening

As a common sense, trust is an important factor that affects the degree of success in e-commerce transaction network (Quelch and Klein, 1996, Keen, 2010). Lu (2011) based on the analysis of the trust formation mechanism, and found through empirical researches that the degree of familiarity between members in a group, the perceived similarity (whether members have common interests or
hobbies), structural guarantee and trust inclination can significantly influence the trust between members in a virtual community, which further impact users’ purchase intention. That is to say, the higher similar taste between members, the more trust one person would perceive. Furthermore, according to informational influence theory (Myers and Arenson, 1972), individual will lean towards the “persuasive” opinion without private information processing, which cause final decision to the same as the one it trust under the effect of it (Kaplan and Martin F, 1977). Tom Hayes (2010) stated that a group set up by similar interest is often very xenophobia and high-interacted. They adopted accordant behaviours and communication methods and often followed group’s opinion instead of their own. This indicated that a user’s behaviour has a significant relationship with the homogeneity level of the group. Therefore, it is logical to launch that if one incline to be affected by others’ behaviour, the person with similar interest and taste would be the preferred. Hence, we get the following hypothesis:

H2: Compared to friends with different music taste, friends with similar music taste have much stronger effects on a user’s online music listening.

3.2.3 Online Feedback from Crowds

Many studies have pointed out that the herding effect resulted from crowd behavior on Internet is an important factor affecting users behavior (Lee et al. 2011; Duan et al. 2009). It is obvious that online music is an experience product. Each user has an anonymous identity on the virtual Internet, which means user’s track listening behavior will not be exposed under others’ view except his or her friends. According to anonymous effect theory proposed by America psychologist Zimbardo in his famous simulated prison experiment (Haney et al. 1973), it is speculated that the loss of identity would make a user feel a lower importance to others, as well as the farther closeness in space and psychological to other users. That is to say, when a user makes a listening decision, the impact of informational social influence from crowd is stronger, but the impact of normative social influence from crowd is relatively weaker. Generally, due to the strong positive informational social influence (although the normative social influence is weaker), it can get the following hypothesis:

H3: Crowd’s online feedback is positively associated with a user’s online music listening behavior.

3.2.4 Crowds and Friends’ Different Effects on User’s listening behavior

From the hypotheses above, we can see that both friends and crowds online feedback can influence users’ behavior. However, in online feedback from crowds and from friends, which one is more powerful? Aronson (2005) has found informational social influence and normative social influence. Informational social influence is defined as an influence to accept information obtained from another as evidence about the reality with regard to his/her own behavior, and normative social influence is defined as an influence to conform with positive expectations of another, leading to feelings of self-esteem or self-approval (Deutsch and Gerard 1955; Aronson et al. 2005). According to the research on informational social influence, the more uncertain the individual is about the correctness of his/her judgment, the more urgent of the decision-making situation is, and the more professional the others are, the more likely he/she is to be susceptible to informational influences and follows the crowd decisions in making his/her judgment (Allen 1965; Tesser et al. 1983; Baron et al. 1996; Aronson et al. 2005). But with the users’ music playing number increasing, the uncertainty may
decrease and the normative social influence may have a greater impact. Thus we have the following hypothesis:

H4: Compared to crowds’ online feedback, friends’ online feedback has much stronger positive impact on users’ music listening behavior.

3.2.5 Crowds and Friends’ Different Effects on Different Types of Music’s online Listening

Since Anderson (2006) proposed Long Tail theory, many studies have focused on this phenomenon both in offline and online area, especially in music world (e.g., Bhattacharjee et al., 2007; Gaffney, 2009; Dewan et al, 2012). Gaffney (2009) pointed out that social networking sites offer ways of making the Long Tail more visible. And as Dewan (2012) demonstrated, the mainstream music is more like a search good, whereas niche music is closer in nature to experience goods. That means when consumers listen to mainstream music, they may have a prior awareness about the track or the singer, so they do not need much more online feedback information when they play the music. But they are less likely to have had prior experience to niche music and they may have much more uncertainty to niche music compared with mainstream music. Therefore, consumers will obtain more information from others (both friends and crowds) to decrease their uncertainty. These lead to the following hypotheses.

H5: Compared to mainstream music, when it comes to niche music, the positive effect of friends’ online feedback on a user’s online music listening turns stronger.

H6: Compared to mainstream music, when it comes to niche music, the positive effect of crowds’ online feedback on a user’s online music listening turns stronger.

4 Research Methodology

4.1 Data collection

Our data is collected from Last.fm website, which was founded in the United Kingdom in 2002 and is one of the world’s largest online music community platforms. The reason why this website is chosen as the data source is that users’ information, online feedback information and track quality related information can be obtained from this website directly and efficiently. Using "Audioscrobbler", Last.fm builds a detailed profile of each user’s musical taste by recording details of the tracks that the user listens to, either from Internet radio stations, or the user’s computer or many portable music devices. Each registered user has a profile which displays the most recent songs they have played, and regularly updated charts of their top artists and songs. In addition, Last.fm also provides several communication mechanisms for those interested in using the site socially. Once the friendship is approved by both relational partners, each appears in the others’ publicly visible friends list and users can see their friends’ most recent listened track list (Baym and Ledbetter, 2009).

A crawling program was developed to get the data by the open API provided from Last.fm. The algorithm of the data crawling program includes three phases. The first phase is to find target tracks and seed users. To minimize history marketing bias, new tracks are initially collected on the top 1000 chart as the target tracks. Next, we seed users who listened to these target tracks were randomly selected by using “*** (username) is listening to this track” function. The second phase is to acquire target users. To observe users who have both influence effect from crowds and friends, those users
who have at least one friend listened to one target track are selected as the target users. The third phase is to get the users’ online music listening behavior data. Meanwhile the user demographic and track characteristic data were also collected.

As in sum, 190 tracks, 1000 target users and more than 40000 related friends of these target users are random selected finally (see Table 1) at March, 2013. Each target user is with independent characteristics and dispersed in the sample. All the listening behavior information of 1000 target users and 42283 friends were extracted during the period of April 19, 2013 to June 28, 2013. A total of 1900000 items of track listening behavior information were collected. After examination, there is no essential difference between the sample users and the overall population. The gender ratio was 62% male, 38% female and the average age 24.4 years.

<table>
<thead>
<tr>
<th>Data items</th>
<th>Description</th>
<th>Observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>User</td>
<td>620 male, 380 female Age: 1-112 (mean 24.4, standard dev. 7.2)</td>
<td>1000</td>
</tr>
<tr>
<td>Tracks</td>
<td>Rank on Last.fm: 5-1000 (mean 415.3, standard dev. 298.1) Related artist: 47 Artists</td>
<td>190</td>
</tr>
<tr>
<td>Time period (units: week)</td>
<td>Week 16-25 in the year of 2013 (10 weeks in total) Week16:1th; Week25:10 th: From April 19, 2013 to June 28, 2013</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 1. General Description of Collected Data (1900000 items in sum)

4.2 Variables and Measures

<table>
<thead>
<tr>
<th>Variable</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Playcntijt</td>
<td>The number of user i played the track j at time t (units: week).</td>
</tr>
<tr>
<td>log_FRIENDLISTENNUMijt</td>
<td>The number of user i’s friend ever listened track j at time t (units: week).</td>
</tr>
<tr>
<td>log_LISTENERSjt</td>
<td>The total number of listeners of music track j at time t (units: week).</td>
</tr>
<tr>
<td>log_(shouts/listeners)jt</td>
<td>The average number of shouts (reviews) peer listener of track j at time t (units: week).</td>
</tr>
<tr>
<td>TasteMatchi</td>
<td>A taste match score between user i and his friends, we consider score&gt;=0.5 as similar music taste and the score&lt;0.5 as different music taste. The variable of TasteMatch means the total number of user i’s friends who have similar music taste with the user.</td>
</tr>
<tr>
<td>Agei</td>
<td>Self-reported age of user i</td>
</tr>
<tr>
<td>Genderi</td>
<td>Self-reported gender of user i (male=1).</td>
</tr>
<tr>
<td>log_UserPlayCnti</td>
<td>The total number user i played on all songs on Last.fm.</td>
</tr>
<tr>
<td>Log_Preferenceij</td>
<td>The preference of user i for the artist of music track j.</td>
</tr>
<tr>
<td>ArtistYearsj</td>
<td>The duration (units: years) from the artist of track j released her first album to time t.</td>
</tr>
<tr>
<td>log_Durationj</td>
<td>The duration from track j released to March 22, 2013 (data collection initial time point).</td>
</tr>
</tbody>
</table>

Table 2. Description of Dependent Variable

Dependent Variable

The data on Last.fm contains listening history of each user, providing a great advantage to track the online user’s choice of listening on each particular track. Intuitively, the play count can indicate the
user’s online music listening behavior, which can show a good reference to predict to the user’s further listening decision. Therefore, in this study, the online user’s choice of listening music is proposed by the variable named \( \text{Playcnt}_{ijt} \), presenting the number of user \( i \) played the track \( j \) at time \( t \).

Independent Variables

In order to validate the proposed hypotheses, the independent variables are considered in this study from two dimensions: friends level and crowds level.

(1) Friend’s variable

The independent variable in friend level indicates the influences from users’ friends on user’s listening choice. We set the number of friend listeners of the track \( \log(\text{FriendListenNum}_{ijt}) \) as the independent variable on friend level, which presents the impact of online feedback from friend on user’s listening behavior.

(2) Crowds’ variables

In this study, the number of track listeners \( \log(\text{Listeners}_{jt}) \) and the average number of track reviews per listener \( \log(\text{Shouts/Listeners}_{jt}) \) are utilized to represent the impact of online feedback from crowd on user’s listening behavior. Moreover, the variable of the influence coming from crowd, measured by the number of the track listeners \( \log(\text{Listeners}_{jt}) \) and the average number of track reviews per listener \( \log(\text{Shouts/listners}_{jt}) \), are acquired dynamically week by week from April 19, 2013 to June 28, 2013, 10 weeks involved. The calculations of the online feedback variables are shown in Table 3.

Control Variables

We set user’s demographic information such as age \( \text{Age}_i \), gender \( \text{Gender}_i \), the total count of tracks ever played by user \( \log(\text{UserPlayCnt}_i) \), user’s listening preference \( \log(\text{Preference}_i) \), the experience of the artist \( \text{ArtistYears}_j \) and duration of track released \( \log(\text{Duration}_j) \) as six control variables.

5 Data Analysis and Results

5.1 Descriptive Analysis

The descriptive statistical analysis of the variables is as shown in Table 3. As seen from the table, the variable of \( \text{Playcnt}_{ijt} \) is quite smaller than the standard deviation value (mean value of \( \text{Playcnt}_{ijt} = 0.476 \); standard deviation value of \( \text{Playcnt}_{ijt} = 5.426 \)), indicating that the data of users’ music play count is over discrete and we choose the Negative Binomial Regression Model is more suitable. Meanwhile, the average age of the users is around 24 with standard deviation value of 7.153, indicating the most of the users in our sample are young listeners. Additionally, 62% of the users in our sample are male (Male =1, mean =0.62). From the general data descriptive statistics, it shows a more apparent trend of younger and masculine users on the Last.fm music website.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std.Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{Playcnt}_{ijt} )</td>
<td>0.476</td>
<td>5.426</td>
<td>0</td>
<td>771</td>
</tr>
<tr>
<td>( \log(\text{FriendListenNum}_{ijt}) )</td>
<td>1.467</td>
<td>4.699</td>
<td>1</td>
<td>3326</td>
</tr>
<tr>
<td>( \log(\text{Listeners}_{jt}) )</td>
<td>90723.1</td>
<td>73087.92</td>
<td>12845</td>
<td>634686</td>
</tr>
<tr>
<td>( \log(\text{Shouts/listners}_{jt}) )</td>
<td>87.135</td>
<td>161.989</td>
<td>1</td>
<td>1554</td>
</tr>
<tr>
<td>( \text{TasteMatch}_i )</td>
<td>29.8</td>
<td>34.438</td>
<td>0</td>
<td>255</td>
</tr>
<tr>
<td>( \text{Age}_i )</td>
<td>24.386</td>
<td>7.153</td>
<td>1</td>
<td>113</td>
</tr>
<tr>
<td>( \text{Gender}_i )</td>
<td>0.620</td>
<td>0.485</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 3. Data Descriptive Statistics

<table>
<thead>
<tr>
<th>Preference$_{ij}$</th>
<th>3.748</th>
<th>8.544</th>
<th>1</th>
<th>36.857</th>
</tr>
</thead>
<tbody>
<tr>
<td>UserPlayCnt$_i$</td>
<td>37130.57</td>
<td>40306.79</td>
<td>1</td>
<td>276421</td>
</tr>
<tr>
<td>ArtistYears$_j$</td>
<td>10.684</td>
<td>12.486</td>
<td>0</td>
<td>49</td>
</tr>
<tr>
<td>Duration$_j$</td>
<td>241352.6</td>
<td>74578.9</td>
<td>31000</td>
<td>481999</td>
</tr>
</tbody>
</table>

5.2 Model Specifications

The dependent variable of this study is an online user’s music listening behavior. We estimate the impact of users and their online feedback, and the impact of music factor on users’ music listening choice to verify the hypotheses proposed above. The general model equation is shown as equation (1):

\[ \text{Playcnt}_{ijt} = \gamma \text{Friend}_{ijt} + \rho \text{Crowds}_{ijt} + \beta \text{Control} + \epsilon_{ijt} \]  

(1)

Where, Playcnt$_{ijt}$ is the number of user $i$ choose to listen music track $j$ at time $t$, Friend$_{ijt}$ is the friend factors influence which effects on user $i$ to listen track $j$ at time $t$; Crowds$_{ijt}$ is the crowds factors influence which effects on user $i$ to listen track $j$ at time $t$; Control means the control variables in the model. $\beta, \gamma, \rho$ are the coefficients of the model and $\epsilon_{ijt}$ is a residual error term, indicating all the potential influencers that the models have not shown.

Because the Playcnt$_{ijt}$ is a discrete count data, we choose count data model as our research model. As usual, the basic count data model is Poisson Model. But the limitation of this model is that the mean value of data should equal to the standard deviation value. So if the standard deviation value is quite bigger than the mean value, the Negative Binominal Mode may be more suitable (The data from Table 3 has proved it). The basic Poisson Regression Model is shown just as equation (2) and (3):

\[ P(Y_{ijt} = y_{ijt} | X_{ijt}) = \frac{e^{-\mu_{ijt}} \mu_{ijt}^y}{y!} \]  

(2)

\[ E(Y_{ijt} | X_{ijt}) = \mu_{ijt} = e^{x_{ijt} \beta} \]  

(3)

To solve the limitation of Poisson Model, we further suppose that $\mu$ satisfies the assumption $\mu \sim \Gamma(r, \delta)$, and $r = e^{x_{ijt} \beta}$. The probability of $y_{ijt}$ is shown as follows:

\[ P(Y_{ijt} = y_{ijt} | X_{ijt}) = \int_0^\infty \frac{e^{-u_{ijt}} u_{ijt}^{y_{ijt}}}{y_{ijt}!} f(u_{ijt}) du_{ijt} = \frac{\Gamma(y_{ijt} + \delta)}{\Gamma(\delta) \Gamma(y_{ijt} + 1)} \left( \frac{\delta}{1 + \delta} \right)^{y_{ijt}} \left( \frac{1}{1 + \delta} \right)^{\delta} \]  

(4)

Finally, we establish equation (5) to examine the influence of friends and crowds on users’ online music playing behavior for the mainstream and tail subsamples (mainstream: Music Rank<=500; Tail: Music Rank>500), the formula is shown as follows:

\[ \text{Playcnt}_{ijt} = \exp(\gamma_1 \log(\text{FriendListNum}_{ijt}) + \gamma_2 \log(\text{ListenerNum}_{ijt}) \times \text{TasteMatch}_i + \gamma_3 \text{TasteMatch}_i + \delta_1 \log(\text{ListenerNum}_{ijt}) + \delta_2 \log(\text{Shouts/listeners})_i + \beta_1 \text{age}_i + \beta_2 \text{gender}_i + \beta_3 \log(\text{Preference})_i + \beta_4 \log(\text{UserPlayCnt})_i + \beta_5 \text{ArtistYears}_j + \beta_6 \log(\text{Duration})_j + \epsilon_{ijt} ) \]  

(5)

5.3 Results

In Table 4, the results of negative binominal regression estimation using equation (5) is presented. The column 1 shows the influence of friends factors on users’ music listening behavior. Column 2 shows the influence of friends on users’ music listening behavior with moderator variable. Column 3 shows the influence of crowds on users’ listening behavior and Column 4 is the full model.
To estimate the effect of friends, we analyze the friend’s listening number (log\_FriendListenNum\_{ijt}). From the full model in column (4) implies the coefficient of log\_FriendListenNum\_{ijt} is 2.068. This means that for one unit increase in the number of friend listeners, the log count of Playcnt\_{ijt} increases by 2.068 times, indicating that higher friend listen number for the track increases the likelihood of the user’s listening behavior of that track. This provides evidence for H1.

Meanwhile, we notice that the variable of TasteMatch\_i scarcely has a moderator impact on users’ friend listeners(Model1: \( \gamma_1 = 2.071 \); Model2: \( \gamma_2 = 2.137 \)). In addition, the interaction impact between friend listeners and the quantity of friends who have preference similarity with user \( i \) (TasteMatch\_i) is negative (\( \gamma_3 = -0.003 \)), hardly supporting H2. This result break the traditional thinking, proving that the friends with different music taste have much stronger effects on a user’s online music listening. It can be explained by Shannon’s information theory, which identified the information of the thing that can eliminate random uncertainty. The different taste friends will give more information quantity and consumers will take more action to decrease the uncertainty, compared to the influences of friends who have similar taste.

And to estimate the effect of crowds, we analysis the listener’s number (log\_Listeners\_{jt}) and effective comments peer listeners (log\_(Shouts/listners)\_{jt}). The results from Table 4 column(4) shows that the influence of crowds, both two factors, are positive( \( \delta_1 \log\_Listeners\_{jt} = 1.224 \); \( \delta_2 \log\_(Shouts/listners)\_{jt} = 0.119 \)). This provides evidence for H3.

To compare the friend’s online feedback with crowd’s online feedback, we also standard the data. The result is shown in Table 5. The analysis results indicate that compared to crowds’ online feedback, friends’ online feedback has much stronger, positive impact on users’ music listening behavior (\( \gamma_1 \log\_FriendListenNum\_{ijt} = 0.414 \); \( \delta_1 \log\_Listeners\_{jt} = 0.384 \); \( \delta_2 \log\_(Shouts/listners)\_{jt} = 0.119 \)), supporting evidence for H4.
Finally, we turn to the estimation results for the mainstream and niche subsamples, presented in Table 6. We find support for both H5 and H6. That is both friend listen number (log_FriendListenNum$_{ijt}$) and listeners (log_Listeners$_{jt}$) have a stronger association with sampling in the tail compared with the mainstream (p<0.01), which means the both friends and crowds are particularly important signals of online feedback for niche music.

<table>
<thead>
<tr>
<th></th>
<th>Mainstream (Rank&lt;=500)</th>
<th>Niche (Rank&gt;500)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log_FriendListenNum</td>
<td>1.946*** (0.032)</td>
<td>2.414*** (0.059)</td>
</tr>
<tr>
<td>TasteMatch</td>
<td>0.005*** (0.000)</td>
<td>0.001*** (0.000)</td>
</tr>
<tr>
<td>Log_FriendListenNumTasteMatch</td>
<td>-0.004*** (0.000)</td>
<td>-0.004*** (0.001)</td>
</tr>
<tr>
<td>Log_Listeners</td>
<td>0.642*** (0.017)</td>
<td>1.623*** (0.024)</td>
</tr>
<tr>
<td>Log(shouts/listeners)</td>
<td>0.147*** (0.005)</td>
<td>-0.027*** (0.007)</td>
</tr>
<tr>
<td>Age</td>
<td>-0.075*** (0.001)</td>
<td>-0.061*** (0.002)</td>
</tr>
<tr>
<td>Gender</td>
<td>0.433*** (0.010)</td>
<td>0.287*** (0.015)</td>
</tr>
<tr>
<td>Log_preference</td>
<td>2.933*** (0.010)</td>
<td>3.040*** (0.014)</td>
</tr>
<tr>
<td>Log_Userplaycnt</td>
<td>1.229*** (0.010)</td>
<td>1.139*** (0.014)</td>
</tr>
<tr>
<td>Artistyears</td>
<td>-0.039*** (0.000)</td>
<td>-0.007*** (0.001)</td>
</tr>
<tr>
<td>Log_Duration</td>
<td>0.258*** (0.030)</td>
<td>-0.549*** (0.041)</td>
</tr>
<tr>
<td>Constant</td>
<td>-10.290*** (0.192)</td>
<td>-12.246*** (0.247)</td>
</tr>
<tr>
<td>Inalpha</td>
<td>2.360*** (0.005)</td>
<td>2.441*** (0.008)</td>
</tr>
<tr>
<td>AIC</td>
<td>0.718</td>
<td>0.564</td>
</tr>
<tr>
<td>Obs.</td>
<td>1,189,952</td>
<td>710,048</td>
</tr>
</tbody>
</table>

Table 6. **Negative Binomial Regression Results——Mainstream vs. Niche**

### 5.4 Correlation Analysis and Robustness Checks

We conducted a correlation analysis and found no highly correlated pairs of variables in our model, with the highest pair-wise Pearson correlation between variables registered at 0.27.

Further, to verify the robustness of our results and associated data operationalization methods, we performed three series of robustness checks whose results are shown in Table 7. In the table below, we consider different sets of partition, which based on a threshold of 300 for the music rank, checking H1 to H5. As can be seen, the results for both sets support hypothesis above.
<table>
<thead>
<tr>
<th></th>
<th>Mainstream (Rank&lt;=300)</th>
<th>Niche (Rank&gt;300)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log_FriendListenNum</td>
<td>1.866*** (0.039)</td>
<td>2.236*** (0.042)</td>
</tr>
<tr>
<td>TasteMatch</td>
<td>0.003*** (0.000)</td>
<td>0.005*** (0.000)</td>
</tr>
<tr>
<td>Log_FriendListenNumXTasteMatch</td>
<td>-0.002*** (0.000)</td>
<td>-0.005*** (0.000)</td>
</tr>
<tr>
<td>Log_Listeners</td>
<td>0.773*** (0.022)</td>
<td>1.080*** (0.018)</td>
</tr>
<tr>
<td>Log(Shouts/listeners)</td>
<td>0.117*** (0.006)</td>
<td>0.057*** (0.005)</td>
</tr>
<tr>
<td>Age</td>
<td>-0.077*** (0.001)</td>
<td>-0.062*** (0.001)</td>
</tr>
<tr>
<td>Gender</td>
<td>0.476*** (0.013)</td>
<td>0.286*** (0.012)</td>
</tr>
<tr>
<td>Log_preference</td>
<td>2.892*** (0.013)</td>
<td>3.009*** (0.011)</td>
</tr>
<tr>
<td>Log_Userplaycnt</td>
<td>1.172*** (0.012)</td>
<td>1.261*** (0.011)</td>
</tr>
<tr>
<td>Artistyears</td>
<td>-0.041*** (0.000)</td>
<td>-0.003*** (0.001)</td>
</tr>
<tr>
<td>Log_Duration</td>
<td>0.361*** (0.033)</td>
<td>-0.582*** (0.037)</td>
</tr>
<tr>
<td>Constant</td>
<td>-11.224*** (0.228)</td>
<td>-9.451*** (0.212)</td>
</tr>
<tr>
<td>lnalpha</td>
<td>2.392*** (0.006)</td>
<td>2.386*** (0.006)</td>
</tr>
<tr>
<td>AIC</td>
<td>0.744</td>
<td>0.600</td>
</tr>
</tbody>
</table>

Table 7. Robustness Checks

6 Discussion and Conclusion

Take online music as an example, we get the antecedents of purchase intention in Web2.0 era and identify its dynamic features. There are several findings of this work:

Firstly, taking online music as an example, we found that friends online feedback and crowds online feedback both have the positive influence on users’ music listening behavior. That is, giving more evidences for the crowds online feedback has a positive influence on user’s purchase behaviour (Jiang and Benbasat, 2004; Kempf and Smith, 1998; Pavlou and Fygenson, 2006), the same as friends online feedback (Kotler, 2000; Voyer, 2000). Secondly, we find that friends’ online feedback has much stronger positive impact on users’ music listening behaviour compared to crowds’ online feedback, which answers the question of what is the different between friends and crowds online feedback on users’ behaviour. However, our studies have the different conclusion with Tan and Hosanagar (2011), who found that the crowds and friends have different influences on users’ rating behaviour. The possible explanation may be because of the research subjects. All in all, our research means that although the crowds can impact the users’ behaviour but the online sales company should pay more attention to the influence of friends’ influences. Thirdly, we found that the friends with different music taste have much stronger effects on a user’s online music listening, breaking the traditional thinking. This phenomenon can be supported by Shannon’s information theory that friends with different taste with us can give much higher information quantity for us, thus they influence us much higher than friends with similar taste. Fourthly, our research shows that the influence of both friends and crowds online feedback will become stronger to niche music, compared to mainstream music. The results suggest that online feedback is a driver of the long tail in users’ listening decision, reminding us of the empirical analysis of Dewan and Ramaprasad (2012), who have shown that social media is a driver of the long tail.

The limitation of the research is obvious. Our results are based on only one set of data on online music listening. Future research on online film, online video, online reading, online games etc. are
needed to explore whether our research results can be extended to all the experience digital good or not. Also, future research on search goods is necessary to compare the difference of value-intention framework in Web 2.0 between experience goods and search goods.

References


