Paying for Content or Paying for Community?
The Effect of Social Computing Platforms on Willingness to Pay in Content Websites[1]

Gal Oestreicher-Singer and Lior Zalmanson
Faculty of Management; Tel Aviv University

ABSTRACT

One of the dominant strategies for content providers in the digital era is the two-tiered business model, in which basic services are offered for free and additional, premium services are offered for a fee. However, converting users 'from free to fee' remains challenging, and an effective conversion strategy is crucial for a firm's success. Previous attempts to understand consumers' willingness to pay for content services have focused on the content consumption experience of the user. We study how consumers' engagement in a website's online community affects their willingness to pay for premium services.

Focusing on the proprietary content industry, we use data from Last.fm, a site offering both music consumption and online community features. The basic use of Last.fm is free, and premium services are provided for a fixed monthly subscription fee. Although the premium services on Last.fm are mainly aimed at improving the content consumption experience, we find that willingness to pay for premium services is strongly associated with the level of community participation of the user.

Drawing from the literature on levels of participation in online communities, we show that consumers' willingness to pay increases as they climb the so-called 'ladder of engagement' with the website. Moreover, we find that willingness to pay is more strongly linked to community participation than to the volume of content consumption. We control for self-selection bias by using propensity score matching. We extend our results by estimating a hazard model to study the effect of community activity on the time between joining the website and the subscription decision.

Our results constitute new evidence of the importance of facilitating user participation in content websites in order to increase willingness to pay. We discuss how firms can effectively leverage online communities in their business strategy to build new and effective business models that can enhance firm performance.

Keywords: Business Models; Online Communities; Content Websites; Digital Business Strategy

[1] We thank seminar participants at UC Irvine, New York University, International Conference on Information Systems, the Marketing Science Conference, and the Workshop on Information in Networks. We would also like to thank Jacob Goldenberg, Vijay Gurbaxani, Arun Sundararajan, and Barak Libai for helpful discussions. Financial support from the Google Inter-University Center for Electronic Markets and Auctions and Rothschild-Caesarea fund is gratefully acknowledged.

Paying for Content or Paying for Community?
The Effect of Social Computing Platforms on Willingness to Pay in Content Websites

INTRODUCTION

The online environment has created new challenges for content providers. Whereas in the past, a newspaper or a music label could simply charge consumers for the content it offered, the widespread availability of free content online has reduced consumers' willingness to pay for the privilege of reading an article or listening to a song. Content providers, in turn, have been compelled to seek new business strategies.

One of the dominant strategies for content providers in the digital era is the *two-tiered* business model, commonly referred to as *freemium*, in which basic services are offered for free and additional, premium services are offered for a fee. This model, which was discussed extensively in Chris Anderson's book, *Free* (Anderson 2009) as well as in recent academic works (see Teece 2010), is well suited for digital products, owing to their unique cost structure and low delivery costs. It can also help mitigate the threat of digital piracy and increase a firm's consumer base. Variations of the two-tiered model have been adopted by firms such as Adobe, Skype, MySpace, and Flickr, as well as by numerous software companies (such as Linux, Firefox, and Apache) that operate in the open-source marketplace. This business model relies on the assumption that a fraction of consumers who use free services will eventually be converted into paying customers. In practice, however, conversion rates vary and are often very low, and firms continue to seek effective strategies for converting consumers 'from free to fee' (Teece 2010). Prevailing existing strategies to increase conversion include restrictions on free content, as well as various marketing efforts such as e-mail targeting or price promotions for limited periods (Pauwels and Weiss 2008).

Here we argue that an alternative conversion strategy may stem from developments in a different domain of the IT strategy arena: the implementation of social computing features. In recent years, online content providers have begun to invest in social computing platforms, which enable consumers to interact socially and to create online communities on websites (for a review of existing social computing features, see Parameswaran and Whinston 2007). Such platforms offer consumers a variety of opportunities not only to consume content but also to actively engage with the website. Some websites merely allow users to post comments, while others also provide an online community, wherein the consumer can create an on-site identity (in many cases, a personal page), make online friends, attend virtual social events, build a reputation, and interact with other consumers. Sites that rely extensively on such community features are also referred to as 'social media websites' (Agichtein et al. 2008).

Previous attempts to understand consumers' willingness to pay for content services have focused on the content consumption experience of the user, i.e., consumption habits, the user's perception of the quality of the product, past experience and available alternatives (more on this later). Herein we conjecture that a consumer's willingness to pay for premium services in a content website is strongly linked to his or her level of participation in the online community offered on the website. While it has been argued that consumers are not willing to directly pay for access to online communities (Nielsen 2010), participation may positively affect consumers' experience and their engagement with the site. Specifically, we find that consumers' willingness to pay increases with their engagement with the community, and that the decision to subscribe is more strongly linked to the consumer's community activity than to the amount of content he consumes.

Our research relies on the literature on participation patterns in online communities. Several studies have described consumers' usage of and involvement in community-based websites in terms of a ladder-type model of growing engagement. The mere act of consuming content—referred to as 'lurking'—can be perceived as the first stage of engagement. In the second stage, the user makes minor contributions to the website community, e.g., by rating or tagging content and by carrying out small-scale contributions. The third stage entails full participation in the site's community, including posting opinions, advice and content, as well as extensive social collaborations. Finally, the highest level of engagement involves taking responsibility by moderating and leading website communities.

As consumers climb up the ladder of engagement, they develop a deeper sense of commitment to the website (Bateman et al. 2010) and perceived ownership (Preece and Schneiderman 2009). Although, to the best of our knowledge, the effects of user engagement in online communities on willingness to pay have not been studied, we propose that such increased engagement with the website may translate into increased willingness to pay. Furthermore, it is possible that by choosing to offer more community features, the content provider gives its consumers a 'ladder to climb on' and hence the opportunity to reach higher levels of engagement, which content consumption alone cannot provide. Hence, providing a community on one's content website may encourage more consumers to switch from free to fee.

In this work we focus on websites that combine proprietary content with an open social arena. We use data from Last.fm, a proprietary content website that serves both as an online radio and as a social networking site and implements a two-tiered business model. Even though the premium services it offers mainly improve the proprietary-content consumption experience (for example, by increasing bandwidth), we find that willingness to pay for premium services is strongly associated with the user's level of social activity. Specifically, consumers who participate in the community (i.e., use features that enable them to contribute to the community) show a higher propensity to pay compared with users who do not use these features. Users who act as leaders in this community show even higher propensity to pay.

One of the main challenges in studying the effect of usage of new IT artifacts is that of selection bias and user heterogeneity. If the use of community features is indeed associated with increased willingness to pay, it may be unclear whether the relationship is one of causation or simply a result of self-selection (that is, consumers who use premium services may simply be more likely to participate in social programs). One of the main problems faced by researchers in the field is the difficulty in separating correlation from causation and untangling endogeneity in situations where treatment assignment (in this case, the use or non-use of social features) is not random. We use a matching algorithm that allows us to overcome the non-random treatment assignment. We are able to match a consumer who uses the community features —for example, posts a blog entry—with a consumer who has not used those features, but who is as likely as the former consumer (based on his or her demographic characteristics, music consumption, and social activity levels) to have posted a blog entry. We therefore create two comparable groups of consumers, with similar propensity to contribute, and compare their willingness to pay. Using propensity score matching we are able to control for some of the self-selection biases created by observational data and provide results as to the directional effect of IT-enabled features on willingness to pay.

We extend these results by estimating a hazard model to study the effect of community activity on the time between joining the website and the subscription decision. We find that users who are more active in the community will make the subscription decision sooner than users who are less active (or not active at all). Moreover, we find a strong association between group leadership and the subscription decision. These results provide yet another dimension to our findings, suggesting that a consumer's community activity is associated not only with an increased willingness to pay for a premium subscription but also with a shorter time window between joining the website and subscribing.

Our results constitute new evidence of the importance of introducing social computing—a new IT artifact—as a means of driving consumers' willingness to pay, providing insights into the causal effects of social engagement on consumers' decisions to purchase premium services.

RESEARCH BACKGROUND

Willingness to Pay for Online Content Services

Our work draws on and adds to the literature on willingness to pay for online content. Scholars and practitioners have noted that digital content companies find it difficult to charge their consumers for access to media services, including proprietary content such as music, movies, and newspaper articles (Dyson 1995; Picard 2000). Consumers' increasing tendencies to seek out better prices (Shankar et al. 1999), widespread piracy (Jain 2008; Rob and Waldfogel 2006), and the introduction of digital sharing platforms (P2P) (Asvanund et al. 2004; Bhattacharjee et al. 2007) have all introduced new challenges for the online content retailer (see also Bhattacharjee et al. 2003; and Gopal et al. 2004). Existing revenue models, which rely mostly on charging for content (also referred to as *Pay per Use)* and advertising, have proven challenging and insufficient in most cases (Lopes and Galletta 2006) as consumers demonstrate low willingness to pay (Chyi 2003; Nielsen 2010), and advertising seems to have declined in both reach and credibility (Clemons 2009).

In light of this literature and the low pricing practices of music (Shiller and Waldfogel 2008) and content retailers such as Apple iTunes, it is not surprising that many content websites operate under a two-tiered business model, wherein basic services are provided for free, and premium services are offered for a fee (Doerr et al. 2010; Hung 2010; Riggins

2003). This two-tiered model, also referred to as the "freemium business model" (Wilson 2006), has received wide attention from the press and is currently prevalent among content websites. It is especially suited for digital products, and specifically content websites, owing to their unique cost structure, low marginal costs, and potentially large consumer base. One should note that this model is different from the free sampling strategy, which has also received much attention in the marketing literature (Bawa and Shoemaker 2004; Chellappa and Shivendu 2005; Gedenk and Neslin 1999; Scott 1976). In the two-tiered business model, the free offer is usually not limited in time and is offered in parallel to the for-pay premium offer.

Switching consumers 'from free to fee' has proven challenging for businesses (Moe and Fader 2004). Previous researchers have tried to identify the factors that drive willingness to pay. Naturally, the prospect of gaining access to better content or service encourages users to subscribe to premium services (Ye et al. 2004). Other researchers have stressed that providing some content for free while limiting access to the rest may result in lower perceived value of the free content, causing lower demand levels (Brynjolfsson et al. 2003; Fitzsimons and Lehmann 2004; for opposing results see Zeithaml 1988), as well as slower growth of the consumer base for the free service (Pauwels and Weiss 2008). We add to the literature on willingness to pay by studying a website that offers premium services without limiting the accessible free content.

Adding Community to Content Websites: The New Value Proposition of Social Computing

Until recently, content websites were considered online manifestations of regular newspaper, radio and television channels. Content providers used the web only as an additional channel for their original yet traditional content offerings (O'Reilly 2005). In recent years, however, content providers have begun to exploit the properties of the online environment in order to enhance the content consumption experience. Web 2.0 is the term used to describe companies, among them content providers, that incorporate social computing platforms in order to increase consumer engagement through participation and interaction. Web 2.0 platforms support multi-directional content experiences by integrating online communities with content consumption. They provide opportunities for visitors to co-produce their own experiences on a website, for example, by creating and sharing their own web-logs, video clips, or personal radio stations.

Naturally, a successful community depends on the participation and contributions of its members (Butler 2001). An extensive stream of research has focused on why people choose to invest time and effort in participating in and contributing to online communities (e.g., Wasko and Faraj 2005). Over the last two decades this stream has yielded diverse and sometimes contradictory explanations for such behavior, including the following: increased recognition (Kollock 1999; Rheingold 1993), reciprocity (Kollock 1999; Wasko and Faraj 2005), sense of community (Kavanaugh 2003; Quan y Hasse et al. 2002) and altruism (Lakhani and von Hippel 2003). Recently, Ma and Agarwal (2007) showed how enabling different technological features in a website community can promote users' willingness to share knowledge. These findings shed light on why people share information, knowledge and advice; however, they do not provide insight into the evolution of user behavior over time or the processes users undergo as they become more engaged in the website, topics that are of key interest in this research. Furthermore, they do not link specific types of participation to willingness to pay or to monetary spending on a website.

Levels of Participation

In their seminal work on learning processes in communities of practice, Lave and Wenger (1991) proposed a characterization of community behavior over time. They noted that newcomers "become more competent as they become more involved in the main processes of the particular community. They move from legitimate peripheral participation to 'full participation'." (Lave and Wenger 1991, 37) More recently, there have been various attempts at creating more thorough frameworks that model users' behavior specifically in online community contexts. Amy Jo Kim (2000), for example, differentiates among several participation roles: (i) the **visitor**, who exhibits unstructured participation; (ii) the **novice,** who invests time and effort in order to become a (iii) **regular**, who displays full commitment; and (iv) the **leader**, who sustains membership participation and guides interactions of others. Li and Bernoff (2008) develop a ladder-type graph known as ‘social technographics profiling’, which uses findings from large-scale surveys to create profiles of online behavior. Preece and Schneiderman (2009) propose a ‘Reader to Leader’ framework with emphasis on different needs and values at different levels of participation. The different approaches are summarized in Table 1.

(Insert 'Table 1' here)

As can be easily noted, all frameworks start from a reader type, who only consumes content, and they progress to users who invest some time and effort in making small contributions and carrying out minor acts of participation and content organization; they continue with users who invest significant time and effort in community participation, and they culminate (in successful cases) with a member who creates significant content, leads, and moderates discussions in the community. In contrast to the inconclusive findings regarding the motivations of users in online communities, it seems that there is a high degree of consensus among academics and practitioners regarding the various stages of the user's membership lifecycle.

Why would one expect users to repeatedly participate in a community and climb the levels of participation within it? In a recent study, Bateman et al. (2010) offered an overarching theory, the commitment-based approach.

Bateman et al.'s study (2010) showed that users' behavior on content sites is directly linked to their commitment levels, as defined by the organizational commitment theory (Meyer and Allen 1991). Content consumption was shown to be linked to *continuance commitment*, commitment based on the calculation of costs and benefits. The few studies that have investigated lurkers—users who strictly consume content—found that these users report mostly information benefits. If a user's total level of benefits is lower than the cost of finding the right content, he or she is likely to discontinue use of the site (Cummings et al. 2002; Nonnecke and Preece 2000).

Community participation was found to be associated with *affective commitment*, which is a positive emotional attachment or 'feeling of belonging' to the community. In the traditional (offline) organizational commitment context, affective commitment was shown to develop through social exchanges and relationships that promote trust (Cook and Wall 1980) and feelings of being treated fairly by the community (Eisenberger et al. 1990). The practical effects of attention from the community have been demonstrated in recent research. Joyce and Kraut (2006) showed how a user's likelihood of posting is related to the properties of the replies he received to his initial posting. Lampe and Johnston (2005) found that a newcomer's probability of returning to a site is affected by the ratings given to her first post. Huberman et al. (2009) showed, in the context of YouTube clips, that increased attention leads to heightened contribution of content. Burke et al. (2009) quantitatively examined photo contributions on Facebook and found that direct feedback on content is one of the factors related to the volume of content that a user subsequently uploads.

Community leadership, the top level of user participation in online communities, was shown to be associated with *normative commitment* (Bateman et al. 2010)[2]. The organizational commitment theory defines normative commitment as a sense of obligation to the community, i.e., the user participates in the community because he feels he 'ought to'. Normative commitment can be influenced by repetitive social exchanges in which a person learns about other community participants' values such as loyalty (Wiener 1982), or it can develop when a person feels indebted to the community because the benefits he receives exceed his own contribution (Bateman et al. 2010). Leaders of online communities have been shown to contribute the largest number of comments and to be the most active (Cassell et al. 2006; Yoo and Alavi 2004). A study of leadership in Wikipedia's community showed that leaders use multiple discourse channels, utilizing many features of the site, in order to broadcast their messages (Forte and Bruckman 2008).

Because users' commitment to a website increases with their levels of contribution and participation, researchers have suggested that website design schemes and business models should strive to increase consumers' engagement with company-sponsored community features (Preece and Schneiderman 2009). However, as most research has dealt with communities that are owned and managed on a voluntary basis, it is not yet clear how increased user engagement can benefit commercial companies. Thus, the relationship between a successful business and a successful online community remains to be elucidated (Wirtz et al. 2010). In particular, the issue of willingness to pay or increased paid consumption as a result of community participation is not discussed in this literature.

[2] Not surprisingly, leadership behavior was also shown by Bateman et al. (2010) to be associated with a degree of affective commitment as well, stressing the cumulative nature of levels of participation.

Broadly related to our work is the marketing literature that links community participation to willingness to pay in the context of brand communities. Studies have shown that such communities create bonds between consumers and brands, and that these bonds are strong and lasting and help promote brand loyalty, in both the offline (Muniz and O'Guinn 2001) and online contexts (Bagozzi and Dholakia 2006; McAlexander et al. 2002). In this type of relationship the connection between the product (or its producer) and the consumer is much stronger than the mere transaction-oriented relationship (McAlexander et al. 2002). Community involvement also affects the individual's decision to purchase services from a brand (Jang et al. 2008), increases the likelihood of adopting a new product from the brand (Thompson and Sinha 2008), and increases willingness to engage in new product development (Füller et al. 2008). While a community on a content provider's website is not necessarily a brand community as defined by Muniz and O'Guinn (2001), the brand community literature provides additional support for our conjecture that community participation may lead to increased willingness to pay[3].

CONTEXT AND HYPOTHESES

The data for this research were taken from Last.fm, an online music radio site that also functions as a social community. The website was purchased by CBS for $280 million in 2007 and is one of the leading proprietary music websites. Last.fm offers music streaming services[4] and differentiates itself from other online radio services with the method it uses to recommend songs to its users (also called 'AudioScrobbler'): After analyzing the user's listening habits, the Last.fm engine searches for other site members with similar tastes and recommends their favorite songs back to the user.

[3] In a somewhat similar context, Algesheimer at el. (2009) studied the effect of community participation on bidding behavior in online auctions on eBay, with mixed results.

[4] Last.fm uploads songs to the website, and a user can listen to them using the site's downloadable radio software, or by using the music streams on the website directly.

While the site's core business is centered around providing music-listening capabilities, Last.fm also enables the user to create a personal profile page (similar to profile pages on other social networking websites), join groups (mostly based on musical taste), contribute to blogs by posting short articles, or take a lead role in groups and moderate content. Users can also add tags to artists, albums, and tracks by using chosen keywords and can create playlists (personalized 'radio stations') for others to enjoy.

Last.fm implements the two-tiered business model by offering its users two levels of membership. The first is regular registration (free service), which enables the user to create a personal profile page, listen to online radio, and use other site functions. The second is the paid subscription, in which subscribers pay a monthly fee of $3 for a package of premium services that include the following[5]:

- Improved infrastructure, including removal of ads from the subscriber's page and top-priority quality-of-service on web and radio servers.
- Extended listening options, including the capacity to listen to unlimited personal playlists on shuffle mode and to create a 'Loved Tracks' radio channel[6].
- Improved social status, including the ability to see who visited one's homepage on Last.fm. In addition, the user's subscription status appears on his or her personal page.

(Insert 'Figure 1' here)

The core conjecture of this paper is that community activity on a content site is linked to willingness to pay for premium services provided by that site. Following the discussion on

[5] In April 2009, Last.fm changed its business model in certain countries outside the US and currently allows only paying subscribers to stream label-owned music. However, in the UK, Germany and USA the model has not changed.

6 This is a playlist created by the site based on a user's tagging of songs as 'loved'.

levels of participation, we expect community participation to positively affect consumers' propensity to subscribe. Hence, our first hypothesis is:

H1: Consumers who participate more in the community are more likely to pay for a premium subscription.

In our specific context, “participation in the community” can entail any one of the following activities: joining groups, leading groups, publishing a post in a forum, and adding an entry to one's blog.

The literature on levels of participation treats content organization as a form of low-level community participation. Content organization is made up of small acts of structured contribution that can be perceived as adding value to the user's own content consumption but that can also add value to the community. These acts include, for example, tagging content with keywords to ease its discovery, or rating content in order to promote its popularity and reputation (Li and Bernoff 2008). Because these activities require only a small investment from the consumer, content organization can be considered as an extension of content consumption behavior, which is characterized by continuance commitment. These activities can also represent the initial steps of "community participation" behavior.

We therefore expect content organization activities to positively affect consumers' propensity to subscribe. Hence, our second hypothesis is:

H2: Consumers who participate more in content organization activities are more likely to pay for a premium subscription.

In our specific context, “content organization” can entail any one of the following activities: attaching tags to songs, tagging favorite songs as 'loved', and creating playlists (a list of songs to be listened to together).

Finally, it seems natural that a consumer’s willingness to pay should be associated with content consumption levels. This proposition is supported by literature on willingness to pay for content, which has shown that "heavy" users in terms of content have higher willingness to pay (Ye et al. 2004). In our context, content consumption is measured by the total number of tracks and the average daily number of tracks the user listens to. Hence, our third hypothesis is:

H3: Consumers who listen to more music are more likely to pay for a premium subscription.

The literature on levels of participation in online communities describes content consumption, content organization, community activity and community leadership as a hierarchy of levels of engagement in a website. Our data also enable us to compare the effects of different levels of website engagement on willingness to pay. Hence, our fourth hypothesis further addresses the different effects that different types of activities have on willingness to pay.

H4(a): Content organization will have a stronger association with the subscription decision than will content consumption.

H4(b): Community participation will have a stronger association with the subscription decision than will content organization and content consumption.

Moreover, the strongest form of commitment, normative commitment, reflects a sense of obligation toward the website, and it is associated with leadership roles in the community. Our data, which include information both on group membership and on group leadership, provide us with a unique opportunity to study the difference between mere community participation and taking a leading role. We can therefore test the following hypothesis:

H4(c): Leadership of groups will have a stronger association with the subscription decision than will participation in groups.

Prior literature on social influence provides some additional explanations for purchase behavior that should be incorporated into our analysis. Service adoption decisions of consumers may be influenced by the actions of their peers (Choi et al. 2010). This may be due to social contagion (Susarla et al. 2011), the informative nature of word-of-mouth communication (Brown and Reingen 1987; Godes and Mayzlin 2004), observational learning (Wang and Xie 2011; Zhang 2010), herding (Huang and Chen 2006) and opinion leadership (Venkatraman 1990). Our fifth hypothesis will test the existence of social influence in our context:

H5(a): Consumers who have friends listed on their personal pages are more likely to pay for premium subscriptions.

H5(b): Consumers who have friends who are paying subscribers are more likely to pay for premium subscriptions themselves.

Finally, demographics may influence consumers' willingness to pay for premium services. In our context we only obtain information about age and gender. We expect that younger users will have less access to payment methods because of lower incomes and age restrictions for credit card ownership in different countries. Furthermore, following Venkatesh and colleagues (2000), we expect that males and females will differ in their adoption of technological services and willingness to pay for them. We therefore control for demographics in our analysis.

DATA ANALYSIS

Data Collection and Preparation

As in most social networks, each user on Last.fm has his or her own webpage (see Figure 1 for illustration)[7]. We collected the following data on Last.fm users:

[7] See Sinkkonen et al. (2007) for an analysis of Last.fm's music social network topology.

- Demographic information: age, gender, and time since registration to the website.
- Music consumption information: the number of tracks listened to and the time since last visit.
- Content organization activities: the number of songs tagged, the number of songs marked as ‘loved’, and the number of playlists (a list of songs to be played together) created.
- Community participation activities: the number of group memberships (groups on Last.fm are formed around genre, artists and other topics), the number of groups led, the number of blog entries, and the number of posts to forums.
- Number of friends listed on the user's page and the number of friends who are paying subscribers.

We collected these data using two specially programmed web crawlers. One web crawler gathered information about a random sample of 150,000 Last.fm users (subscribers and non-paying users). For this dataset, we omitted data on subscribers and used only data on non-paying users. A second web crawler collected information about new paying subscribers at the time that they purchased their subscriptions. We were able to identify these users because Last.fm features a list of recent subscribers, which is continually updated. By limiting our analysis to new subscribers and omitting members with previously established subscriptions, we control for increased activity that might result from the membership benefits of the premium subscription. Thus far we have collected information on close to 5,000 new subscribers.

Data collection was done over a period spanning 3 months starting in January 2009. In order to omit inactive users from our analysis, we removed data on users who had not visited the site during the 3 months prior to data collection. We also omitted users and subscribers who had in the past used a ‘Reset’ option that reset the logs of their personal site usage. Our final dataset consisted of 39,397 non-paying users and 3,612 new subscribers. Some descriptive statistics for our data are presented in Table 2.

(Insert 'Table 2' here)

The descriptive statistics clearly suggest that the usage pattern of subscribers is quite different from that of regular users. Table 3 and Figure 2 summarize the average activity levels of the consumers in our sample, who are categorized as either (paying) subscribers or (non-paying) users. For each type of activity, the third column of Table 3 shows the ratio between subscriber activity level and user activity level. We used the *t*-test and the Mann-Whitney U-test to compare non-paying users with subscribers, as the two populations are not normally distributed (Mann and Whitney 1947).

(Insert 'Table 3 and Figure 2' here)

We observe that subscribers consume 23% more music than do their non-paying peers. Interestingly, subscribers carry out a significantly larger number of content–organization activities. On average, subscribers create 67% more playlists, they choose to mark 218% more tracks as ‘loved’, and they create 140% more tags ($P < 0.01$).

Most intriguingly, subscribers are substantially more involved in the site’s community: compared with nonpaying users, paying subscribers write 199% more posts on the site’s forums, join 70% more groups, lead on average 142% more groups, and publish 111% more blog entries ($P < 0.01$).

Moreover, paying subscribers have more friends listed on their pages. Table 4 shows that whereas the average non-paying user has slightly more than 14 friends, the average subscriber has 21 friends, i.e., subscribers have on average 45% more friends ($P < 0.01$). As expected, paying subscribers have many more friends who are subscribers than non-paying users do; the average subscriber has 2.82 subscriber friends, compared to only 0.42 subscriber friends for the average non-paying user ($P < 0.01$)[8].

There are also demographic differences between subscribers and non-paying users. We did not observe a significant difference in activity levels or in propensity to subscribe based on gender. We did, however, find that subscribers are on average 6 years older than non-paying users (see Table 2). Given the relatively small subscription fee of $3 per month, we think it is likely that this difference is caused by differences in income level or accessibility to payment methods. Interestingly, we also find that subscribers make their subscription decisions after using the site for 652 days on average. This suggests that the typical subscription decision is not spontaneous. Rather, it requires deep familiarity with the website and its features. This indicates that converting users from free to fee is a long process that requires patience from website owners.

Moreover, we find that 99.1% of all users (paying and non-paying) have listened to music, 77.6% have engaged in content organization behavior, 57.9% have participated in the community and 5.2% have led a group, taking leadership role in the community. Interestingly, only 8.7% of the users who have engaged in a community activity has not used the content organization features of the website. This supports the notion of a hierarchy of activities.

Methodology and Results

To better understand the interplay of content consumption, content organization, community activity, and willingness to pay for a subscription, we estimated a logistic

[8] As we collect the data at the moment of subscription, we can know that the friends paid before the focal user did.

(binary) choice equation, predicting the probability of paying for a subscription[9]. Formally, we estimated the following block equation:

$$U_i(Subscribe) = \alpha_0 + \alpha_1 ContentConsumption_i + \sum_{j=1}^{J} \beta_{ij} ContentOrganization_i + \alpha_2 FriendsCount_i + \alpha_3 SubscriberFriendsCount_i + \sum_{k=1}^{K} \gamma_{ik} CommunityParticipation_i + \alpha_4 CommunityLeadership + \sum_{l=1}^{L} \delta_{il} Demographics_i + \varepsilon_i = V_i + \varepsilon_i$$

Content consumption is estimated using the total number of tracks (in thousands) that user *i* listened to, as well as the average daily number of tracks that user *i* listened to. The content organization activities include tagging of songs, creating playlists, and marking songs as ‘loved’. $FriendsCount$ is the number of friends listed on a user's personal page, and $SubscriberFriendsCount$ is the number of friends listed on the user's personal page who became subscribers prior to the focal user's decision. The community participation activities include joining groups, leading groups, posting in a forum, and adding an entry to a personal blog. Demographics include age, gender, and the number of days since the user started using the website. The error terms ε_i are assumed to follow an extreme value distribution (i.e., we use the logit model). Thus, the conditional probability, Pr_i, that consumer *i* chooses to pay for a premium subscription is given by the usual expression

$$Pr_i = \frac{\exp(V_i)}{1 + \exp(V_i)}$$

[9] Since premium services are offered for a fixed monthly fee, we use a logistic regression model with a binary dependent variable.

Estimating this model presented us with two econometric challenges:

First, we needed a control for increased use of the site due to the actual subscription decision. It is possible that after subscribing to premium services, consumers tend to use the site more because of the benefits a subscription provides. For that reason, we limited our analysis to non-paying users and to new subscribers whose data had been collected immediately following the time of subscription, that is, before their usage could be influenced by the subscription itself. We therefore merged two sets of data: one consisting of randomly chosen non-paying users, and one consisting of users who had just purchased a subscription.

Second, when we looked at the random set of users on whom we collected information, we noticed that subscribers made up only 0.89% of the site population. If we used this correct ratio in composing our dataset, the occurrence of ones in our dependent variable (*Subscribe*) would be a *rare event*. The biases that rare events create in estimating logit models have been discussed in the literature (Ben-Akiva and Lerman 1985). Briefly, this poses a problem when estimating a logit model, because the model would predict that everyone would be a regular, non-subscribing user while still obtaining a 99% level of accuracy. To overcome the problem of misclassification, one should re-estimate the model while deliberately under-sampling the non-paying users, so that a more balanced sample of ones and zeros in the dependent variable is obtained. This sampling technique is called *choice-based sampling* (Ben-Akiva and Lerman 1985). To this end, we used our collected set of 3,437 new subscribers and only 9,537 non-paying users. However, using choice-based sampling leads to inconsistent intercept estimation when traditional maximum likelihood methods are used. Two alternative solutions have been suggested in the literature: Manski and Lerman (1977) developed a weighted endogenous sampling maximum likelihood (WESML) estimator, which accounts for the different weights in the zeros and ones from the population of interest. However, this estimator has the undesirable property of increasing the standard errors of the estimates (Greene 2000; Manski and Lerman 1977). A second approach, which we follow, is to adjust the estimated intercepts for each alternative by subtracting the constant $ln(S_i/P_i)$ from the exogenous maximum likelihood estimates of the intercept, where S_i is the percentage of observations for alternative *i* in the sample, and P_i is the percentage of observations for alternative *i* in the population (Manski and Lerman 1977; see Villanueva et al. 2008 for a similar implementation).

The correlation matrix is presented in Table 4, and the estimation results using the choice-based sample are reported in Table 5, each column representing an additional block being added to the estimation.

(Insert Tables 4 & 5 here)

Estimation Results

The number of different community activities, the number of content organization activities, and the level of content consumption are strongly and significantly associated with the likelihood of subscription.

Community Participation: Joining a group, leading a group and posting a blog entry are each associated with a significant increase in the odds of subscribing to premium services, supporting *H1*, (*Odds Ratio* = 1.007 for each group membership, *Odds Ratio* = 1.226 for each group leadership, and *Odds Ratio* = 1.051 for each blog entry). Note that posting a comment in a forum does not have a significant association with the subscription decision.

Community Leadership: Group leadership has a much stronger association with the subscription decision than group membership does. Specifically, our results suggest that being a leader of one more group has a stronger effect on the odds ratio than being a member of 30 additional groups. Hence, *H4(c)* is clearly supported.

Content Organization: We also find that content-organization activities, including marking tracks as 'loved' and creating playlists, are positively correlated with subscription behavior, partially supporting *H2*. Creating tags for songs was not found to be statistically significant in the full model. While tagging songs as 'loved' has a weaker association with the subscription decision compared with participation in community activities, creating a playlist has a very strong effect on the odds ratio. Hence, *H4(b)* is only partially supported.

Content Consumption: As expected, content consumption has a positive association with the subscription decision, supporting *H3*. Interestingly, content consumption is associated with a relatively low effect on the subscription decision and is not significant in all models. Looking at our full model, it seems that the effect of posting an additional entry to a blog is equal to that of listening to over 10,000 more tracks. Similarly, being a member in one more group has a stronger effect on the odds ratio than listening to 100,000 more tracks. These findings support *H4(a)* and suggest that willingness to pay is more strongly linked to community activity and to content organization activities than to content consumption. These results are especially interesting given that the core business of the website is providing content, and that most of the features provided to the paying subscribers are closely related to the content-consumption experience.

Social Influence: We also find that the number of subscriber friends (i.e., friends who have already purchased a paid subscription) listed on a user's page is associated with a strong positive effect on the user's propensity to pay for premium services (supporting *H5(b))*. When we control for the number of subscriber friends, we find that the number of friends without a subscription has a small negative association with the subscription behavior. This could indicate that non-subscribing friends create negative word of mouth regarding the subscription decision, either verbally or through observational learning.

Demographics: The age of the user is positively associated with the likelihood of subscription, but gender has no significant effect. More interestingly, the number of days since the user started using the website is found to be negatively associated with the subscription decision.

The Effect of Community Participation on Time until Subscription

We find that subscribers make their subscription decisions after using the site for 652 days on average. This suggests that the typical subscription decision is made by a user who is deeply familiar with the website and its features. In what follows, we study the factors that influence the time between joining the website and the decision to subscribe. Therefore, the event of interest is the conversion from free to fee. Specifically, we are interested in the effect of content consumption, content organization and community activity on the hazard of consumers to convert. We therefore estimate a hazard (survival) model, using the following equation:

$$H_i(t) = \exp\left\{\alpha_0 + \alpha_1 ContentConsumption_i + \sum_{j=1}^{J} \beta_{ij} ContentOrganization_i \right.$$
$$+ \alpha_2 FriendsCount_i + \alpha_3 SubscriberFriendsCount_i$$
$$+ \sum_{k=1}^{K} \gamma_{ik} CommunityParticipation_i + \alpha_4 CommunityLeadership$$
$$\left. + \sum_{l=1}^{L} \delta_{il} Demographics_i \right\}$$

This model allows us to study how the different covariates are associated with the 'hazard' (in this case, a positive hazard in the form of a subscription decision). We use the Cox regression to estimate these effects. The results of this estimation are presented in Table 6.

(Insert 'Table 6' about here)

The results show that community activity and content organization activity variables are each positively associated with the hazard rate. That is, users who are more active in the community or who actively organize content will make the subscription decision sooner than users who are less active (or not active at all). Moreover, we again see a significant positive association between group leadership and the subscription decision. Similarly, users with more subscriber friends have a higher hazard rate. These results provide yet another dimension to our previously reported results: not only is community activity associated with a greater willingness to pay for a premium subscription, it is also associated with a shorter time window between joining the website and subscribing.

Propensity Score Matching

Although the preceding econometric analysis provides support for a positive and statistically significant association between online community activity and propensity to purchase a premium-service subscription, the nature of observational data raises concerns about the causal interpretation of our findings. As mentioned above, through our sampling technique, we control for possible post-subscription increases in site usage. However, we do not control for the bias caused by self-selection. That is, since we did not randomly assign users to "treatment" groups (increased community activity), we are unable to control for observed and unobserved variables that drive users to self-select themselves into a particular treatment group. It is easy to think of variables that might influence users' community activity levels and simultaneously increase their propensity to pay for premium services, hence creating a self-selection bias.

A solution to the self-selection bias is to use a *proportional outcome approach*. Selection bias due to correlation between the observed characteristics of a user and the user's level of social activity (his "treatment" level) can be addressed by using a matching technique based on propensity scores (Rosenbaum and Rubin 1983; for a recent use of propensity scores in the marketing context, see Aral et al. 2009; Mithas and Krishnan 2009). The fundamental problem in identifying treatment effects is one of incomplete information. Though we observe whether the treatment occurs and whether the outcome is conditional on the treatment assignment, the counterfactual is not observed. In a nutshell, propensity matching techniques enable us to investigate heterogeneous treatment effects in non-experimental data, based on observed variables[10]. The objective of propensity score matching is to assess the effect of a treatment by comparing observable outcomes (in our case, subscription behavior) among treated observations (in our context, users who participate in the website's community) to a sample of untreated observations (in our context, users who did not participate in the website's community) matched according to the propensity of being treated (that is, the propensity to participate).

Mathematically, let $y_{i,1}$ denote the outcome of observation *i*, if the treatment occurs (given by $T_i=1$), and $y_{i,0}$ denote the outcome if the treatment does not occur ($T_i=0$). If both states of the world were observed, the average treatment effect, τ, would equal $y_1 - y_0$, where y_1 and y_0 represent the mean outcomes for the treatment group and control group, respectively. However, given that only y_1 or y_0 is observed for each observation, unless assignment into the treatment group is random, generally, $\tau \neq y_1 - y_0$.

Propensity score matching attempts to overcome this problem by finding a vector of covariance, *Z*, such that $(y_1, y_0) \perp T|Z,\ pr(T=1|Z) \in (0,1),$ where $\perp$ denotes independence. That is, the treatment assignment is independent of the outcome conditional on a set of

[10] In contrast, selection bias stemming from correlation between unobserved variables and the user's social activity level is a more difficult problem. Previous literature has often used the strong ignitability assumption (Rosenbaum and Rubin 1983).

attributes *Z*. Moreover, if one is interested in estimating the average treatment effect, only the weaker condition, $E\left[y_0 \middle| T=1, Z\right] = E[y_0|T=0,Z] = EE[y_0|Z],\ \ pr(T=1|Z) \in (0,1),$ is required.

To implement the matching technique, we define the "treatment" group as the set of people who participated in community activity. Since most propensity score matching techniques use a binary treatment, we grouped user participation in community activities into <u>four distinct binary treatments</u> and repeated the following exercise for each treatment separately:

- *GroupLead*, which is equal to one if the user has ever led a group;
- *BlogEntry,* which is equal to one if the user has ever posted an entry to a blog;
- *GroupMember*, which is equal to one if the user has ever joined a group;
- *ForumPost,* which is equal to one if the user has ever posted an entry to a forum page.

Additionally, we group all of the user's community activities into one binary variable, *CommunityActivity*, which is equal to one if the user has ever posted an entry to a blog, joined a group, or posted an entry to a forum page.

In our context, we are able to identify a number of observed variables that might influence a consumer's propensity to engage in social activity and should therefore be included in the covariates in *Z*. We estimate the propensity to participate or contribute to the community based on demographic information (including gender and age), music consumption patterns (including the number of tracks listened to, and the number of days on the Last.fm site), and the number of friends listed on the user's page[11].

[11] For robustness, we repeated the estimations using the other activities as covariates as well. That is, when estimating a person's propensity to perform a certain activity, we included the other activities of the person in the propensity estimations. For example, when estimating the propensity to write a blog entry, we included group membership and posts to forums into the score estimations.

Consequently, we should match observations that have identical values for all variables included in *Z*. For example, in the case of *GroupLead* treatment, we should match a 22-year-old male consumer who listened to 1,000 tracks, had been using Last.fm for a year, and is a group leader, with another 22-year-old male who listened to 1,000 tracks and had been using Last.fm for a year, but who is not a group leader. However, if we do that, we might find very few exact matches. Since exact matching is often untenable, Rosenbaum and Rubin (1983) prove that conditioning on $p(Z)$ is equivalent to conditioning on Z, where $p(Z)=pr(T=1|Z)$ is the propensity score. That is, for each consumer we estimate $p(Z)$—the propensity of being treated (in the previous example, the propensity of leading a group)—using a probit model. We thereafter match consumers not according to their exact attributes but according to their propensity scores. One of the advantages of propensity score methods is that they easily accommodate a large number of control variables.

Upon estimation of the propensity score, a matching algorithm is defined in order to match the treated and untreated cases. We used the kernel matching estimator matching technique (Heckman 1997)[12]. We were then able to compare the percentage of subscribers between the treated and the matched untreated groups. For the *CommunityActivity* variable we repeated the estimations using the Mahalanobis matching technique, a method specifically designed for multiple treatments (Rubin, 1980). Using this method, one estimates a different propensity score for each treatment included in the *CommunityActivity* variable (i.e., posting to a forum, group membership, and blog entry), and users are then matched on the basis of these multiple scores.

[12] We chose the kernel matching technique because of its treatment of the "distance" between the matched and unmatched cases through weights. Kernel matching gives more weight to close neighbors while still assigning some weight to the more distant neighbors. The potential benefit is that these estimators are less sensitive to a mismatch along unmeasured dimensions, but the cost is that they introduce an added mismatch along measured dimensions. For robustness, we repeated the analysis using the nearest neighbor matching algorithm, with very similar results.

The results of our comparisons for each of the treatments are presented in Table 7. Column A in Table 7 corresponds to the case in which the treatment is defined as *GroupMember*. In this case each consumer who has a group membership is matched with a consumer who does not have a group membership, according to the above-mentioned covariates (including demographics, music listening, and friends). Out of the 29,941 consumers with group memberships, 8.5% were found to have a subscription. However, out of the 29,941 consumers who were matched to those consumers (but were not group members) only 6.9% had a subscription. Since this difference is statistically significant ($P < 0.001$), we are able to conclude that, controlling for the observed differences between the groups, consumers who are group members are more likely to pay for a premium subscription. Similar analysis for the other four treatments (group leadership, forum posting, blog entries and any community activity) is presented in columns B to F of Table 7. Note that *CommunityActivity* was estimated twice, once using the kernel matching approach (column E) and once using the Mahalanobis matching approach (Column F). All these estimates provide similar conclusions: After controlling for self-selection bias based on demographics, music consumption, and number of friends, we observe a significant difference between the treated and untreated conditions in the mean percentage of users who subscribe to premium services.

These differences emphasize the effect of community participation on the propensity to subscribe to the website and strengthen the findings of the binary logistic model. (A comparison of covariate means both before and after the matching are presented in appendix 1.)

(Insert 'Table 7' about here)

DISCUSSION

The unique characteristics of digital products have made the two-tiered business model a widespread choice for online content providers. However, the profitability of this model depends strongly on firms' ability to convert non-paying users into paying subscribers.

Concurrently, with the success of social networking sites and the rise of user-generated culture, many businesses have begun to incorporate social platforms into their websites, with the intent of leveraging such features to create and retain new value. Such community features may provide firms with a variety of indirect benefits, such as the ability to listen to consumers' conversations, energize word-of-mouth communication, learn how consumers interact with their products and exploit intelligence to provide personalized services (Park and El-Sawy 2008). Until now, however, the use of IT-enabled social features has not been shown to directly impact firms’ profitability. Moreover, it has been shown that consumers are not willing to pay directly for the use of these community features. (Nielsen 2010).

We propose that a number of findings from this research can provide new insights regarding the ongoing debate on the optimal business model for online content consumption and social computing integration in a firm's digital strategy.

First, we suggest that there is another type of benefit that an online community may provide a firm. A firm that creates a community on its website offers its consumers a platform through which they can become increasingly engaged with and committed to the site. In our empirical analysis, we find that users who are more active in the community are substantially more likely to pay for premium services, and this effect is observed even after accounting for content consumption, demographics, and social influence. We also find that, in the context of music content, community activity is more strongly associated with the likelihood of subscription than is the music consumption itself.

Among all the social attributes we examined, the number of subscriber friends, the number of playlists, the number of groups led, and the number of blog entries are most strongly associated with the purchase decision. The first two observations are not surprising in our context. Past research has already shown how social interactions in online environments can influence purchasing decisions (Godes and Mayzlin 2004; Huang and Chen 2006). The effect of playlist creation, in turn, might be a fairly obvious outcome of the extended playlists option that a premium subscription provides in the website we study. However, none of the premium services directly improves the user's ability to lead groups or to post to blogs. In fact, most of the benefits associated with a subscription—including higher bandwidth, access to new music features, and removal of ads from the user's page—are not directly related to the community aspects of the website.

Our findings support the notion of a hierarchy, portrayed in the literature on levels of participation in online communities. According to this hierarchy, group leadership and blog postings are at the top end of user participation behavior, whereas acts of content organization and consumption reflect lower levels of participation. The group leader is in charge of moderating the group's discussions and adding new members to its community. The active blogger creates his own space and frequently shares his written thoughts with the entire Last.fm community. An explanation for the correlation between these activities and the purchase decision can stem from the connection between these activities and levels of commitment. While consuming content reflects a continuance commitment based on cost–benefit analysis, engagement created by social computing might increase affective and normative commitments. Highly involved users might feel the need to pay tribute to a website they enjoy and want to actively support. They like the website and, in some cases, feel a sense of obligation towards it as they take on more responsibility and participate intensively in the community. Among such users, the presence of the community might introduce a sense of social reciprocity, associated with monetary payments to the website. While analysts have noted that people report that they are not willing to pay for online content (Nielsen 2010), involvement in a community on a content website might serve as a key to overcoming that obstacle.

We extend our results in two directions. First, we use a hazard model to study the effect of community activity on the time between joining the website and the subscription decision. We find that users who are more active in the community will make the subscription decision sooner after joining compared with users who are less active (or not active at all). Moreover, we again see the strong association between group leadership and the subscription decision. These results suggest that a consumer's community activity is associated not only with increased willingness to pay for a premium subscription but also with a shorter time window between joining the website and subscribing. This indicates that community participation can act as a catalyst for purchasing decisions in online content websites.

Second, we extend our results by using propensity score matching, a method of estimating treatment effects from non-experimental data (Aral et al. 2009; Mithas and Krishnan 2009). Previous research on willingness to pay has used surveys or interviews in order to assess purchasing intent (Riggins 2003; Srinivasan et al. 2002; Ye et al., 2004). By using a dataset of users who are currently active on the content website, we were able to study actual purchasing decisions without the biases commonly associated with surveys. The featured list of recent subscribers, updated in real time, allowed us to avoid the influences of post-subscription behavior and to properly compare a subscriber's profile to that of a non-paying consumer.

Although we did not control for unobserved heterogeneity in treatment assignment, propensity score matching allowed us to control for self-selection bias based on consumption patterns, demographics, and social influence levels. We show that the contribution of content to the community increases contributors' willingness to pay for premium services. This provides the first evidence as to the causal effect of community activity on consumers' willingness to pay.

Managerial Implications

This research suggests that future fee-paying subscribers of a content website are not necessarily the most avid content consumers but instead may be the most avid participants in the website's online community. This finding implies the importance of community-building in a content website. However, managers should consider their options carefully when attempting to plan an effective community that could impact revenue streams. This research supports the notion that a website that offers only content and does not support community activity is not enough to engage consumers and motivate them to pay for subscriptions.

As content-organization activities were not found, in most cases, to have a great impact on the subscription decision, we propose that it is also not enough to give users the option to rate or tag content in the hope they will become committed subscribers. Similarly, friend-making was not found to substantially influence subscription decisions, unless one's friends are paying members.

Taken together, these results highlight the importance of creating a community environment that facilitates different levels of participation. Two of the activities that were most strongly linked to the subscription decision—blog creation and moderation of content (by group leadership)—are of a high-participation nature and are likely to occur in advanced stages of community membership. By offering a variety of social features, a website can create the full ‘ladder’ of participation and encourage users to advance towards this high level of involvement, potentially increasing the chances that they will subscribe. A website owner should make such features available and easy to use, while making sure users are aware of their existence. Our research suggests that content providers should not ask themselves "How will I make my users pay?" but rather "How will I make my users more engaged?" The solution to this question may increase free-to-fee conversion rates.

Researchers as well as practitioners have noted that many users of content websites ignore community features and stay at the first level of participation (i.e., ‘lurking’), whereas only a few make their way to the highest level of participation (Li and Bernoff 2008). Hence, merely offering a community might not be enough; websites may need to actively help users move on to the next level of participation. Previous research has indicated that consumers move up the ladder starting at activities that require low levels of participation, such as content organization activities. Therefore, it could be wise to not immediately invite consumers to participate in activities requiring high levels of participation (such as group leadership), but rather offer incremental changes in the levels of participation. This can be done in different ways. One approach is to suggest a consequent activity of a higher level upon completion of an activity. For example, a user who consumes content might be asked to tag it, a user who tags content might subsequently be asked to also review it in a forum, a user who is active in discussions might be asked to lead the forum, and so on. This might help increase the percentage of users who reach high levels of participation.

Another question of interest in this context is whether the two-tiered model is effective, given the low subscription fee and the possibility that subscriptions might detract from ad revenue: As noted above, one of the benefits of a premium subscription on Last.fm is the removal of ads from one’s personal page. Last.fm, like most firms, does not disclose exactly how many paying subscribers it has or how much revenue it receives from advertisements. However, in our data set, which included 150,000 randomly chosen users of Last.fm, there were 1,335 paying subscribers. This implies a conversion rate of about 0.9%. This number is in line with numbers reported by other websites, whose conversion rates are between 0.5% to 15%, but are often on the low side (Anderson 2009). Given that Last.fm has about 30 million registered users and the monthly subscription fee is $3, we estimate that the revenue from premium subscriptions is about $9.6 million a year. Since these are all digital services, with low marginal costs, the profit margins on this amount are estimated to be very high. Hence, even with a low conversion rate and a relatively low monthly subscription fee, subscriptions are a substantial source of income for the website. Moreover, given the vast number of registered users, even a small change in users' propensity to subscribe will result in a substantial increase in profit. For example, a 10% increase in the conversion rate, from 0.9% to 1%, will result in an additional $1,188,000 per year.

While there are no official reports on the profitability of advertising business models, the convention is that the advertising conversion rates on search engines such as Google are about 2%[13], whereas the conversion rates reported by social networks (such as Facebook) are about 0.051–0.063%[14]. The reported average payoff of a click-through on a Google ad is 5 cents. Of course, this conversion rate is with regard to page views. A simple calculation therefore shows that for an average click-through rate of about 0.05%, a $3 monthly fee is equivalent to about 120,000 page views a month. While this is a very rough estimate, it is clear to see that a paying member generates much more profit than a nonpaying member who is exposed to ads. Therefore, given the challenges of the advertising business model, a careful discussion of new strategic means by which firms can increase, even by a small fraction of a

[13] As reported on the Google Help page for AdWords (http://www.google.com/support/forum/p/AdWords/thread?tid=7aeb3290fd8feccb&hl=en)

[14] WebTrends Report (http://f.cl.ly/items/2m1y0K2A062x0e2k442l/facebook-advertising-performance.pdf)

percentage, users' willingness to pay for premium subscriptions is of great importance to this industry.

It is important to note that the strategy of promoting community participation is likely to work best in content sites that achieve high readership, such as successful mainstream news or music websites that cater to a variety of users. This is true for two reasons. The first is that such sites have substantial numbers of users who start at the first stage of participation. Even if just a small percentage of these users progress to become highly engaged and eventually contribute payments to the site, they might still constitute a large population that can benefit the site’s overall income. Second, websites that implement social computing features are also prone to network externalities, and thus, a consumer’s value is greatly affected by fellow consumers' behavior. A site with high readership in which some users progress to content organization and contribution can affect other people's experience of the website, their satisfaction, and ultimately their retention. For similar reasons, websites that begin with a small number of content readers might have problems implementing such a model, as only a few users will eventually pay, and the cost of community building may be unsustainable. Such websites might prefer to use the services of existing social media companies, for example, by building a fan page on Facebook or on Twitter.

Limitations and Future Work

This research was carried out on the Last.fm website, which allowed exploration of different social computing features. Last.fm is a leading music-providing website and also has a relatively active community, in which a variety of social features are offered to the users, making it a fruitful source of data for research of this type. Nevertheless, future research should investigate other websites, providing different types of content such as news or video. Furthermore, Last.fm is an intermediary and not a content creator. Content creators, such as *The New York Times*, deliver original content. As there are no perfect substitutes for original, unique content, some may argue that consumers' willingness to pay for such content will be higher, and therefore that content creators may not need to add community features to their websites. However, Last.fm has a unique (patented) music recommendation system that creates a unique experience for the user. Furthermore, original content creators face similarly low willingness to pay, which in turn creates financial difficulties (Nielsen 2010). Investment in social computing features may therefore be beneficial for those websites as well.

We focus on proprietary content websites. While it is possible that our findings can be extended to websites that offer user-generated content as well, we have no data on such websites. This would be another interesting direction for future work.

Moreover, we used real-world data, in which the subscription package offered to Last.fm users included one set of premium services. It is impossible to know which premium service, if any, appealed most to the new subscribers. Future research should consider a controlled experimental setting, where different bundling packages can be explored. Such research should aim to unbundle the service packages and link the willingness to pay for different services to different community activities.

As in other, similar empirical studies, it is impossible to account for the unobserved consumer characteristics that might influence the subscription decision. In this case, our rich data set has allowed us to control for different behaviors and attributes observed online. We have also implemented a propensity score matching technique to further control for observable variables. Nevertheless, there are still correlated unobservables that should be handled in future work, perhaps using an experimental setting. Specifically, richer data about the local (person-to-person) social activity of consumers might provide interesting insights into the extent and nature of peer influence on the subscription decision. Finally, our research focuses on consumers' usage levels in the period prior to the subscription decision. An extension of the research to post-purchasing behavior, e.g., through the use of panel data, could have provided additional support to our findings. We encourage fellow researchers to further investigate how new social possibilities can be incorporated into digital business strategies.

REFERENCES

Agichtein, E., Castillo, C., Donato, D., Gionis, A., and Mishne, G. 2008. "Finding High-Quality Content in Social Media," in *WDSM'08 Proceedings of the International Conference on Web Search and Web Data Mining*, NewYork: ACM.

Algesheimer, R., Borle, S., Dholakia, U. M., and Singh, S. 2009. "The Impact of Customer Community Participation on Customer Behavior: Am Empirical Investigation," working paper.

Anderson, C. 2009. *Free: The Future of a Radical Price*. New York: Hyperion.

Aral, S., Muchnick, L., and Sundararajan, A. 2009. "Distinguishing Influence-based Contagion from Homophily-driven Diffusion in Dynamic Networks," *Proceedings of the National Academy of Sciences of the United States of America* (106:51), pp. 21544–21549.

Asvanund, A., Clay, K., Krishnan, R., and Smith, M. D. 2004. "An Empirical Analysis of Network Externalities in Peer-To-Peer Music Sharing Networks," *Information Systems Research* (15:2), pp. 155–174.

Bagozzi, R. P., and Dholakia, U. M. 2006. "Antecedents and Purchase Consequences of Customer Participation in Small Group Brand Communities," *International Journal of Research in Marketing* (23:1), pp. 45–61.

Bateman, P. J., Gray, P. H., and Butler, B. S. 2010. "The Impact of Community Commitment on Participation in Online Communities," *Information Systems Research*, Published online in Articles in Advance, January 27, DOI: 10.1287/isre.1090.0265.

Bawa, K., and Shoemaker, R. 2004. "The Effects of Free Sample Promotions on Incremental Brand Sales," *Marketing Science* (23:3), pp. 345–363.

Ben-Akiva, M., and Lerman, S. R. 1985. *Discrete Choice Models*, Cambridge, MA: MIT Press.

Bhattacharjee, S., Gopal, R. D., Lertwachara, K., Marsden, J. R., and Telang, R. 2007. “The Effect of Digital Sharing Technologies on Music Markets: A Survival Analysis of Albums on Ranking Charts,” *Management Science* (53:9), pp. 1359–1374.

Bhattacharjee, S., Gopal, R. D., and Sanders, G. L. 2003. “Digital Music And Online Sharing: Software Piracy 2.0?” *Communications of the ACM* (46:7), pp. 107–111.

Brown, J., and Reingen, P. H. 1987. “Social Ties and Word-of-Mouth Referral Behavior,” *Journal of Consumer Research* (14), pp. 350–362.

Brynjolfsson, E., Hu, Y., and Smith, M. D. 2003. “Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers,” *Management Science* (49:11), pp. 1580–1596.

Burke, M., Marlow, C., and Lento, T. 2009. “Feed Me: Motivating Newcomer Contribution in Social Network Sites,” ACM CHI 2009: Conference on Human Factors in Computing Systems, Boston, MA.

Butler, B. S. 2001. "Membership Size, Communication Activity, and Sustainability: A Resource-Based Model of Online Social Structures," *Information Systems Research* (12:4), pp. 346–362.

Cassell, J., Huffaker, D., Tversky, D., and Ferriman, K. 2006. “The Language of Online Leadership: Gender and Young Engagement on the Internet,” *Developmental Psychology* (22), pp. 436-449.

Chellappa, R. K., and Shivendu, S. 2005. “Managing Piracy: Pricing and Sampling Strategies for Digital Experience Goods in Vertically Segmented Markets,” *Information Systems Research* (16:4), p. 400.

Chen, Y., and Seshadri, S. 2007. “Product Development and Pricing Strategy for Information Goods under Heterogeneous Outside Opportunities,” *Information Systems Research* (18:2), pp. 150–172.

Choi, J., Lee, S. M., and Soriano, D. R. 2009. “An Empirical Study of User Acceptance of Fee-Based Online Content,” *Journal of Computer Information Systems* (49:3), 60–70.

Chyi, H. I. 2005. “Willingness to Pay for Online News: An Empirical Study on the Viability of the Subscription Model,” *Journal of Media Economics* (18:2), pp. 131–142.

Clemons, E. K. 2009. “The Complex Problem of Monetizing Virtual Electronic Social Networks,” *Decision Support Systems* (48:1), pp. 46–56.

Cook, J., and Wall, T. 1980. "New Work Attitude Measures of Trust, Organizational Commitment and Personal Need Non- Fulfillment," *Journal of Occupational Psychology* (53:1), pp. 39–52.

Cummings, J., Butler, B., and Kraut, R. 2002. "The Quality of Online Social Relationships," *Communications of the ACM* (45:7), pp. 103-108.

Doerr, J., Benlian, A., Vetter, J., and Hess, T. 2010. “Pricing of Content Services–An Empirical Investigation of Music as a Service,” *Sustainable e-Business Management* (58), pp. 13–24.

Dyson, E. E. 1995. “Intellectual Value,” *Wired Magazine*, (July), (http://www.wired.com/wired/archive/3.07/dyson.html).

Eisenberger, R., Fasolo, P., and Davis-LaMastro, V. 1990. "Perceived Organizational Support and Employee Diligence, Commitment, and Innovation," *Journal of Applied Psychology* (75), pp. 51-59.

Fitzsimons, G. J., and Lehmann, D. R. 2004. “Reactance to Recommendations: When Unsolicited Advice Yields Contrary Responses,” *Marketing Science* (23:1), pp.1–16.

Forte, A., and Bruckman, A. (2008) “Why Do People Write for Wikipedia? Incentives to Contribute to Open-Content Publishing,” in *Proceedings of 41st Annual Hawaii International Conference on System Sciences (HICSS)*.

Füller, J., Matzler, K., and Hoppe, M. 2008. “Brand Community Members as a Source of Innovation,” *Journal of Product Innovation Management* (25:6), pp. 608–619.

Gedenk, K., and Neslin, S. A. 1999. “The Role of Retail Promotion in Determining Future Brand Loyalty: Its Effect on Purchase Event Feedback,” *Journal of Retailing* (75:4), pp. 433–459.

Godes, D., and Mayzlin, D. 2004. “Using Online Conversations to Study Word of Mouth Communication,” *Marketing Science* (23:4), pp. 545–560.

Godes, D., Ofek, E., and Sarvary, M. 2009. “Content vs. Advertising: The Impact of Competition on Media Firm Strategy,” *Marketing Science* (28:1), pp. 20–35.

Gopal, R. D., Sanders, G. L., Bhattacharjee, S., Agrawal, M. K., and Wagner, S. C. 2004. “A Behavioral Model of Digital Music Piracy, Journal of Organizational Computing and Electronic Commerce,” (14:2), pp. 89–105

Greene, W.H. 2000. *Econometric Analysis*. 4th ed., Upper Saddle River, NJ: Prentice Hall.

Heckman, J. J. 1997. “Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations,” *Journal of Human Resources* (32:3), pp. 441–462.

Huang, J., and Chen, Y. 2006. “Herding in Online Product Choice, Psychology and Marketing,” (23:5), pp. 413–428.

Huberman, B. A., Romero, D. M., and Wu, F. 2009. "Crowdsourcing, Attention and Productivity," *Journal of Information Science* (35:6), pp. 758-765.

Hung, J. 2010. "Economic Essentials of Online Publishing with Associated Trends and Patterns," *Publishing Research Quarterly* (26:2), pp. 79–95.

Jain, S. 2008. “Digital Piracy: A Competitive Analysis,” *Marketing Science* (27:4), pp. 610-626.

Jang, H., Olfman, L., Ko, I., Koh, J., and Kim, K. 2008. “The Influence of Online Brand Community Characteristics on Community Commitment and Brand Loyalty,” *International Journal of Electronic Commerce* (12:3), pp. 57–80.

Joyce, E., and Kraut, R. E. 2006. “Predicting Continued Participation in Newsgroups,” *Journal of Computer-Mediated Communication* (11:3), pp. 723–747.

Kavanaugh , A. 2003 . “When Everyone is Wired,” in *The Wired Homestead*, J. Turow and A. Kavanaugh (eds.), Cambridge, MA: MIT Press.

Kim, A. J. 2000. *Community Building on the Web*, Berkeley, CA: Peachpit Press.

Kollock, P. 1999. “The Economies of Online Cooperation: Gifts and Public Goods in Cyberspace,” in *Communities in Cyberspace*, M. A. Smith and P. Kollock (eds.), London: Routledge.

Lakhani, K. R., and von Hippel, E. 2003. "How Open Source Software Works: "Free" User-to-User Assistance," *Research Policy* (32:6), pp. 923–943.

Lampe, C., and Johnston, E. 2005. “Follow the (Slash) Dot: Effects of Feedback on New Members in an Online Community,” *ACM Conference on Supporting Group Work*, New York: ACM Press, pp. 11–20.

Lave, J., and Wenger, E. 1991. *Situated Learning: Legitimate Peripheral Participation*, Cambridge, UK: Cambridge University Press.

Li, C., and Bernoff, J. 2008. *Groundswell: Winning in a World Transformed by Social Technologies*, Boston, MA: Harvard Business Review.

Lopes, A., and Galletta, D. 2006. “Consumer Perceptions and Willingness to Pay for Intrinsically Motivated Online Content,” *Journal of Management Information Systems* (23:2), pp. 203-231.

Ma, M., and Agarwal, R. 2007. “Through a Glass Darkly: Information Technology Design, Identity Verification, and Knowledge Contribution in Online Communities,” *Information Systems Research* (18:1), pp. 42–67.

Mann, H. B., and Whitney, D. R. 1947. “On a Test of whether One of Two Random Variables Is Stochastically Larger than the Other,” *Annals of Mathematical Statistics* (18), pp. 50–60.

Manski, C., and Lerman, L. 1977. “The Estimation of Choice Probabilities from Choice-Based Samples,” *Econometrica* (45:8), pp. 1977–1988.

McAlexander, J. H., Schouten, J. W., and Koenig, H. F. 2002. “Building Brand Community,” *Journal of Marketing* (66:1), pp. 38–54.

Meyer, J. P., and Allen, N. J. 1991."A Three-Component Conceptualization of Organizational Commitment," *Human Resource Management Review* (1:1), pp 61-89.

Mithas, S., and Krishnan, M. S. 2009. “From Association to Causation via a Potential Outcomes Approach,” *Information Systems Research* (20:2), pp. 295-313.

Moe, W. M., and Fader, P. 2004. “Dynamic Conversion Behavior at E-Commerce Sites,” *Management Science* (50:3), pp. 326–335.

Muniz, A. M., and O’Guinn, T. C. 2001. “Brand Community,” *Journal of Consumer Research* (27), pp. 412–432.

Nielsen. 2010. "Changing Models: A Global Perspective for Paying for Content Online," (http://in.nielsen.com/site/documents/PaymentforOnlineContent.pdf).

Nonnecke, B., and Preece, J. 2000. “Lurker Demographics: Counting the Silent,” in *Proceedings of Annual ACM Conference on Human Factors in Computing Systems*, New York: ACM, pp. 73-80.

O'Reilly, T. 2005. "What is Web 2.0 – Design patterns and Business Models for the Next Generation of Software," (http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html).

Parameswaran, M., and Whinston, A. B. 2007. "Research Issues in Social Computing," *Journal of the Association for Information Systems* (8:6), pp. 336-350.

Park, Y., and El Sawy, O. A. 2008. "Towards a Design Theory for Hedonic Systems: Delivering Superior User Experience in the Digital Home Entertainment Context," in *Proceedings of the Third International Conference on Design Science Research in Information Systems and Technology*, V. Vaishnavi and R. Baskerville (eds.), Atlanta, GA: Georgia State University.

Pauwels, K., and Weiss, A. 2008. "Moving from Free to Fee: How Online Firms Market to Change Their Business Model Successfully," *Journal of Marketing* (72:3), pp. 14-31.

Picard, R. G. 2000. "Changing Business Models of Online Content Services—Their Implications for Multimedia and Other Content Producers," *International Journal on Media Management* (2:2), pp. 60–68.

Preece, J., and Schneiderman, B. 2009. "The Reader-to-Leader Framwork: Motivating Technology-Meditated Social Participation," *AIS Transactions on Human-Computer Interaction* (1:1), pp. 13–32.

Quan y Hasse , A. , Wellman, B., Witte, J., and Hampton, K. 2002. "Capitalizing on the Internet: Social Contact, Civic Engagement, and Sense of Community," in *The Internet in Everyday Life* , B. Wellman and C. Haythornthwaite (eds.), Oxford, UK: Blackwell, pp. 291–394.

Rheingold, H. (1993). *The Virtual Community: Homesteading on the Electronic Frontier*, New York: HarperCollins.

Riggins, F. J. 2003. “Market Segmentation and Information Development Costs in a Two-tiered Fee-based and Sponsorship-based Web Site,” *Journal of Management Information Systems* (19:3), pp. 69-81.

Rob, R., and Waldfogel, J. 2006. “Piracy on the High C’s: Music Downloading, Sales Displacement, and Social Welfare in a Sample of College Students,” *Journal of Law and Economics*, (49:1), pp. 29–62.

Rosenbaum P. R., and Rubin, D. B. 1983. “The Central Role of the Propensity Score in Observational Studies for Causal Effects,” *Biometrika* (70:1), pp. 41–55.

Rubin, D.B. 1980. “Bias Reduction Using Mahalanobis–Metric Matching,” *Biometrics* (36), pp. 293-298.

Scott, C. A. 1976. “The Effects of Trial and Incentives on Repeat Purchase Behavior,” *Journal of Marketing Research* (13), pp. 263–269.

Shankar, V., Rangaswamy, A., and Pusateri, M. 1999. “The Online Medium and Customer Price Sensitivity,” working paper.

Shiller, B., and Waldfogel, J. 2008. “Music for a Song: An Empirical Look at Uniform Song Pricing and its Alternatives,” working paper.

Sinkkonen J., Aukia, J., Kaski, S. 2007. “Inferring Vertex Properties from Topology in Large Networks,” *Working Notes of the 5th International Workshop on Mining and Learning with Graphs (MLG'07)*, Florence, Italy: Universita degli Studi di Firenze.

Srinivasan, S. S., Anderson, R., and Ponnavolu, K. 2002. “Customer Loyalty in E-commerce: An Exploration of Its Antecedents and Consequences,” *Journal of Retailing* (78:1), pp. 41–50.

Susarla, A., Oh, J., and Tan, Y. 2011. "Social Networks and the Diffusion of User-Generated Content: Evidence from YouTube," *Information Systems Research*, Published online in Articles in Advance, April 8, DOI: 10.1287/isre.1100.0339.

Teece, D. J. 2010. "Business Models, Business Strategy and Innovation," *Long Range Planning* (43:2-3), pp. 172–194.

Thompson, S. A., and Sinha, R. K. 2008. “Brand Communities and New Product Adoption: The Influence and Limits of Oppositional Loyalty,” *Journal of Marketing* (72:6), pp. 65–80.

Venkatesh, V., Morris, G. M., and Phillip L. A. 2000. “A Longitudinal Field Investigation of Gender Difference in Individual Technology Adoption Decision-Making Processes,” *Organizational Behavior and Human Decision Processes* (83:1), pp. 33– 60.

Venkatraman, M. P. 1990. “Opinion Leadership: Enduring Involvement and Characteristics of Opinion Leaders: A Moderating or Mediating Relationship,” in *Advances in Consumer Research*, vol. 17, M. E. Goldberg, G. Gorn, and R. W. Pollay (eds.), Provo, UT: Association for Consumer Research, pp. 60–67.

Villanueva, J., Yoo, S., and Hanssens, D. M. 2008. “The Impact of Marketing-induced vs. Word-of-mouth Customer Acquisition on Customer Equity,” *Journal of Marketing Research* (45:1), pp. 48–59.

Wang, Q., and Xie, J. 2011. “Online Social Interactions: A Natural Experiment on Word of Mouth versus Observational Learning," *Journal of Marketing Research*, forthcoming.

Wasko, M. M., and Faraj, S. 2005. "Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice," *MIS Quarterly* (29:1), pp. 35–57.

Wenger, E. 1998. *Communities of Practice: Learning, Meaning and Identity*, Cambridge, UK: Cambridge University Press.

Wiener, Y. 1982. "Commitment in Organizations: A Normative View," *Academy of Management Review* (7), pp 418- 428.

Wilson, F. 2006. “My Favorite Business Model,” *VC Blog* (March), (http://www.avc.com/a_vc/2006/03/my_favorite_bus.html).

Wirtz, B.W., Schilke, O., and Ullrich, S. 2010. “Strategic Development of Business Models: Implications of the Web 2.0 for Creating Value on the Internet,” *Long Range Planning* (43:2–3), pp. 272–290.

Ye, L. R., Zhang, Y., Nguyen, D. D., and Chiu, J. 2004. “Fee-based Online Services: Exploring Consumers' Willingness to Pay,” *Journal of Technology and Information Management* (13:2), pp. 134–141.

Yoo, Y., and Alavi, M. 2004. “Emergent Leadership in Virtual Teams: What do Emergent Leaders Do?” *Information and Organization* (14), pp. 27-58.

Zeithaml, V. A. 1988. “Consumer Perceptions of Price, Quality, and Value: A Means-End Model and Synthesis of Evidence,” *Journal of Marketing* (52), pp. 2–22.

Zhang, J. 2010. “The Sound of Silence: Observational Learning from the U.S. Kidney Market," *Marketing Science* (29:2), pp. 315–335.

Table 1. Levels of Participation

	Communities of Practice (Wenger 1998)	Participation Levels (Kim 2000)	Social Technographics Tool (Li and Bernoff, 2008)	Reader-to-Leader Framework (Preece and Schneiderman 2009)
Content Consumption	*Peripheral* Does not participate in the community	*Visitor* Outside, unstructured participation	*Joiners and Spectators* Reading content and creating a user page	*Reader* Only consumes articles/content
Content Organization	*Inbound* Initial participation activity on the way to full participation	*Novice* Newcomer is becoming invested in the community	*Collectors* Tagging content, voting and simple ratings	*Contributor* Contributes some content to the website's community
Community Involvement	*Insider* Full participation in the community	*Regular* Fully committed community participant	*Critics* Posting comments, critique, participating in discussions	*Collaborator* Participates in group projects and cooperation
Community Leadership	*Boundary* Spans boundaries and links communities of practice	*Leader* Sustains membership participation and brokers interactions	*Creators* Publishing original user-generated content, publishing a blog	*Leader* Leads the community, moderates discussions

This table depicts the different frameworks of community behavior over time. The Communities of Practice model (Wenger 1998, based on early work by Lave and Wenger 1991) focuses on communities of practice in which a participant becomes increasingly involved and progresses to the center of the community. Kim (2000) focuses on online behavior over time and stresses the user's ongoing effort. Li and Bernoff (2008) develop their levels by categorizing different participation activities of the Web 2.0 era, differentiating between content organization (collectors), participation (critics), and full involvement in the form of creation. Preece and Schneiderman (2009) emphasize that at the stage of full community participation there are also more collaboration and socialization roles.

Table 2. Descriptive Statistics

Type of Membership:		Non-paying user			Subscriber		
		Mean	Median	Variance	Mean	Median	Variance
Content Consumption	Tracks listened to	17,616	11,265	477,622.677	21,688	11,039	998,060.194
Content Organization	Playlists created	0.77	1	0.47	1.29	1	7.15
	'Loved' tracks tagged	65.97	11	41,872	210.34	83	314,062
	Tags created	9	1	1,400.19	21.27	2	5,298.45
Friends	No. of friends	14.56	9	640.923	21.19	10	1,196.87
Subscriber Friends	No. of subscriber friends	.42	0	1.86	2.82	1	31.812
Community Participation	Posts published to forums	9.12	0	7,596.37	27.31	0	75,401.53
	Groups joined	5.27	2	168.69	8.98	3	463.08
	Blog entries published	0.42	0	2.24	0.89	0	5.62
Community Leadership	Groups led	0.07	0	0.165	0.17	0	0.452
Demographics	Age	23.08	21	39.15	29.43	27	88.41
	Gender (0= Male, 1= Female)	0.34	0	0.22	0.29	0	0.20
	Usage (Days)	720.53	662.33	98,666.55	652.08	600	335,075.46

Table 3. Comparing Activity Levels of Subscribers and Non-paying Users

	Variable Name	Subscriber Mean	User Mean	Ratio	U-test *P* Value	*t*-Test *P* Value
Content Consumption	No. of tracks listened to	21,689	17,617	1.23	0.427	0.00***
Content Organization	No. of playlists	1.29	0.77	1.67	0.00***	0.00***
	No. of loved tracks	210.34	65.97	3.18	0.00***	0.00***
	No. of tags created	21.27	9	2.40	0.00***	0.00***
Friends	No. of friends	21.19	14.56	1.45	0.00***	0.00***
Subscriber Friends	No. of subscriber friends	2.82	.42	6.71	0.00***	0.00***
Community Participation	No. of group memberships	8.98	5.27	1.70	0.00***	0.00***
	No. of posts to forums	27.31	9.12	2.99	0.00***	0.00***
	No. of blog entries	0.89	0.42	2.11	0.00***	0.00***
Community Leadership	No. of groups led	0.17	0.07	2.42	0.00***	0.00***
Demographics	User's age	29.43	23.08	1.27	0.00***	0.00***
	Days since joining the website	652.08	720.53	1.10	0.00***	0.00***

*** - Significant at the 0.01 level

Table 4. Correlation Matrix

	Gender	Age	Days	Number of Friends	Number of Subscriber Friends	Tracks Listened To	Playlists Created	'Loved' Tracks Tagged	Forum Posts Published	Groups Joined	Groups Led	Blog Entries Written	Tags Created	Subscriber
Gender	1.000													
Age	-.186**	1.000												
Days	-.063**	-.022*	1.000											
Num. of Friends	.062**	-.063**	.172**	1.000										
Num of Sub. Friends	.021*	.149**	.097**	.717**	1.000									
Tracks Listened To	-.080***	-.059**	.367**	.343**	.245**	1.000								
Playlists Created	.003*	.139**	-.034**	.146**	.238**	.079**	1.000							
'Loved' Tracks Tagged	-.008	.115**	.047**	.208**	.284**	.179**	.350**	1.000						
Forum Posts Published	-.009	.019*	.063**	.134**	.155**	.161**	.009**	.091**	1.000					
Groups Joined	-.028**	-.043**	.126**	.373**	.312**	.242**	.065**	.165**	.148**	1.000				
Groups Led	-.044**	-.014	.127**	.236**	.185**	.189**	.021**	.067**	.122**	.376**	1.000			
Blog Entries Written	-.002**	.028**	.173**	.293**	.263**	.251**	.063**	.130**	.144**	.267**	.251**	1.000		
Tags Created	-.035**	.066**	.078**	.172**	.178**	.159**	.110**	.216**	.101**	.221**	.161**	.204**	1.000	
Subscriber?	-.051**	.363**	-.074**	.121**	.327**	.068**	.144**	.186**	.055**	.112**	.088**	.124**	.122**	1.000

**. Correlation is significant at the 0.01 level (2-tailed).

Table 5. Binary Logistic Regression Model for Subscribing Decision

	Content Consumption	+ Content Organization	+ Friends	+ Subscriber Friends	+ Community Participation & Leadership	+ Usage & Demographics
	A	B	C	D	E	F
	B (S.E) / EXP(B)	B (S.E) / EXP(B)	B (S.E) / EXP(B)	B (S.E) / EXP(B)	B (S.E) / EXP(B)	B (S.E) / EXP(B)
Number of tracks listened to (in thousands)	.005*** (.001)	.002*** (.001)	.000 (.001)	.000 (.001)	-.001 (.001)	.007*** (.001)
	1.005	1.002	1.000	1.000	.999	1.007
Number of playlists	___	.323*** (.024)	.320** (.025)	.250*** (.026)	.249*** (.026)	.169*** (.026)
		1.381	1.377	1.284	1.282	1.184
Number of 'loved' tracks	___	.002*** (.000)	.002*** (.000)	.001*** (.000)	.001*** (.000)	.001*** (.000)
		1.002	1.002	1.001	1.001	1.001
Number of tags	___	.003*** (.001)	.002*** (.001)	.002*** (.001)	.002*** (.001)	.001 (.001)
		1.003	1.002	1.002	1.002	1.001
Number of friends	___	___	.006*** (.001)	-.062*** (.002)	-.064*** (.003)	-.047*** (.003)
			1.006	.940	.938	.954
Number of subscriber friends	___	___	___	.908*** (.026)	.905*** (.026)	.784*** (.027)
				2.480	2.472	2.375
Number of group memberships	___	___	___	___	.004** (.002)	.007*** (.002)
					1.004	1.007
Number of groups led	___	___	___	___	.184*** (.058)	.204*** (.059)
					1.201	1.226
Number of blog entries	___	___	___	___	.038** (0.15)	.049*** (.015)
					1.039	1.051
Number of posts to forums	___	___	___	___	.000 (.000)	.000 (.000)
					1.000	1.000
Age	___	___	___	___	___	.082*** (.003)
						1.086
Gender	___	___	___	___	___	-.079 (.055)
						.924
Days	___	___	___	___	___	-.001*** (.000)
						.999
Constant	-1.122 *** (0.25)	-.1.600*** (.035)	-1.651*** (.036)	-1.411*** (.039)	-1.410*** (.039)	-2.956*** (.109)
	.326	.202	.192	.244	.244	.052
Revised Constant						-6.355

Observations: 13004 ; Log likelihood: 10,812.498

Cox & Snell R-Square: 0.280, Nagelkerke R-Square: 0.408

- significant at the 0.05 level ; *- significant at the 0.01 level

Table 6. Cox Regression Model for Subscribing Decision

		B	S.E.	Wald	df	Hazard Exp(B)
Content Consumption	Number of music tracks (In thousands)	-.007***	.001	106.492	1	.993
Content Organization	No. of playlists	.027***	.005	38.712	1	1.027
	No. of Loved tracks	.000***	.000	21.396	1	1.000
	No. of tags created	.000	.000	.412	1	1.000
Friends	No. of friends	-.013***	.001	143.943	1	.987
Subscriber Friends	No. of sub. friends	.116***	.005	466.304	1	1.123
Community Participation	Posts published	.000	.000	.002	1	1.000
	Groups joined	.002***	.001	7.134	1	1.002
	Blog entries published	.017	.008	2.311	1	1.018
	Groups led	.051**	.023	5.605	1	1.053
Demographics	Age	.060***	.002	1211.184	1	1.062
	Gender	.169	.039	18.852	1	1.184

N (non-paying users) = 37,480, N (subscribers) = 3,430

Overall Model Estimation: $\chi 2$ = 5,058.890. df = 11, p = 0.00,

-2 Log likelihood = 63,387.610

- significant at the 0.05 level ; *- significant at the 0.01 level

Table 7 – Propensity Score Matching

Treatment	A Group Membership	B Group Leadership	C Forum Postings	D Blog Entries	E Community Activity (Heckman)	F Community Activity (Mahalanobis)
Number of Matched Cases	29,941	2,423	16,375	6,097	30,882	30,882
Percentage of subscribers among treated cases	8.5%	15.2%	10%	12.5%	8.4%	8.4%
Percentage of subscription among non-treated cases	6.9%	9.8%	7%	9.8%	7%	6.2%
Diff Mean	1.6%	5.4%	3.0%	2.6%	1.4%	2.2%
t-test (Diff Mean > 0)	7.38***	5.78***	9.83***	4.79***	6.61***	11.07***
Diff Mean (Std. Err)	.002	.009	.003	.005	.001	.001
Std.Dev	.37	.45	.39	.43	.25	.35

Figure 1. Last.fm Screen Shot (User Page)

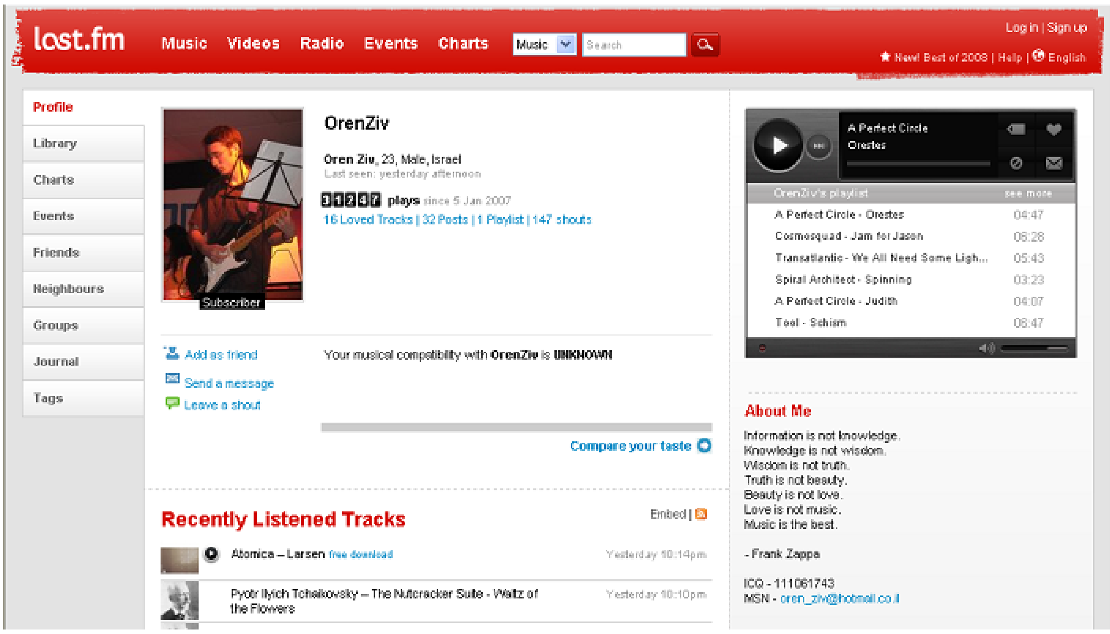
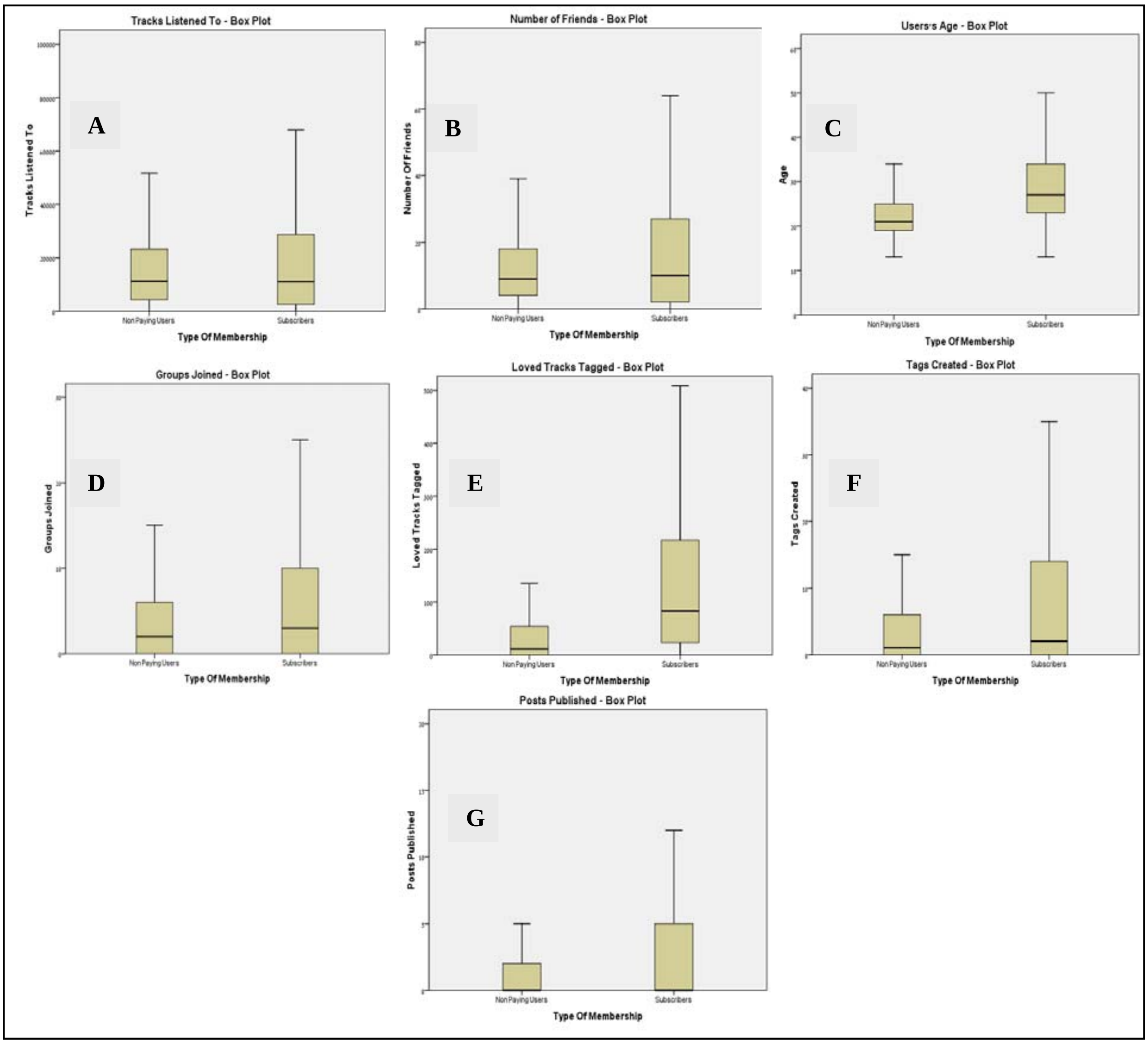


Figure 2. Box Plot Graphs



Panel A presents the statistical distribution differences between non-paying users (on the right) and subscribers (on the left) for the 'music tracks listened to' variable. Similarly, Panel B presents the distribution of the variable 'user's number of friends'; Panel C the distribution of the user's age; Panel D the distribution of the number of groups joined by the user; Panel E the distribution of the number of tracks that were tagged as 'loved'; Panel F the distribution of tags created; and Panel G the distribution of the number of posts published to forums.

Appendix 1. Comparison of Means Before and After Propensity Score Matcing

	Treatment 1 – Group Membership			Treatment 2 – Group Leadership		
	Treatment 1	Control		Treatment 2	Control	
		Pre	Post		Pre	Post
No. of tracks listened to	21,644.59	9961.55	23,247	34737.95	16.957.24	37944
No. of playlists	.85	.73	.84	.92	.80	.80
No. of 'loved' tracks	92.91	45.95	80.25	148.25	73.90	134.63
No. of tags created	13.14	3.29	13.665	32.22	8.71	27.27
No. of friends	19.21	6.23	18.16	35.17	13.92	31.98
No. of sub. Friends	.82	.19	.55	1.92	.54	1.38
No. of group memberships	8.16	0	0	21.39	4.64	10.09
No. of posts to forums	15.37	.41	1.56	73.76	6.89	18.27
No. of blog entries	.62	.1	.2	1.88	.37	.84
No. of groups led	.12	0	0	1.41	0	0
Users' age	23.11	24.70	23	23.16	23.64	22.44
Users' gender	0.31	0.36	0.39	0.2	0.34	0.32
Days since joining the website	761.613	602.783	699.05	921.56	701.82	814.94

* Note, that for the treatment group, there is not difference in mean before and after the matching process

Treatment 3 – Group Postings　　　Treatment 4 – Journal Postings

	Treatment 3	Control		Treatment 4	Control	
		Pre	Post		Pre	Post
No. of tracks listened to	27,036.60	12,377.88	27,901	28,537.9	16,211.6	31,860
No. of playlists	.87	.77	.85	.97	.78	.94
No. of 'loved' tracks	113.51	56.32	112.4	143.61	67.27	132.57
No. of tags created	18.18	5.03	15.42	25.70	7.44	20.25
No. of friends	24.08	9.61	23.60	27.57	13.06	26.9
No. of sub. Friends	1.11	.32	.93	1.52	.47	1.19
No. of group memberships	10.80	2.38	5.04	12.60	4.43	8.71
No. of posts to forums	27.98	0	0	41.33	5.59	21.02
No. of blog entries	1.04	.10	.15	3.21	0	0
No. of groups led	.19	.01	.02	.27	.05	.1
Users' age	22.84	24.09	22.62	23.59	23.62	23.27
Users' gender	0.3	0.35	0.36	0.33	0.33	0.33
Days since joining the website	831.618	638.351	724.24	842.872	692.834	781.06

* Note, that for the treatment group, there is not difference in mean before and after the matching process

	Treatment 5 – Community Activity			Treatment 6 – Community Activity (Mahalanobis matching)		
	Treatment 5	Control		Treatment 6	Control	
		Pre	Post		Pre	Post
No. of tracks listened to	21,429.62	9,120.73	23,604	21,429.62	9,120.73	22,927
No. of playlists	.85	.72	.83	.85	.72	.84
No. of loved tracks	92.26	42.01	84.69	92.26	42.01	70.37
No. of tags created	12.87	2.81	16.55	12.87	2.81	14.15
No. of friends	18.82	5.70	17.10	18.82	5.70	16.25
No. of sub. friends	.80	.17	.52	.80	.17	.49
No. of group memberships	7.78	0	0	7.78	0	0
No. of posts to forums	14.84	0	0	14.84	0	0
No. of blog entries	.63	0	0	.63	0	0
No. of groups led	.11	0	0	.11	0	0
Users' age	23.16	24.76	23.13	23.16	24.76	23.05
Users' gender	0.32	0.36	0.4	0.32	0.36	0.41
Days since joining the website	760.053	586.872	675.9	760.053	586.872	678.27

* Note, that for the treatment group, there is not difference in mean before and after the matching process