MIS Quarterly

SPECIAL ISSUE: DIGITAL BUSINESS STRATEGY

CONTENT OR COMMUNITY? A DIGITAL BUSINESS STRATEGY FOR CONTENT PROVIDERS IN THE SOCIAL AGE[1]

Gal Oestreicher-Singer and Lior Zalmanson

Recanati Graduate School of Business, Tel Aviv University, Tel Aviv 69978 ISRAEL

{galos@post.tau.ac.il} {zalmanso@post.tau.ac.il}

The content industry has been undergoing a tremendous transformation in the last two decades. We focus in this paper on recent changes in the form of social computing. Although the content industry has implemented social computing to a large extent, it has done so from a techno-centric approach in which social features are viewed as complementary rather than integral to content. This approach does not capitalize on users' social behavior in the website and does not answer the content industry's need to elicit payment from consumers. We suggest that both of these objectives can be achieved by acknowledging the fusion between content and community, making the social experience central to the content website's digital business strategy.

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 aimed primarily 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 participation" on 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 suggest that firms whose digital business models remain viable in a world of "freemium" will be those that take a strategic rather than techno-centric view of social media, that integrate social media into the consumption and purchase experience rather than use it merely as a substitute for offline soft marketing. We provide new evidence of the importance of fusing social computing with content delivery and, in the process, lay a foundation for a broader strategic path for the digital content industry in an age of growing user participation.

Keywords: Premium services, social media, online communities, propensity score matching, UGC, digital business strategy, ladder of participation

[1] Anandhi Bharadwaj, Omar A. El Sawy, Paul A. Pavlou, and N. Venkatraman served as the senior editors for this special issue and were responsible for accepting this paper.

Introduction

Rapid technology changes over the past two decades have presented the content industry with a vast number of opportunities as well as new challenges. The relative ease of digitizing text, music, and video, coupled with the ubiquity of content consumption technologies such as personal computers and MP3 players, have encouraged content providers to rely increasingly on electronic offerings, thereby reducing their production and operational costs considerably. Correspondingly, distribution costs have been lowered by the process of net-enablement—a term that refers to the incorporation of digital networks into content delivery, management, and marketing (Straub and Watson 2001; Wheeler 2002). Indeed, consumers have largely switched to online consumption of content: Americans bought 1.27 billion digital tracks online in 2011, which accounted for more than half of all music sales (Nielsen 2012); in 2010, 59 percent of Americans consumed online news on a regular basis (Pew Research Center 2010). However, at the same time, these changes have also lowered switching costs for consumers, increased piracy, and created masses of new free content, raising a need for content providers to adopt new strategic thinking in order to sustain competitive advantage.

The recent widespread adoption of social computing marks another dramatic step in the evolution of the online content industry.[2] Social computing technologies are giving rise to new forms of user interaction and cocreation; their implications for business strategy have yet to be fully elucidated. In particular, researchers and practitioners must cope with issues such as "how business can generate value through social networks, how communities in these initiatives can gain value, and how to assess the costs and benefits of social computing initiatives" (Parameswaran and Whinston 2007, p. 344).

The growing literature on digital business strategy has stressed the transformational role that IT plays in contemporary business processes (Banker et al. 2006; El Sawy 2003; Lu and Ramamurthy 2011; Sambamurthy et al. 2003). In the same spirit, we assert that social computing can be a transformational force for the content industry. Initially perceived as a threat and as a potential source of piracy and disintermediation, many incumbents now regard social offerings as complementary to content and as an integral part of content delivery strategy (Adams 2011). Nevertheless, the benefits of social computing features are still under debate, and the industry often questions their potential. It is our conjecture that firms fail to reap tangible benefits from social computing because they have largely implemented social features using a techno-centric approach rather than a strategic one: They view social computing features as add-ons to traditional content. These implementations are useful, but they miss the broader promise of social computing for content websites, one that can only be realized by taking a strategic approach.

In this paper we suggest a first step toward this strategic approach. We rely on the notion of a "fusion view of IT" (El Sawy 2003) and contend that social computing should no longer act as a merchandising complement to the firm's value proposition; it is not a technological enhancement to the product, nor is it simply an innovative marketing tool. Rather, it is an inherent part of the firm's product, the core of its digital business strategy. The adoption of this digital business strategy and its value proposition transforms the main role of the firm from providing content to establishing content-related and IT-enabled social experiences for its users (we refer to these experiences as *social content*). This paradigm is constructed to reflect users' changing expectations in a "social" age characterized by the rise of social media platforms and corresponding changes in online behavior.[3]

We continue and add to the literature on digital business strategy by focusing on the "ladder of participation" paradigm—a user segmentation scheme based on the evolving nature of participatory behavior in online communities. We believe this to be the key difference between social computing and other net-enabled technologies. We discuss the value that users at different participatory levels derive from social content websites. We then define how value is captured: We conjecture that business models that cater to different participatory levels in the ladder are inherently well suited to social content websites, as they offer the firm a method to capitalize on users' evolving commitment to the firm's offerings. We focus on the "freemium" model, in which a website offers most of its services for free while restricting only some premium features to fee-paying consumers, and discuss how this model is suitable for capturing the value created in socially active users.

Next, we empirically demonstrate the relations between social content consumption, users' participation patterns, and their willingness to pay, using data from Last.fm. Last.fm is a proprietary content website that serves both as an online radio and as a social networking site and operates under the

[2]Social computing is a collective name for IT technologies that facilitate collective action and social interaction online (Parameswaran and Whinston 2007).

[3]Yoo (2010) discusses "digital natives," users who spend most of their lives surrounded by computers, mobile devices, and video games. For them, technological acceptance is taken for granted and questions of digitization seem irrelevant.

freemium business model. Even though the premium services it offers mainly improve the proprietary-content consumption experience (for example, by increasing bandwidth), we find that acts of payment are strongly associated with social computing-based features. 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 social computing features. Users who act as leaders in this community show even higher propensity to pay.

Our results underline the importance of viewing social computing as being integral to the product strategy of content providers. We provide new and unique evidence of the importance of introducing social computing as a means of both value creation and capture and as a source of competitive advantage for content providers, leading to insights into the causal effects of social engagement on consumers' buying decisions, and laying the foundation for a broader strategic path that content providers can follow in an age of growing user participation.

The Fusion of Social Computing and Content

In the past, most information technologies adopted by organizations were perceived as tools added to boost productivity or lower operational costs. IT was connected to people's work as an artifact that could be pushed aside if necessary, while work might still continue (El Sawy 2003). Consequently, the broad strategic view was that IT strategy must be aligned with the firm's business strategy (Henderson and Venkatraman 1993). However, during the last two decades, the digital infrastructure of business and society has shifted radically, and researchers and managers alike have acknowledged that the role of IT has undergone a transformation. IT has become immersed in the workspace and in homes, developing into an unavoidable part of both daily routines and business processes. A newer view has been suggested in which technology is not only immersed in the business environment but is fused with it, such that IT and business strategy are indistinguishable to our perception and form a unified fabric (El Sawy 2003). This change has called attention to the need for firms to develop digital business strategies, an approach that takes into account the embeddedness of technology in the business process and daily lives.

Social computing technologies are a recent example of technology that has become deeply embedded in our daily routines and personal interactions, to the point where it is nearly impossible to disentangle business and social processes from their underlying IT infrastructures. Moreover, social computing creates a shift in which computing power is transferred from organizations to individuals, empowering users with relatively low technological sophistication to use the web to "manifest their creativity, engage in social interaction, contribute their expertise, share content, collectively build new tools and disseminate information" (Parameswaran and Whinston 2007, p. 753).

The content industry in particular has been transformed by social computing technologies. The evolution of the content industry's view on social computing can be understood in light of El Sawy's (2003) three phases of IT strategy: connection, immersion, and fusion. Table 1 summarizes the three phases of IT in the context of social computing in the content industry. Until recently, many content providers were in the so-called *connection* phase of IT strategy, perceiving the web only as an additional channel for traditional newspaper, radio and television content offerings (O'Reilly 2005). As such, these providers viewed social computing as a source of competition and therefore as a threat. Their perception led them to adopt strategies that emphasized quality and the trustworthy attributes of traditional and proprietary content over new alternatives (Posner 2005; Neuberger and Nuernbergk 2010).

With the immense success of social computing, many content providers entered the *immersion* phase of IT strategy (El Sawy 2003; Table 1), embracing social computing features as a means of stimulating consumption of traditional content (for example, by allowing word of mouth) and prolonging users' length of stay on their websites. Users were encouraged to actively engage with the content and with one another, by posting comments, conversing on user forums, and sharing user-generated content, either within the websites themselves or through existing popular social computing platforms (Clemons 2009; e.g., many websites feature buttons, such as Facebook's "Connect" button and Google Plus's "+1" button, which enable users to share content with friends on the corresponding social computing platforms). Nevertheless, this positive approach toward social computing still puts the emphasis on the content offered rather than on the social experience (i.e., firms still perceive social computing features as complementary rather than as integral to the firm's offerings). For example, the *New York Times* website (NYTimes.com) now includes social computing features that allow users to comment on articles, rate them, and share them via social networking sites. However, in essence, the site still functions as a traditional newspaper—the user's fundamental experience of the site revolves around the consumption of proprietary news content, presented in accordance with the

Table 1. Three Phases of IT in the Context of Social Computing in the Content Industry

Phase	Broader Industry View on IT (Based on El Sawy 2003)	Content Industry View on Social Computing
Connection	IT is used as a tool to help people in their work. It is a separable artifact that can be connected to people's work actions and behaviors	**Social Computing is a mere tool and its use is optional**. Many ignore social computing altogether, and some perceive it a threat and promote against it.
Immersion	IT is immersed as part of the business environment and cannot be separated from work and the systemic properties of interorganizational relationships.	**Social Computing is a valuable complementary offer.** Social Computing platforms are being widely used to attract users and differentiate websites from their competitors. Practically, social features are add-ons next to traditional content, which is still the main focus of the offer.
Fusion	IT is not only immersed but is fused with the business environment such that they are indistinguishable to our perception and form a unified fabric.	**Social Computing cannot be differentiated from the content experience**. Content is inherently a social experience. Content providers create social experiences in which the user creates a personal online identity and interacts with others. This social experience takes center stage on the website, replacing content.

vision of the editors or site administrators, and this experience remains more or less consistent regardless of the presence of social features.

A different approach for social computing in the content industry is based on the *IT fusion* view proposed by El Sawy (2003; Table 1). According to this view, content websites are inherently social, and as such they cannot separate content from social computing elements. Indeed, many content-related experiences are, at their core, social. People derive great pleasure from watching films together with friends, attending concerts in groups, and discussing news articles and texts in organized "knitting circles" or other informal gatherings. Thus, the offer made by the website should emphasize the social aspect of content consumption: meaning the creation and enhancement of relationships.

In recent years we have seen the rise of many sites that have been described as social media platforms, such as Facebook, Digg, and LiveJournal, among others. These platforms enable users to create an on-site identity (in many cases, by designing a personal page), make online friends, curate content for others to enjoy, attend virtual social events, participate in social games, create collaborative user-generated content and build ongoing reputations. Social media platforms have understood that consuming content, and forming relationships around it by discussing, sharing, and reacting to it, are parts of the same experience.

The content industry, in contrast, has not fully grasped the implications of this reality. The next step for the content industry is to build an arena in which social interaction can occur. To do so, content providers must supply users with social experiences based on shared content, not merely add a "social" layer to their traditional offerings. This implies that users should be able to interact with fellow users through the website, not merely interact with content itself. This approach is user-centric, positioning the users' personal experience rather than the content itself at the core of the online product. It creates a shift in the role of the content industry, rendering it an enabler of experiences rather than a mere provider of content. The result is a hybrid between "content provider" and "virtual community" business models (Weill and Vitale 2001) that can be referred to as a *social content website*. This use of social computing as an inherent part of the value proposition is unique to the IT fusion phase discussed above. Table 2 provides a detailed comparison of the content website characteristics under each of the three phases of IT discussed (connection, immersion, and fusion).

As content consumption becomes a social experience, value creation becomes dependent on the social environment as well. For example, users browsing the website can be informed in real time who is consuming which content or how popular different content items are. Similarly, enabling ratings and comments allows users to influence the navigation and consumption decisions of other users. Clearly, providing a platform in which users can organize discussions around different topics will inform other users, and allowing users to moderate content can improve content quality. By constructing an array of value-creating features based on social computing, firms can encourage user participation and contri-

Table 2. Comparison of Content Website Approaches

	Traditional Content Website (Connection Phase)	Content Website with Social Computing (Immersion Phase)	Social Content Website (Fusion Phase)
Value Proposition	Users derive value from consuming firm-delivered content.	Users derive value from consuming firm-delivered content and from interaction with other users on the website via social computing features.	Users derive value from an ongoing content-based social experience in which they can fulfill different roles in the site and form meaningful relationships.
Value Creation	Created by the firm by producing/delivering content.	Created mainly by the firm by producing/delivering content and also by social interaction.	Created by both firm and users through a ladder of participation.
Value Capture	Advertisements, paying for access to content.	Advertisements, paying for access to content.	Advertisements, freemium.
Segmentation scheme	Content consumption levels.	Content consumption levels, content taste and valuation (via social computing).	Content and social consumption based on the ladder of participation.
Pattern of Interaction between firm and users	Feedback in the form of targeted messages or controlled guestbook/ feedback forum.	Interaction throughout various variations of social computing add-ons—talkbacks, forum/blog postings.	Interaction throughout a unified social platform.
Pattern of Interaction between users	Not available on site.	Interaction through conversations using social computing features-forums, blogs.	Socializing around content, social curation of content through user pages.

bution. However, we propose that in order to produce a value-creating environment that also facilitates successful value capture and profits, firms must first understand users' behavioral dynamics in a social context.

The Ladder of Participation: The Dynamics of Social Content

Past research has investigated participation patterns in a communal setting, both offline and online. 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'" (p. 37). More recently, there have been various attempts at creating more thorough frameworks that model users' behavior specifically in online community contexts. Kim (2000), for example, differentiates among several participation roles: (1) the **visitor**, who exhibits unstructured participation; (2) the **novice,** who invests time and effort in order to become a (2) **regular**, who displays full commitment; and (4) 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 3.

It seems that there is a high degree of consensus among academics and practitioners regarding the various stages of the user's membership life cycle. 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. Clearly, users who "move up the ladder" invest more effort in the website and create more value than users who just consume content. It is also clear that each level is associated with different social

Table 3. Levels of Participation

	Communities of Practice (Wegner 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.

computing features. Content organization includes the option to tag content and recommend it. Community participation includes joining affinity groups, posting comments, and contributing content. Community leadership entails moderation of user groups and their respective content.

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. (2011) offered an overarching theory: the commitment-based approach. In their study, Bateman et al. showed that users' behavior on content sites is directly linked to their commitment levels, as defined by 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 benefit 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 receives in response 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 users whose videos attract more attention subsequently contribute greater quantities 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. 2011).[4] 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 repeated 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. 2011). 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). Granted, not all users will end up being community leaders, and not everyone will be involved in the community; it is actually not necessary for everyone to do so. The value proposition depends on having a critical mass of users carrying out different contributing acts.

Linking Participation to Value Capture and Willingness to Pay

Value capture has proved to be challenging for the traditional online content provider. 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' increased tendency 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 introduced new challenges for online content retailers (see also Bhattacharjee et al. 2003; Gopal et al. 2004).

When content providers first adopted social computing features, they resorted to advertising as their base revenue model. However, advertising is essentially "flat"; it does not utilize the insights that come with better understanding of users' behavioral dynamics in a social context. The different levels of participation call for a business model that better allows for user segmentation.

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

An emerging business model that allows for such segmentation is the *freemium* (or two-tiered) model, wherein basic services are provided for free, and premium services are offered for a fee (Doerr et al. 2010; Hung 2010; Riggins 2003; Teece 2010). The underlying assumption of the freemium model is that delivering a product for free can attract a large number of users and encourage participation, and a small fraction of participants will pay for the premium offer.

A careful strategy for user segmentation and a tailored attractive premium offer are the key to the success of the freemium model. One widespread approach is offering a portion of the content for free and the rest for a fee. However, researchers have stressed that this 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).

Linking this to the previous discussion on levels of participation, we suggest that a successful strategy for firms using the freemium model should incorporate a new segmentation scheme. That is, premium offers should be aimed at users with higher levels of participation. As discussed, those users exhibit higher levels of commitment to the website. Marketing scholars have noted that commitment can yield loyalty, which encourages payment (Beatty and Kahle 1988; Dick and Basu 1994). Loyalty is defined as a composite blend of both brand attitude and behavior and is associated with increased purchases (Pritchard and Howard 1997). It is also associated with the conscientious willingness to pay a premium price, or alternatively the exhibition of price indifference (Fornell 1992; Raju et al. 1990; Zeithaml et al. 1996).

In the context of organizational commitment theory, Fullerton (2003, 2005) found that only consumers who exhibited affective commitment expressed their loyalty in the form of "willingness to pay more," while consumers who exhibited continuance commitment were not willing to pay for premium service.

Connecting this to the ladder of participation discussed above, we propose that willingness to pay for premium services is not associated with content consumption alone but rather is associated with content organization and community participation as well. Specifically, by testing the following hypotheses, we aim to show that users on higher rungs of the ladder of participation are more willing to pay compared with users on lower rungs of the ladder.

First we hypothesize that any level of participation on the website—content consumption, content organization, as well as community participation—will be associated with propensity to subscribe:

H1: User participation in the website is positively associated with the likelihood of subscribing to premium services.

The ladder of participation suggests that "higher" levels of participation will be associated with higher willingness to pay. For example, leadership roles in the community are associated with the strongest form of commitment, normative commitment, which reflects a sense of obligation to the website. Thus, our second hypothesis compares the effects of different levels of website engagement on willingness to pay.

H2(a): Content organization will have a stronger association with the decision to subscribe to premium services than will content consumption.

H2(b): Community participation will have a stronger association with the decision to subscribe to premium services than will content organization and content consumption.

H2(c): Community leadership will have a stronger association with the decision to subscribe to premium services than will community participation, content organization, and content consumption.

Nearly by definition, commitment is a long process and cannot happen overnight. Users who are climbing the ladder of participation are experimenting with new content and social activities in which they invest increasing time and effort. It is, therefore, reasonable to expect that participating users will become more committed faster. In the context of content websites, this means they are likely to make the decision to subscribe to premium services sooner. Hence, our third hypothesis is

H3: User participation is positively associated with a shorter period of time of free usage.

As with the subscription decision (H2), we can compare the effects of different levels of participation on time to subscription. We formulate this comparison in our fourth hypothesis:

H4(a): Content organization will have a stronger association with a shorter period of time of free usage than will content consumption.

H4(b): Community participation will have a stronger association with a shorter period of time of free usage than will content organization and content consumption.

H4(c): Leadership of groups will have a stronger association with a shorter period of time of free usage than will participation in groups.

Data Collection and Preparation

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[5] 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.

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), link to friends' pages, 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 (see Figure 1 for illustration).

Last.fm implements the freemium 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:[6]

- Improved infrastructure: removal of ads from the subscriber's page and top-priority quality-of-service on web and radio servers.

[5]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.

[6]In April 2009, Last.fm changed its business model in certain countries and currently allows only paying subscribers to stream label-owned music. However, in the United Kingdom, Germany, and the United States, the model has not changed.

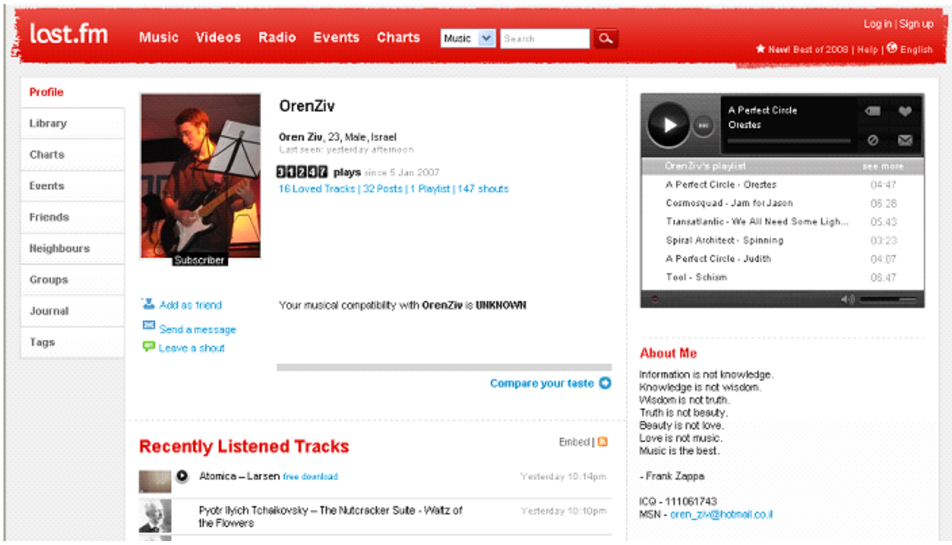

Figure 1. Last.fm Screen Shot (User Page)

- Improved content organization: capacity to listen to unlimited personal playlists on shuffle mode and to create a "Loved Tracks" radio channel.[7]

- 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.

Note that none of these "premium offers" changes the music consumption option. That is, there is no limitation on the content available to nonsubscribers. Similarly, the premium subscription does not change the functionality for community participation and leadership. That is, there are no features that are blocked to nonsubscribers, and nonsubscribers can participate in and lead social groups and contribute to blogs in exactly the same manner as subscribers.

We collected data about a random sample of 150,000 Last.fm users (subscribers and nonpaying users). The data for each user include music listening behavior, number of friends, community activity levels, and demographics. Table 4 details the data available for each user.

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 nonpaying users). For this data set, we omitted data on subscribers and used only data on nonpaying 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.[8] 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 data set consisted of 39,397 nonpaying users and 3,612 new subscribers. Some descriptive statistics for our data are presented in Table 5.

Last.fm's various social computing options can be sorted into a ladder of participation. In the context of music listening on Last.fm, content consumption is measured either by the total

[7]This is a playlist created by the site based on a user's tagging of songs as "loved."

[8]Our crawler collected the data from the "welcoming new subscribers" page once an hour. We have tested and found that the page is updated practically instantly after the subscription is paid for. Hence, the data collected reflect activity levels of the users before subscribing and up to an hour after subscribing.

Table 4. Data Description

Data Type	Description
Demographic Information	Age, gender, time since registration to the website
Music Consumption Information	Number of song plays, time since last play.
Content Organization Activities	Number of songs tagged, number of songs marked as "Loved," number of playlists created
Community Participation Activities	Number of group memberships, number of groups led, number of posts to groups and last.fm official forum, number of blog entries
Friends	Number of nonsubscriber friends, number of subscriber friends

Table 5. Descriptive Statistics

Type of Membership		Non-Paying User			Subscriber		
		Mean	Median	Variance	Mean	Median	Variance
Content Consumption	Song Plays	17616	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	Number of friends	14.56	9	640.923	21.19	10	1,196.87
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	335,075.6

Table 6. Comparing Activity Levels of Subscribers and Nonpaying Users

	Variable Name	Subscriber Mean	User Mean	Ratio	U-test *P* Value	*t*-Test *P* Value
Content Consumption	No. of song plays	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.

number of plays or by the average daily number of song plays. One rung above is content organization, which 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). The next level of engagement, community participation, 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. Finally, community leadership is measured by the number of groups led by a user.

The descriptive statistics clearly suggest that the usage pattern of subscribers is quite different from that of regular users. Table 6 summarizes the average volume of activity attributed to different rungs of the ladder of participation for paying subscribers and for nonpaying users. For each type of activity, the third column of Table 6 shows the ratio between subscriber activity level and user activity level. We used the *t*-test and the Mann-Whitney U-test to compare nonpaying users with subscribers, as the two populations are not normally distributed (Mann and Whitney 1947).

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

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

Moreover, paying subscribers have more friends listed on their pages. Table 7 shows that whereas the average nonpaying user has slightly more than 14 friends, the average subscriber has 21 friends, that is, subscribers have on average 45 percent more friends ($P < 0.01$). 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. 2009). Indeed, we find that paying subscribers have many more friends who are subscribers than nonpaying users do; the average subscriber has 2.82 subscriber friends, compared to only 0.42 subscriber friends for the average nonpaying user ($P < 0.01$).[9]

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

There are also demographic differences between subscribers and nonpaying 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 nonpaying users (see Table 5). 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 access to payment methods. Interestingly, we also find that subscribers make their subscription decisions after using the site for 652 days on average.[10] 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 percent of all users (paying and nonpaying) have listened to music, 77.6 percent have engaged in content organization behavior, 57.9 percent have participated in the community, and 5.2 percent have led a group, taking a leadership role in the community. Interestingly, only 8.7 percent of the users who have engaged in a community activity have 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 (binary) choice equation, predicting the probability of paying for a subscription.[11] Formally, we estimated the following block equation:

$$\begin{aligned} U_i(\mathit{Subscribe}) &= \alpha_0 + \alpha_1 \mathit{ContentConsumption}_i \\ &+ \sum_{j=1}^{J} \beta_{ij} \mathit{ContentOrganization}_i + \alpha_2 \mathit{FriendsCount}_i \\ &+ \alpha_3 \mathit{SubscriberFriendsCount}_i + \sum_{k=1}^{K} \gamma_{ik} \mathit{CommunityParticipation}_i \\ &+ \alpha_4 \mathit{CommunityLeadership}_i + \sum_{l=1}^{L} \delta_{il} \mathit{Demographics}_i + \varepsilon_i = V_i + \varepsilon_i \end{aligned}$$

[10]Given recent changes in the content industry, it is likely that the consumer's attitude to subscribing will change over time and that this time period will become shorter on average. However, our data seem to suggest that the paying consumer is well acquainted with the website before the subscription decision.

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

Table 7. Correlation Matrix

	Gender	Age	Days	Number of Friends	Number of Subscriber	Song Plays	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									
Song Plays	-.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 (two-tailed).

Content consumption is estimated using the total number of song plays (in thousands) to which user *i* listened. We also repeated the analysis using the average daily number of song plays to which user *i* listened with similar results.[12] 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.[13] 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

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

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 nonpaying 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 nonpaying 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 percent of the site population. If we used this cor-

[12]To clarify, the number of songs is the number of plays. That is, if a user listened to a song twice, it will be counted as two "tracks listened to."

[13]The goal of separation between total number of friends and subscriber friends is to capture possible peer effects in the subscription decision. See Bapna and Umyarov (2012) for a discussion of peer effects in the context of Last.fm.

rect ratio in composing our data set, 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, nonsubscribing user while still obtaining a 99 percent level of accuracy. To overcome the problem of misclassification, one should re-estimate the model while deliberately under-sampling the nonpaying 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 nonpaying 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; for a similar implementation, see Villanueva et al. 2008).

The correlation matrix is presented in Table 7, and the estimation results using the choice-based sample are reported in Table 8, each column representing an additional block being added to the estimation.

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, supporting H1.

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 (*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 has. 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 10 additional groups ($P = 0.02$). Hence, *H2(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 (*Odds Ratio* = 1.001 for each track marked as loved, and *Odds Ratio* = 1.184 for each playlist created). 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, *H2(b)* is only partially supported.

Content Consumption: As expected, content consumption has a positive association with the subscription decision, supporting *H1*. 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 playing over 10,000 more songs ($P < 0.01$). Similarly, being a member in one more group has an effect on the odds ratio equal to listening to 100,000 more songs ($P < 0.01$). These findings support *H2(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: As expected, 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 (Bapna and Umyarov 2012). 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 nonsubscribing 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.

Table 8. 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)	B (S.E)	B (S.E)	B (S.E)	B (S.E)	B (S.E)
	EXP(B)	EXP(B)	EXP(B)	EXP(B)	EXP(B)	EXP(B)
Number of song plays (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
Log Likelihood	15,025.902	14,096.893	14,053.363	11,755.238	11,728.094	10,812.496
Cox & Snell R-Square	.004	.073	.076	.226	.227	.280
Nagelkerke R-Square	.006	.103	.111	.339	.332	.408

Observations: 13,004. **Significant at the 0.05 level. ***Significant at the 0.01 level.

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. Figure 2 presents the consumption and participation patterns of different users as a function of time. Notably, in the first year of using the website, a user's music consumption decreases until it reaches a relatively stable level. However, there seems to be a consistent and stable increase over time in the likelihood of participating in the different content organization and community activities.

In what follows, we investigate the effect of content consumption, content organization, and community activity on the likelihood of consumers to purchase a paid subscription. We therefore estimate a hazard (survival) model, using the following equation:

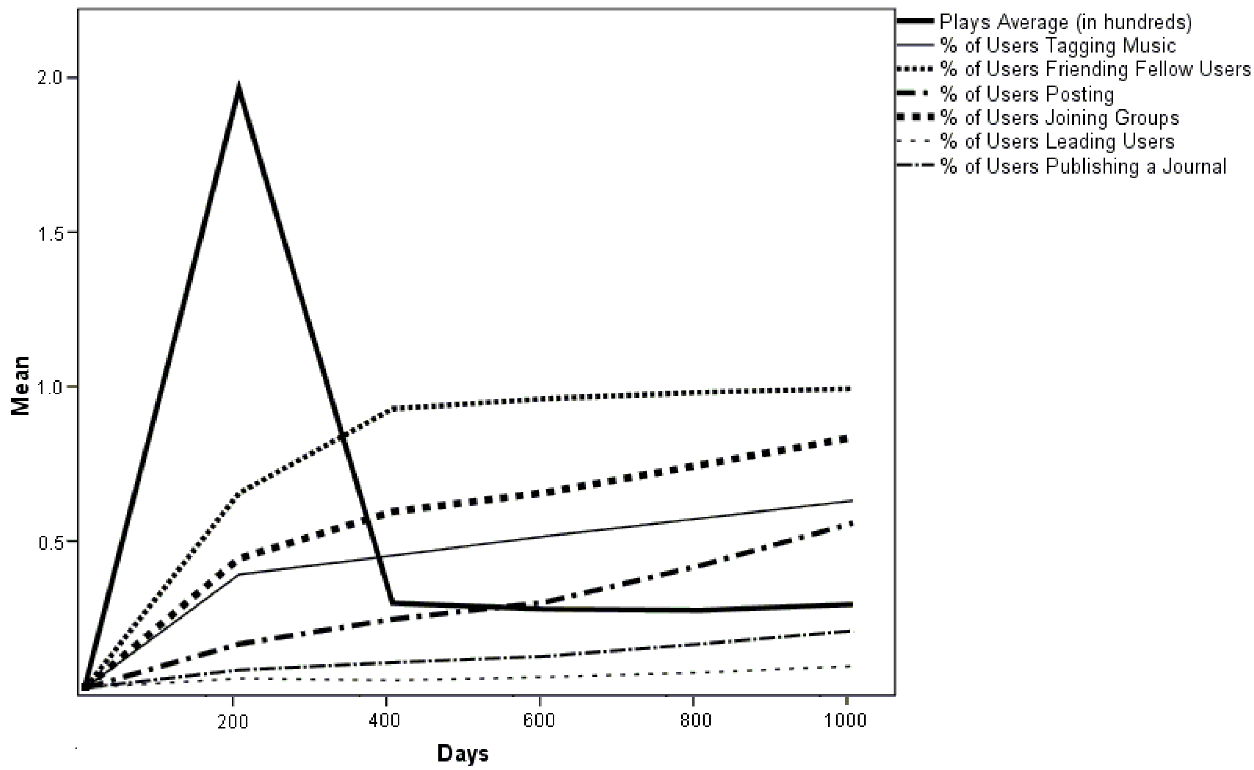

Figure 2. Content Consumption Levels and Usage of Social Computing Features over Time

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

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 9.

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 (supporting *H3*). Moreover, the strong and significant positive association between group leadership and the subscription decision again stands out, supporting *H4(c)*.

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.

Given the long period of time in question and the potential for exogenous changes in consumer taste, for robustness we repeated our analysis with a few subsamples of users who had joined the site more recently before subscription (subscribers who had been on the website less than 800, 600, and 400 days prior to subscription). The results are very similar, both in sign and magnitude. Note that as freemium models become more prevalent in the content industry, consumers may become more receptive to paying, and the time window to subscription may become shorter.

Propensity Score Matching

Although the preceding econometric analysis provides support for a positive and statistically significant association between

Table 9. Cox Regression Model for Subscribing Decision

		B	S.E.	Wald	df	Hazard Exp(B)
Content Consumption	Number of song plays (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	Groups joined	.002***	.001	7.134	1	1.002
	Groups led	.051**	.023	5.605	1	1.053
	Blog entries published	.017	.008	2.311	1	1.018
	Posts published	.000	.000	.002	1	1.000
Demographics	Age	.060***	.002	1211.184	1	1.062
	Gender	.169	.039	18.852	1	1.184

N (nonpaying 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

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. Although 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 nonexperimental data, based on observed variables.[14] 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 covariates, Z, such that $(y_1, y_0) \perp T|Z$, $pr(T = 1|Z) \in (0, 1)$, where $\perp$ denotes independence. That is,

[14] 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).

the treatment assignment is independent of the outcome conditional on a set of attributes Z. Moreover, if one is interested in estimating the average treatment effect, only the weaker condition, $E[y_0|T = 1, Z] = E[y_0|T = 0, Z] = EE[y_0|Z]$, $pr(T = 1|Z) \in (0, 1)$, 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 *four distinct binary treatments* 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 song plays, and the number of days on the Last.fm site), and the number of friends listed on the user's page.[15]

Consequently, we should match observations that have identical values for all variables included in Z. For example, in the case of the *GroupLead* treatment, we should match a 22-year-old male consumer who listened to 1,000 songs, has been using Last.fm for a year, and is a group leader, with another 22-year-old male who listened to 1,000 songs and has 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).[16] 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.

The results of our comparisons for each of the treatments are presented in Table 10. Column A in Table 10 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 percent 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 percent 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 10. Note that *Commu-*

[15]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.

[16]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.

Table 10. Propensity Score Matching

Treatment	A Group Membership	B Group Leadership	C Blog Entries	D Forum Postings	E Community Activity (Heckman)	F Community Activity (Mahalanobis)
Number of matched cases	29,941	2,423	6,097	16,375	30,882	30,882
Percentage of subscribers among treated cases	8.5%	15.2%	12.5%	10.0%	8.4%	8.4%
Percentage of subscriptions among nontreated cases	6.9%	98.%	9.8%	7.0%	7.0%	6.2%
Diff Mean	1.6%	5.4%	2.6%	3.0%	1.4%	2.2%
t-test (Diff Mean > 0)	7.38***	5.78***	4.78***	9.83***	6.61***	11.07***
Diff Mean (Std. Err.)	.002	.009	.005	.003	.001	.001
Std. Dev.	.37	.45	.43	.39	.25	.35
Rosenbaum upper bounds significant for Gamma (Γ)	1.5	1.7	1.5	1.8	1.5	

nityActivity was estimated twice, once using the kernel matching approach (column E) and once using the Mahalanobis matching approach (Column F). All of 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 they strengthen the findings of the binary logistic model. (A comparison of covariate means both before and after the matching are presented in Appendix A.)

Rosenbaum Bounds Sensitivity Analysis

Propensity-score matching operates on a strong assumption that observable characteristics fully account for the selection of users into the treatment and control conditions. However, there could still be hidden bias due to unobservable characteristics. We next conduct a sensitivity analysis by estimating Rosenbaum bounds (Rosenbaum 2002), which measure how strongly an unobservable must influence the selection process in order to completely nullify the causal effects identified in the propensity-matching analysis (for recent applications of this method, see Sen et al. 2012; Sun and Zhu 2010). If p_i denotes the subscription probability of a user who conducted community activity (i.e., a user in the treatment group), and p_j denotes the subscription probability of a user with no community activity (i.e., a user in the control group), Rosenbaum shows the following bounds on the odds ratio for the two matched users

$$\frac{1}{\Gamma} \leq \frac{p_i \,/\, (1 - p_i)}{p_j \,/\, (1 - p_j)} \leq \Gamma$$

where $\Gamma \geq 1$. Γ measures the level of selection effects from unobservable factors. When $\Gamma = 1$, users with the same propensity scores have the same probability of subscribing, and there are no unobserved selection effects. When $\Gamma > 1$, an unobservable causes the odds ratio of treatment assignment to differ between treatment and control groups. This method is based on the intuition that Γ should be close to 1 if the unobservable does not play a significant role in selection. Test statistics are developed to show how far away Γ has to be from 1 in order for the unobservable to nullify the treatment effect. This, of course, depends on the context of the research—if the Rosenbaum bound for unobserved selection (Γ) appears too large to be true in reality in the specific context, a researcher may conclude that the qualitative results from propensity matching hold.

We present the results of the Rosenbaum bounds analysis in the bottom row of Table 10.[17] We report the critical values of Γ at which the community activity effect becomes insignificant. Those values range between 1.4 and 1.8 and are similar to the findings of Sun and Zhu (2012) and DiPrete and Gangl (2004). In other words, an unobservable variable would have

[17]We only include this for the five binary treatments.

to change the odds of selection into the treatment group by at least 50 percent to nullify the effect of community activity on the subscription decision.

Discussion

Our empirical analysis supports our conjecture that users' levels of participation are linked to their willingness to pay for premium service. 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, and community leadership is more strongly associated with the likelihood of subscription than is mere community participation.

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 the factors 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. In line with previous research that links commitment to willingness to pay, we find that among such users, the presence of the affective community may be associated with monetary payments to the website. Analysts have noted that people report that they are not willing to pay for online content (Nielsen 2010); our observations suggest that 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 nonexperimental data. 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 data set 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 nonpaying 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. The additional Rosenbaum bounds sensitivity tests showed that an unobservable variable would have to change the odds of selection into the treatment group by at least 50 percent to nullify the effect of community activity on the subscription decision. 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.

Implications for Digital Business Strategy

This study proposes a new perspective on digital business strategy for the content industry in an age of social computing. Prior transitional processes of content digitization and net-enablement caused the content industry to move from offline to online platforms, where content providers now conduct most of their business. The social era we live in is bringing about new changes in business practices and models and is raising new questions that were not part of the discussion on net-enablement. For example, past research on net-enablement looked at methods of attracting customers by deploying net-based technologies and encouraging interaction between the firm and users (Straub and Watson 2001; Wheeler 2002). However, these studies did not take into account the role that technologies fulfill in facilitating and enhancing users' onsite relationships with other users. As social computing becomes increasingly widespread, these formed relationships are likely to become fundamental components of digital business strategies.

Moreover, prior discussions on business models for the content industry by academics and practitioners have focused mainly on the choice of revenue sources, frequently mentioning examples such as advertising, fixed subscriptions and paying for content items. This stresses a techno-centric approach that views social features as add-ons, enhancements to the core offerings. We suggest that social technologies should be fused to the business processes of content providers, whose role will be to provide interactive content consumption experiences, or *social content*. This fusion blurs the previously acknowledged dichotomy between the business models and strategies of content providers and those of virtual communities (for example, as presented in Weill and Vitale 2001).

We propose that content websites should take user participation dynamics into consideration and create platforms in which users can be encouraged to go beyond passive content consumption and climb up the ladder of participation. Such platforms should incorporate technical features that enable users to organize and curate content and form communities with fellow users, and that offer the ability to personalize the website's content and social outlook. Adopting social content as digital business strategy allows incumbents to differentiate themselves as providers of social experiences that incorporate different sets of activities and appeal to different sets of users. This would render firms less vulnerable to imitation in an age of increased competition in the content industry.

Still, the firm's digital business strategy should be aligned with the firm's core values and identity. This means that the social offer itself will have to take into account the characteristics of the user base, of the attributes and type of content offered by the specific provider, and the provider's ethics and values. For example, a music website that focuses on contemporary music might choose to implement a social experience that encourages the sharing and discovery of new music while helping users express their unique identities. This might mean relatively low restriction on the language and style of interaction. A traditional news website, on the other hand, might want to adopt a social experience that revolves around discussions of content and helps users to act as "news aggregators." This website might restrict user content generation to maintain the reliability of the news source and to maintain a more upscale style of interaction. Clearly, the choice of features will affect the nature of the resulting social environment in the website, which in turn will affect the segment of consumers drawn to the website, their valuation of the website, and their retention. This is not to say that one will be more "social" than the other, but that the kind of social environment induced by one's choice of features should align with the overall strategy of that website.

Hence, digital business strategy of content providers should align incentives for users to move up the rungs of the participation ladder and, in parallel, use the same participation ladder as a segmentation mechanism in order to capitalize on the different levels of participation. Existing strategies that focus on limiting the amount of content consumed before paying (such as NYTimes.com) or segmenting the content itself into free and premium (such as WSJ.com, the *Wall Street Journal* online) do not capitalize on the fusion of content with the social environment. In such cases, even if the content provider succeeds in creating a vibrant community, it still lacks a strategic approach for turning users' emergent patterns of participation into profits.

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 using social computing in content websites. 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. However, adding social components as merely add-ons would not necessarily yield a profit either.

Taken together, our results highlight the importance of creating a community environment that facilitates different levels of participation to create an ongoing and varied experience. 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 toward this high level of involvement, potentially increasing the chances that they will subscribe. A website owner should make 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 participate more?" The solution to this question may increase free-to-fee conversion rates.

While Last.fm is a good example for the inclusion of social features as an inherent part of the website experience, it seems that the website has not yet fully capitalized on it. First, to the best of our knowledge, Last.fm undertakes little initiative to encourage users to climb the ladder of participation. 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.

Second, while our results show a clear relation between social behavior and willingness to pay, Last.fm and other freemium websites currently choose not to base the premium offer on social "perks." Changing the premium subscriber benefits to reflect his social nature might improve conversion rates.

Another question of interest in this context is whether the model suggested in this paper is effective, given 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 percent. This number is in line with numbers reported by other websites, whose conversion rates are between 0.5 and 15 percent, 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 percent increase in the conversion rate, from 0.9 to 1.0 percent, 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 percent,[18] whereas the conversion rates reported by social networks (such as Facebook) are about 0.051 to 0.063 percent.[19] 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 percent, 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 percentage, users' willingness to pay for premium subscriptions is of great importance to this industry.

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

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

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 websites that provide 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.

In addition, we focused on a proprietary content website. 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.

We also focused on an on-site community. However, we do not have data on sites that implement their ladder of participation using external communities such as Facebook. It is possible that such websites could still capture value from users' commitment. Those extensions would both serve as interesting directions for future work.

Our study 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. In addition, we substantiated our conclusions by computing the Rosenbaum bound of unobservable effects. Nevertheless, there are still correlated unobservables, such as aversion to advertising, that should be handled in future work, perhaps using an experimental setting. Furthermore, 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.

Acknowledgments

We thank Ravi Bapna, Jacob Goldenberg, Vijay Gurbaxani, Arun Sundararajan, and Barak Libai, and seminar participants at the University of California, Irvine; New York University; Indiana University; the International Conference on Information Systems; the Marketing Science Conference; and the Workshop on Information in Networks for their feedback. We would also like to thank Oren Ziv for introducing us to Last.fm. Financial support from the Google Inter-University Center for Electronic Markets and Auctions, support of the Henry Crown Institute of Business Research and the Rothschild-Caesarea fund is gratefully acknowledged.

References

Adams, P. 2011. *Grouped: How Small Groups of Friends are the Key to Influence on the Social Web*, Berkeley, CA: New Riders Press.

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.

Banker, R. D., Bardhan, I. R., Chang, H., and Lin, S. 2006. "Plant Information Systems, Manufacturing Capabilities, and Plant Performance," *MIS Quarterly* (30:2), pp. 315-338.

Bateman, P. J., Gray, P. H., and Butler, B. S. 2011. "The Impact of Community Commitment on Participation in Online Communities," *Information Systems Research* (22:4), pp. 841-854.

Bapna, R., amd Umyarov, A. 2012. "Are Paid Subscriptions on Music Social Networks Contagious? A Randomized Field Experiment," SOBACO Working Paper, Carlson School of Management, University of Minnesota, Minneapolis, MN.

Beatty, S. E., and Kahle, L. R. 1988. "Alternative Hierarchies of the Attitude–Behavior Relationship: The Impact of Brand Commitment and Habit," *Journal of the Academy of Marketing Science* (16:2), pp. 1-10.

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.

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," in *Proceedings of the 27th Annual SIGCHI Conference on Human Factors in Computing Systems*, Boston, April 4-9.

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.

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), pp. 60-70.

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.

Dick, A. S., and Basu, K. 1994. "Customer Loyalty: Toward an Integrated Conceptual Framework," *Journal of the Academy of Marketing Science* (22:1), pp. 99-113.

DiPrete, T. A., and Gangl, M. 2004. "Assessing Bias in the Estimation of Causal Effects: Rosenbaum Bounds on Matching Estimators and Instrumental Variables Estimation with Imperfect Instruments," *Sociological Methodology* (24), pp. 271-310.

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.

El Sawy, O. A. 2003. "The 3 Faces of Information Systems Identity: Connection, Immersion and Fusion," *Communications of the AIS* (12), pp. 588-598.

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

Fornell, C. 1992. "A National Customer Satisfaction Barometer: The Swedish Experience," *Journal of Marketing* (56), pp. 6-21.

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*, Los Alamitos, CA: IEEE Computer Society Press.

Fullerton, G. 2003. "When Does Commitment Lead to Loyalty?," *Journal of Service Research* (5:4), pp. 333-344.

Fullerton, G. 2005. "The Service Quality–Loyalty Relationship in Retail Services: Does Commitment Matter?," *Journal of Retailing and Consumer Services* (12:2), pp. 99-111.

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

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," *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.

Henderson, J. C., and Venkatraman, N. 1993. "Strategic Alignment: Leveraging Information Technology for Transforming Organizations," *IBM Systems Journal* (32:1), pp. 4-16.

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.

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

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

Lampe, C., and Johnston, E. 2005. "Follow the (Slash) Dot: Effects of Feedback on New Members in an Online Community," in *Proceedings of the 2005 International 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.

Lu, Y., and Ramamurthy, K. 2011. "Understanding the Link Between Information Technology Capability and Organizational Agility: An Empirical Examination," *MIS Quarterly* (35:4), pp. 931-954.

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.

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.

Neuberger, C., and Nuernbergk, C. 2010. "Competition, Complementarity or Integration?," *Journalism Practice* (4:3), pp. 319-332.

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

Nielsen. 2012. "Nielsen SoundScan Website" (http://nielsen.com/us/en/industries/media-entertainment.html).

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, NY: ACM Press, 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.reillynet.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.

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.

Pew Research Center. 2010. "Understanding the Participatory News Consumer," (http://pewresearch.org/pubs/1508/internet-cell-phone-users-news-social-experience).

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.

Posner, R. A. 2005. "Bad News," *New York Times*, July 31 (http://www.nytimes.com/2005/07/31/books/review/31POSNER.html?pagewanted=print).

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

Pritchard, M. P., and Howard, D. R. 1997. "The Loyal Traveler: Examining a Typology of Service Patronage," *Journal of Travel Research* (35:4), pp. 2-10.

Raju, J., Srinivasan, S. V., and Lal, R. 1990. "The Effects of Brand Loyalty on Competitive Price Promotional Strategies," *Management Science* (36:3), pp. 276-304.

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. 2002. *Design of Observational Studies*, New York: Springer.

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.

Sambamurthy, V., Bharadwaj, A., and Grover, V. 2003. "Shaping Agility through Digital Options: Reconceptualizaing the Role of Information Technology in Contemporary Firms," *MIS Quarterly* (27:2), pp. 237-263.

Sen, B., Shin, J., and Sudhir, K. 2011. "Demand Externalities from Co-Location: Evidence from a Natural Experiment," Working Paper, Yale University, New Haven, CT.

Shankar, V., Rangaswamy, A., and Pusateri, M. 1999. "The Online Medium and Customer Price Sensitivity," eBusiness Research Center Working Paper 04-1999, Pennylvania State University, College Park, PA.

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.

Straub, D., and Watson, R. 2001. "Transformational Issues in Researching IS and Net-Enabled Organizations," *Information Systems Research* (12:4), pp. 337-345.

Sun., M., and Zhu, F. 2012. "Ad Revenue and Content Commercialization: Evidence from Blogs," Working Paper (available at http://dx.doi.org/10.2139/ssrn.1735696).

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

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.

Weill. P., and Vitale, M. R. 2001. *Place to Space*, Boston: Harvard Business School Press.

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

Wheeler , B. C. 2002. "NEBIC: A Dynamic Capabilities Theory for Assessing Net-Enablement," *Information Systems Research* (12:4), pp. 337-345.

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

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. 2010. "Computing in Everyday Life: A Call for Research on Experiential Computing," *MIS Quarterly* (34:2), pp. 213-231.

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.

Zeithaml, V. A., Berry, L. L., and Parasuraman, A. 1996. "The Behavioral Consequences of Service Quality," *Journal of Marketing* (60), pp. 31-46.

About the Authors

Gal Oestreicher-Singer is an assistant professor at Tel Aviv University's Recanati Business School. Her research studies the effects of visible networks on electronic markets and the economics of digital rights management. Her prior research has been published at leading journals including *Management Science*, *Information Systems Research*, and *Journal of Marketing Research*. She received the 2008 ACM SIGMIS Best Dissertation Award, a European Union Marie Curie Early Career Award, an INFORMS CIST Best Paper Award, an ICIS Best Overall Paper award, a MSI-WIMI User Generated Content Research Competition Award, and the Google-WPP Marketing Award. She received her Ph.D. from New York University in 2008, and holds degrees in law and electrical engineering from the Hebrew University in Jerusalem and Tel Aviv University.

Lior Zalmanson is a doctoral student at Tel Aviv University's Recanati Business School. He has a B.Sc. in computer science and an M.Sc. in information systems and technology management from Tel Aviv University. His experience includes the management of social computing and Web 2.0 initiatives in a large governmental organization in Israel. His research interests include online communities and economic behavior in virtual environments, in particular Internet business models and pricing of digital goods.

Appendix A

Comparison of Means Before and After Propensity Score Matching

	Treatment 1: Group Membership			Treatment 2: Group Leadership		
	Treatment 1	Control		Treatment 2	Control	
		Pre	Post		Pre	Post
No. of song plays	21,644.90	9961.55	23,247.00	34737.95	16,957.24	37,944.00
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	.10	.20	1.88	.37	.84
No. of groups led	.12	0	0	1.41	0	0
Users' age	23.11	24.70	23.00	23.16	23.64	22.44
Users' gender	0.31	0.36	0.39	0.20	0.34	0.32
Days since joining the website	761.61	702.78	699.05	921.56	701.82	814.94

*Note that for the treatment group there is no difference in the means before versus after the matching process.

	Treatment 3: Group Postings			Treatment 4: Journal Postings		
	Treatment 3	Control		Treatment 4	Control	
		Pre	Post		Pre	Post
No. of song plays	27,036.60	12,377.88	27,901.00	28,537.90	16,211.60	31,860.00
No. of playlists	.87	.77	.85	.97	.78	.94
No. of loved tracks	113.51	56.32	112.40	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	26.90
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	.10
Users' age	22.84	24.09	22.62	23.59	23.62	23.27
Users' gender	0.30	0.35	0.36	0.33	0.33	0.33
Days since joining the website	831.61	638.35	724.24	842.87	692.83	781.06

*Note that for the treatment group there is no difference in the means before versus 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 song plays	21,429.62	9,120.73	23,604.00	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	2.16	24.76	23.05
Users' gender	0.32	0.36	0.40	0.32	0.36	0.41
Days since joining the website	760.05	587.87	675.90	760.05	586.87	678.27

*Note that for the treatment group there is no difference in the means before versus after the matching process.