View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by AIS Electronic Library (AISeL)

Association for Information Systems
AIS Electronic Library (AISeL)

MCIS 2010 Proceedings

Mediterranean Conference on Information Systems (MCIS)

9-2010

WHO PAYS "PREMIUM" IN THE AGE OF FREE SERVICES? FINDINGS FROM A MEDIA WEBSITE

Gal Oestreicher - Singer,
Tel Aviv University, Israel, galos@post.tau.ac.il

Lior Zalmanson
Tel Aviv University, Israel, zalmanso@post.tau.ac.il

Follow this and additional works at: http://aisel.aisnet.org/mcis2010

Recommended Citation

Oestreicher - Singer,, Gal and Zalmanson, Lior, "WHO PAYS "PREMIUM" IN THE AGE OF FREE SERVICES? FINDINGS FROM A MEDIA WEBSITE" (2010). *MCIS 2010 Proceedings*. 65.
http://aisel.aisnet.org/mcis2010/65

This material is brought to you by the Mediterranean Conference on Information Systems (MCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in MCIS 2010 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

WHO PAYS "PREMIUM" IN THE AGE OF FREE SERVICES? FINDINGS FROM A MEDIA WEBSITE

Gal Oestreicher - Singer, Tel Aviv University, Israel, galos@post.tau.ac.il
Lior Zalmanson, Tel Aviv University, Israel, zalmanso@post.tau.ac.il

Abstract

The challenge for many media websites is converting users from free to fee. In order to encourage user participation and engagement with the websites many of them have provided consumers with a virtual community wherein the user can create an on-site identity, make friends, and interact with other consumers.

We study the interplay between users' functional and social behavior on media sites and their willingness to pay for premium services. We use data from Last.fm, a site offering both music consumption and social networking features. The basic use of Last.fm is free and premium services are provided for a fixed subscription fee. While the premium services mainly improve the content consumption experience, we find that willingness to pay for premium services is strongly associated with the level of social activity of the user, and specifically, the community activity of the user. Our results represent new evidence of the importance of introducing community and social activities as drivers for consumers' willingness to pay for online services.

Keywords: Electronic Commerce, Social Media, Business Models, Propensity Score Matching

1 BACKGROUND AND OVERVIEW

Academic scholars and practitioners have noted that digital media companies find it difficult to charge their users for access to content services (Clemons et al. 2003, Srinivasan et al. 2002, *inter alia*). Media Websites are now encouraging user participation and engagement, for example, by allowing users to post comments to news stories (talkbacks). Many sites that enable users to contribute content also provide consumers with a virtual community, wherein the user can create an on-site identity (often by having a personal page), make online friends, attend virtual social events, build a reputation, and interact with other consumers. These 'extras' render the user's consumption experience increasingly interactive and social.

This interactive and social model of online content consumption brings with it new challenges for site owners and users. By encouraging users to contribute, site owners lose some of their control over the content that consumers experience, particularly in cases where owners cannot eliminate negative reviews or delete uninteresting or offensive posts. Correspondingly, the consumers themselves have greater influence on their fellow consumers' consumption experience. Despite this, many site owners encourage user participation because it can add interesting content that other consumers find valuable.

In this paper, we conjecture that there is a less obvious yet important effect of virtual socialization that is facilitated by offering user-generated content and developing a community on one's site. It is likely that in addition to benefiting other consumers, the act of participation positively affects the experience of the contributing consumer. By contributing content and becoming active in the site's social community, the consumer is likely to feel more involved with the site. This involvement might lead to increased brand loyalty, deceased churn, lower defection to competing sites, and more willingness to pay for (additional) premium services.

We investigate the interplay between users' functional behavior (content consumption) and their social behavior on media sites, as well their willingness to pay for premium services. We focus on websites that combine structured content (in this case, music tracks owned by commercial labels) with an open social arena in which users can add content such as comments, reviews, and 'tags'.

We divide consumers' use of such sites into three groups of activities:

- Functional use, which includes content consumption as well as all activities entailed in content organization.
- Local social network activities, which include on-site interaction with one's friends.
- Community (or global social network) activities, which include publishing user-generated content that can be consumed by the entire site audience, memberships to discussion groups, or comment posting.

Our research questions are as follows:

1. Are consumers who use social networking features in media websites more likely to pay for premium services?
2. If so, what is the marginal effect of local social network activities versus global (community wide) activities on the propensity to pay for those services?

We use data from Last.fm, a media site that serves both as an online radio and as a social networking site. Similar to other media websites, Last.fm allows users to access a set of basic services for free, and provides additional premium services in exchange for a fixed monthly subscription fee. Even though the premium services mainly improve the content consumption experience (for example, by increasing bandwidth), we find that willingness to pay for premium services is strongly associated with the level of social activity of the user. Specifically, consumers who use global social network features (i.e., features that enable the user to publish content and to engage with the entire network) show a higher propensity to pay for premium services compared with users who do not use these features. Our results represent new evidence of the importance of introducing community and social activities as a means of driving consumers' willingness to pay for online services. To the best of our knowledge, this study is the first to examine the influence of social involvement on consumers' decisions to purchase premium services.

Our work adds to two branches of literature: that on willingness to pay for online services, and that on the economic effects of a brand community on online businesses.

Many media sites operate under a two-tiered business model, wherein basic services are provided for free, and premium services are offered for a fee (Picard 2000; Riggins 2003). This business model has received wide attention from the press — including the coining of the term "freemium business model" by Fred Wilson [1]—. Convincing users to switch to a for-pay service is the main challenge of the two-tiered business model. Naturally, providing better content or service encourages users to subscribe to premium services (Ye et al. 2004). However, a user's choice might be influenced by his or her level of engagement in the site's virtual community.

Brand communities are defined as online communities built around commercialized products or shared services. Studies have shown that a user's participation in a community that is linked to a brand can increase strong and lasting bonds with that brand and promote brand loyalty, both in the offline and online context (Mael & Ashforth 1992 in the context of offline communities; McAlexander et al. 2002 and Jang et al. 2008 in the context of online communities).

One of the dimensions of brand loyalty is the consumer's willingness to repurchase (Aaker 1991). Loyal customers have lower price elasticity than do nonloyal customers, and they are willing to pay a premium to continue doing business with their preferred retailers (Reichheld and Sasser 1990). In the e-commerce context, Srinivasan et al. (2002) surveyed 1,211 online customers and identified the existence of an online community as one of eight factors significantly influencing brand loyalty and willingness to purchase in online stores. Our work adds to this literature by providing empirical evidence of the effect of social activity on consumers' willingness to pay for online services in media and content websites.

More broadly, our work also adds to the growing literature surveying the effects of social networks on consumption patterns. Marketing literature has long acknowledged the importance of social networks on the diffusion and adoption of new products and services (see Nair et al. 2006 for a detailed survey of the literature on social effects in marketing). Researchers have also attempted to separate social effects from marketing effects, thus requiring the identification of differing social effects (Trusov et al. 2007; Goh et al. 2008). Recently, researchers have focused on separating between local and global network effects when examining the influence of social factors on the adoption decision (for example, see Tucker 2004 on the adoption of a video messaging system in an organization). However, those works study the diffusion of products for which network effects are an inherent characteristic, such as communication technologies. Our work adds to this literature by emphasizing the importance of introducing local and global social networking features even to websites that offer traditional (professionally generated) content.

The rest of this paper is organized as follows: Section 2 provides an overview of the data and methodology. Section 3 presents the results and discussion; and section 4 concludes.

2 OVERVIEW OF DATA

We collected data from Last.fm, a social media site in which users can listen to music online and create personalized 'radio stations', or playlists. Last.fm also offers its users a social community. Currently, Last.fm has more than 30 million registered users based in more than 200 countries. While the site's main goal is to provide music listening capabilities, it also enables the user to create a personal user profile page, join groups (mostly based on musical taste), contribute to blogs (journals) by posting comments, or to take a lead role in those groups and journals. Users can also add tags to artists, albums, and tracks by using chosen keywords.

Last.fm offers 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's functions. The second is the paid subscription, in which subscribers pay a monthly fee of €2.5 for a package of premium services that include the following:

---

[1] http://www.avc.com/a_vc/2006/03/the_freemium_bu.html

- Improved infrastructure, including removal of ads from the subscriber's page and top-priority quality-of-service on web and radio servers.
- Extended listening options, including the capacity to listen to unlimited personal playlists on shuffle mode, and to create a 'Loved Tracks' radio channel.
- Improved social status, including an icon added to a subscriber's account and the ability to see who has visited one's profile page.

2.1 Data Collection and Preparation

We collected the following data on Last.fm users:

- Demographic information such as age and gender.
- Music consumption information such as number of tracks listened to; number of tracks tagged as 'Loved'; number of user-generated playlists; and time since last visit
- Virtual community activity information such as number of friends; number of blog (journal) posts; number of group memberships; number of groups led; number of user postings to the site's groups

Type Of Membership:	Non paying user			Subscribers		
	Mean	Median	Variance	Mean	Median	Variance
Age	23.08	21	39.156	29.43	27	88.415
Gender (1= Male, 2= Female)	1.34	1	0.223	1.29	1	0.204
Tracks listened to	17,616.99	11,265.00	477,622,677.54	21,688.83	11,039.50	998,060,194.11
Playlists created	0.77	1	0.47	1.29	1	7.15
'Loved' tracks tagged	65.97	11	41,872.72	210.34	83	314,062.36
Tags created	9	1	1,400.19	21.27	2	5,298.45
No. of friends	14.56	9	640.923	21.19	10	1,196.87
Posts published	9.12	0	7,596.37	27.31	0	75,401.53
Groups joined	5.27	2	168.69	8.98	3	463.08
Groups led	0.07	0	0.165	0.17	0	0.452
Journal entries published	0.42	0	2.244	0.89	0	5.623

Table 1. Descriptive Statistics

We collected these data using two specially programmed web crawlers. One web crawler gathered information about a random sample of 150,000 Last.fm users (subscribers and non-paying users). For this dataset, we omitted data on subscribers and used only data on non-paying users. A second web crawler collected information about new paying subscribers at the time that they purchased their subscriptions. We were able to access this set of users thanks to a continually updated list of recent subscribers that is featured on Last.fm. 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 10,000 new subscribers.

Data collection was done over a period spanning 3 months starting in January 2009. In order to omit inactive users from our analysis, we removed data on users who had not visited the site during the 3 months prior to data collection. We also omitted users and subscribers who had in the past used a "Reset" option that reset the logs of their personal site usage. Our final dataset consisted of 39,397 non-paying users and 3,612 new subscribers. Some descriptive statistics about our data are presented in Table 1.

3 DATA ANALYSIS AND RESULTS

The descriptive statistics clearly suggest that the usage pattern of subscribers is quite different from that of regular users. Table 2 summarizes the average activity levels of the consumers in our sample, which we divided into (paying) subscribers and (non-paying) users. For each type of activity, the third column of Table 2 shows the ratio between subscriber activity level and user activity level. To test whether the activity levels of the two populations are sufficiently distinct, a t-test would normally be in order. However in this case, the populations are not normally distributed and as such do not obey the assumption of the independent samples t-test. Therefore we used the Mann-Whitney U-test, where $P < 0.05$ shows that the two populations' medians and means are distinct.

We observe that subscribers consume 23% more music than do their non-paying peers; this difference is not statistically significant, however (Mann-Whitney with $P = 0.427$). Interestingly, subscribers invest significantly more in organizing their pages. On average, subscribers create 67% more playlists on their sites; they choose to tag 218% more tracks as 'Loved'; and create 140% more tags ($P < 0.01$). Since the tags and playlists are available on one's page, it is not clear whether these activities are motivated by the increased level of music consumption, or should be treated as social activities.

Moreover, we observed differences when we compared the social activity levels of subscribers with those of non-paying users. Our measure of local social network activity is the number of friends listed on one's page. In Table 2, one can see that while regular users have an average of 14 friends, subscribers have an average of 21 friends, i.e., subscribers have on average 45% more friends ($P < 0.01$). Most intriguingly, subscribers are substantially more involved in the site's virtual social community: compared with nonpaying users, paying subscribers post 199% more posts on the site's forums, join 70% more groups, lead on average 142% more groups, and publish 111% more blog entries ($P < 0.01$). A possible explanation for the evident differences in activity levels might be demographic differences between subscribers and non-paying users. The two demographic variables we obtained were gender and age. We did not observe a significant difference in activity levels or in propensity to subscribe based on gender. We did, however, find that subscribers are on average 6 years older than non-paying users (see Table 1).

U-test P Value	Ratio	User mean	Subscriber mean	
0.427	1.23	17,616.99	21,688.83	No. of tracks listened to
0.00***	1.45	14.56	21.19	No. of friends
0.00***	1.67	0.77	1.29	No. of playlists
0.00***	3.18	65.97	210.34	No. of Loved tracks
0.00***	2.40	9	21.27	No. of tags created
0.00***	2.11	0.42	0.89	No. of journals / blog entries
0.00***	2.99	9.12	27.31	No. of posts
0.00***	1.70	5.27	8.98	No. of group memberships
0.00***	2.42	0.07	0.17	No. of groups led
0.00***	1.27	23.08	29.43	Users' age
0.00***	1.10	720.53	652.08	Days of use
*** - Significant at the 0.01 level				

Table 2. Comparing Subscribers to Non-Paying Users

3.1 Model Estimation

To better understand the interplay between music consumption, local social activity, social involvement in the site's social community, and willingness to pay for a subscription, we estimate a logistic (binary) choice equation, predicting the probability of paying for a subscription. Formally, we estimated the model:

$$\log \frac{\Pr(Subscribe)}{1-\Pr(Subscribe)} =$$

$$\alpha_0 + \alpha_1 * Age + \alpha_2 * TracksDivThousand + \alpha_3 * PlayListCnt + \alpha_4 * LovedTracksCnt$$

$$+ \alpha_5 * FriendsCnt + \alpha_6 * GroupCnt + \alpha_7 * GroupLeadCnt + \alpha_8 * JournalCnt$$

Note that by controlling for the music consumption characteristics of the user, we are able to measure and quantify the marginal contribution of the social activity levels to the propensity to pay for premium services.Estimating this model presented us with two econometric challenges:

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

Second, when we looked at the random set of users on whom we collected information, we noticed that subscribers made up only 0.89% of the site population. If we used this correct ratio in composing our dataset, the occurrence of ones in our dependent variable (*Subscribe*) would be a *rare event*. The biases that rare events create in estimating logit models have been discussed in the literature (Ben-Akiva and Lerman 1985). In a nutshell, this poses a problem when estimating a logit model in that the model would predict that everyone would be a regular, non-subscribing user while still obtaining a 99% level of accuracy. To overcome the problem of misclassification, one should re-estimate the model while deliberately under-sampling the non-paying users, so that a more balanced sample of ones and zeros in the dependent variable is obtained. This sampling technique is called *choice-based sampling* (Ben-Akiva and Lerman 1985). To this end, we used our collected set of 3,612 new subscribers and only 5,000 non-paying users. However, using choice-based sampling leads to inconsistent intercept estimation when traditional maximum likelihood methods are used. Two alternative solutions have been suggested in the literature: Manski and Lerman (1977) developed a weighted endogenous sampling maximum likelihood (WESML) estimator, which accounts for the different weights in the zeros and ones from the population of interest. However, this estimator has the undesirable property of increasing the standard errors of the estimates (Manski and Lerman 1977; Greene 2000). A second approach, which we follow, is to adjust the estimated intercepts for each alternative by subtracting from the exogenous maximum likelihood estimates of the intercept the constant $Ln(S_i/P_i)$, where S_i is the percentage of observations for alternative i in the sample, and P_i is the percentage of observations for alternative i in the population (Manski and Lerman 1977; see Villanueva et al. 2008 for a similar implementation).

The correlation matrix is presented in Table 3 and the estimation results using the choice-based sample are reported in Table 4[2]. The odds of a user subscription decision are positively associated with the number of (thousands) of tracks the user listens to (*Odds Ratio = 1.003*). We also find that content organizing activities, such as creating a playlist and tagging music tracks as 'Loved', are positively correlated with the subscription behavior (*Odds Ratio =1.245* for *PlaylistCnt* and *Odds Ratio = 1.002* for *LovedTracksCnt*). However, this is understandable given that a premium service subscription gives users extra playlist listening capabilities and the possibility to listen to "loved tracks" as if they were a

---

[2] The equation includes only the coefficients in the regression that are statistically significant. The Tags (*TagsCnt*) and Postings (*PostsCnt*) are not found to be significant predictors of a user's subscription decision.

"radio station". It is therefore natural to assume that heavy users of those features will be more inclined to pay for premium services.

	Gender	Age	Num. Of Friends	Tracks Listened	Playlist Created	Loved Tracks Tagged	Posts Pub.	Groups Joined	Groups Led	Journal Entries Written	Tags Created
Gender	1.000	-.181**	.053**	-.097**	.023**	.005	-.015**	-.025**	-.051**	.000	-.035**
Age	-.181**	1.000	-.067**	-.057**	.101**	.097**	.004	-.057**	-.008	.019**	.041**
Number Of Friends	.053**	-.067**	1.000	.289**	.094**	.194**	.111**	.310**	.184**	.219**	.126**
Tracks Listened To	-.097**	-.057**	.289**	1.000	.042**	.130**	.127**	.216**	.164**	.212**	.119**
Playlist Created	.023**	.101**	.094**	.042**	1.000	.269**	.014**	.066**	.025**	.069**	.100**
Loved Tracks Tagged	.005	.097**	.194**	.130**	.269**	1.000	.070**	.183**	.064**	.123**	.209**
Posts Published	-.015**	.004	.111**	.127**	.014**	.070**	1.000	.195**	.194**	.159**	.102**
Groups Joined	-.025**	-.057**	.310**	.216**	.066**	.183**	.195**	1.000	.370**	.233**	.219**
Groups Led	-.051**	-.008	.184**	.164**	.025**	.064**	.194**	.370**	1.000	.223**	.166**
Journal Entries Written	.000	.019**	.219**	.212**	.069**	.123**	.159**	.233**	.223**	1.000	.180**
Tags Created	-.035**	.041**	.126**	.119**	.100**	.209**	.102**	.219**	.166**	.180**	1.000

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

Table 3. Correlation Matrix

Interestingly, we find that after controlling for content consumption and the use of content organization features (the activities that are most enhanced by premium services), the number of friends the user has listed on his or her page (i.e., the user's level of local social network activity) is positively associated with the user's propensity to pay for premium services (*Odds Ratio = 1.002*).
Within the community-wide activities, writing a blog (journal) entry is positively associated with the subscription decision. Similarly, joining a group or leading a group are associated with significant increases in the odds of subscribing to premium services (*Odds Ratio = 1.047* for *JournalCnt*; *Odds Ratio = 1.004* for *GroupCnt* and *Odds Ratio = 1.432* for *GroupLeadspCnt*). These results are especially interesting, given that the premium services provided to subscribers generally relate to music consumption and not to other forms of interaction on the site.

	B	S.E.	Wald	df	Sig.	Exp(B)
Age	0.112	0.004	877.053	1	0.000***	1.118
TracksDiv1000	0.003	0.001	7.824	1	0.005***	1.003
PlaylistCnt	0.219	0.029	56.185	1	0.000***	1.245
LovedTracksCnt	0.002	0.000	177.530	1	0.000***	1.002
TagsCnt	0.000	0.001	0.177	1	0.674	1.000
FriendsCnt	0.002	0.001	5.897	1	0.015**	1.002
PostsCnt	0.000	0.000	2.017	1	0.156***	1.000
GroupCnt	0.004	0.002	5.048	1	0.025**	1.004
GroupLeadsCnt	0.359	0.067	28.682	1	0.000***	1.432
JournalCnt	0.046	0.015	9.524	1	0.002***	1.047
Constant	-3.820	0.106	1,301.040	1	0.000***	0.022
Revised Constant	-8.20	After estimated intercept adjustment				
N (non-paying users) = 5,000, N (subscribers) = 3,612 Overall Model Estimation: chi-square = 2,108.086. *df* = 10, *p* = 0.00 -2 Log likelihood = 9,605.997, Cox & Snell R Square = 0.217, Nagelkerke R Square = 0.292 **- significant at the 0.05 level ; ***- significant at the 0.01 level						

Table 4. Binary Logistic Regression Model for Subscribing Decision

Our findings seem to indicate that social activity has an important role in subscription behavior. This can also be seen from Table 5: the model correctly predicts 67.4% of the non-paying users and 75.9% of the subscribers.

Observed		Predicted by Membership Type		
		Non-paying	Subscribers	% correct
Membership type	Non-paying	3,370	1,630	67.4
	Subscribers	872	2,740	75.9
	Overall %			70.9

Table 5. Predicted Values of Logit Model

3.2 Propensity Score Matching

Although the preceding econometric analysis provides support for a positive and statistically significant association between social online activity and propensity to purchase a premium services 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, like most other papers on the topic of brand community, 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 score in the marketing context, see Mithas and Krishnan 2008). The fundamental problem in identifying treatment effects is one of incomplete information. Though we observe whether the treatment occurs and whether the outcome is conditional on the treatment assignment, the counterfactual is not observed. In a nutshell, propensity matching techniques enable us to investigate heterogeneous treatment effects in non-experimental data, based on observed variables[3]. 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 contribute to the website's community) to a sample of untreated observations (in our context, users who did not contribute to the community) matched on the propensity of being treated (that is, the propensity to contribute).

Mathematically, Let $y_{i,1}$ denote the outcome of observation i, if the treatment occurs (given by $T_i=1$), and $y_{i,o}$ 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 are observed for each observation, unless assignment into the treatment group is random, generally, $\tau \neq y_1 - y_0$.

Propensity score matching attempts to overcome this problem by finding a vector of covariance, Z, such that $(y_1, y_0) \perp T | Z,\ pr(T=1|Z) \in (0,1)$, where $\perp$ denotes independence. That is, the treatment assignment is independent of the outcome conditional on a set of attributes Z. Moreover, if one is interested in estimating the average treatment effect, only the weaker condition $E[y_0|T=1,Z] = E[y_0|T=0,Z] = EE[y_0|Z],\quad pr(T=1|Z) \in (0,1)$, 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;
- *BlogPost,* 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;
- *GroupPost,* which is equal to one if the user has ever posted an entry to a group page.

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

Consequently, we should match observations that have identical values for all variables included in *Z*. For example, in the case of *GroupLead* treatment, we should match a 22-year-old male consumer who listened to 1000 tracks, had been using LastFM for a year, and is a group leader, with another 22-year-old male who listened to 1000 tracks and had been using LastFM 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 logit model. We thereafter match consumers not according to their exact attributes but according to their propensity score. One of the advantages of propensity score methods is that they easily accommodate a large number of control variables.

---

[3] 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).

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). All treated individuals were matched with untreated individuals with the nearest propensity scores. We were then able to compare the percentage of subscribers within the treated and the matched untreated groups.

The results of our comparison for each of the treatments are presented in Table 6. Column A on Table 6 corresponds to the case where the treatment is defined as *Group_Membership*. In this case each consumer with group membership is matched with a consumer without group membership according to the above-mentioned covariates (including demographics, music listening, and local social activity). Out of the 29,941 consumers with group membership, 8.5% were found to have a subscription. However, out of the 29,941 consumers that were matched to those consumers (but were not group members) only 6.9% had a subscription. Since this difference is statistically significant ($P < 0.001$), we are able to conclude that, controlling for the observed differences between the groups, consumer who are group members are more likely to pay for a premium subscription. Similar analysis for the other three treatments (group leadership, group posting, and blog posting) is presented in columns B, C, and D of Table 6 and provides similar conclusions. After controlling for self-selection bias based on demographics, music consumption, and local social activities, we observe a significant difference between the treated and untreated conditions in the mean percentage of users who subscribe to premium services. That is, we show that consumers who contribute to the community, such as group leaders, group members, and blog writers, have a higher propensity to subscribe to premium services[4]. Moreover, one could consider leading a group to be a variable that represents a higher level of engagement with the site's community (compared with group membership or journal postings). Indeed, both in our logistic regression estimation and in our propensity score analysis, we see a strong correlation between group leading and subscription behavior.

Treatment 4: Blog Postings	Treatment 3: Group Postings	Treatment 2: Group Leadership	Treatment 1: Group Membership	Treatment
6,097	16,375	2,423	29,941	Number of Matched Cases
12.5%	10%	15.2%	8.5%	Percentage of subscription among treated cases
9.8%	7%	9.8%	6.9%	Percentage of subscription among non-treated cases
2.6%	3.0%	5.4%	1.6%	Diff Mean
4.79^{***}	9.83^{***}	5.78^{***}	7.38^{***}	T test (Diff Mean > 0)
.005	.003	.009	.002	Diff Mean (Std. Err)
.43	.39	.45	.37	Std.Dev

Table 6. Propensity Score Analysis

---

[4] We observe statistically significant differences for all treatments.

4 CONCLUDING REMARKS

Our paper emphasizes an important and yet somewhat overlooked role of social activity on websites that provide traditional content. We show an association between community activity and the propensity to pay for premium services. We show that after accounting for content consumption and demographics, both the use of local social network activity features and the use of global network (community wide) activity features are associated with a substantial increase in the probability of paying for premium services.

We extend those results by using propensity score matching, which has been shown to estimate treatment effects from non-experimental data. Through these matching techniques, we provide additional support to our findings. Although we do not control for unobserved heterogeneity in treatment assignment, propensity score matching allows us to control for self-selection bias based on consumption patterns, demographics, and social activity levels and to show that the use of global network features increases users' willingness to pay for premium services.

This study makes an important contribution to the literature of virtual communities and social networks and their influence on electronic commerce. It also provides researchers as well as practitioners with insights into the importance of adding social activities and building virtual communities as part of the media website

References

Aaker, D. (1991). Managing Brand Equity: Capitalizing on the Value of a Brand Name, New York: Free Press.

Ben-Akiva, M. and Lerman, S.R. (1985). Discrete Choice Models, MIT Press.

Clemons, E. K. Gu, B. and Lang, K. R. (2003). Newly vulnerable markets in an age of pure information products: An analysis of online music and online news. Journal of Management Information Systems 19(3), 17-42.

Coupey, E. (2001). Marketing and the Internet, Upper Saddle River, N.J: Prentice-Hall.

Goh, K., Hui, K., and Png, I.P.L. (2008). Social Interaction, Observational Learning and Privacy: The "Do Not Call" Registry. working paper.

Greene, W. H. (2000). Econometric Analysis, 4th ed. Upper Saddle River, N.J: 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), 441-462.

Hill, S. Provost, F. and Volinsky, C. (2006). Network-based Marketing: Identifying Likely Adopters via Consumer Networks. Statistical Science, 21, 256-276.

Jang, H. Lorne, O. Ko, I. Koh, J. and Kim, K. (2008). The influence of on-line brand community characteristics on community commitment and brand loyalty. International Journal of Electronic Commerce, 12(3), 57–80.

Mael, F. and Ashforth, B. E. (1992). Alumni and their alma mater: a partial test of the reformulated model of organizational identification. Journal of Organizational Behavior, 13, March, 103–123.

Manski, C. and Lerman, S. (1977). The Estimation of Choice Probabilities from Choice-Based Samples. Econometrica, 45(8), 1977–88.

McAlexander, James H. Schouten J. W. and Koenig H. F. (2002). Building Brand Community. Journal of Marketing, 66(1), 38-54.

Mithas, S. and Krishnan, M. S. (2008). From Association to Causation via a Potential Outcomes Approach. Information Systems Research, December (published online before print).

Muniz, A. M. and. O'Guinn T. C. (2001) Brand Community. Journal of Consumer Research (27), 412–432.

Nair, H. Manchanda, P. and Bhatia, T. (2006). Asymmetric peer effects in prescription behavior: the role of opinion leaders. Mimeo, Stanford University.

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), 60–68.

Reichheld F. and Sasser, Jr. F.W. (1990). Zero defections: quality comes to services. Harvard Business Review, (68), 105–111.

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), 69-81.

Rosenbaum P.R. and D.B. Rubin. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1), 41-55.

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), 41-50.

Trusov, M. Bucklin, R. E. and Pauwels, K. H. (2008). Effects of Word-of-Mouth versus Traditional Marketing: Findings from an Internet Social Networking Site. Robert H. Smith School Research Paper No. RHS 06-065.

Tucker, C. (2004). Network effects and the role of influence in technology adoption. Mimeo, Stanford University.

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), 48–59.

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), 134-141.