Free vs. For a Fee: The Impact of Information Pricing Strategy on the Pattern and Effectiveness of Word-of-Mouth via Social Media

Hyelim Oh, Animesh Animesh, Alain Pinsonneault
Desautels Faculty of Management
McGill University

June 2013

The bulk of the work in this paper was done by a student

Free vs. For a Fee: The Impact of Information Pricing Strategy on the Pattern and Effectiveness of Word-of-Mouth via Social Media

Abstract

Given the new realities of the digital age, information goods providers like print newspapers are experimenting with different pricing models for their online content. However, it is not clear how information pricing strategy may influence word of mouth (WOM) via social media that has emerged as a dominant channel for raising awareness about a newspaper's articles and acquiring new visitors to newspaper's website. Using NYT's paywall rollout as a natural experiment, our study examines how a firm's information pricing policy (i.e., a shift from "free" to "for a fee") influences the pattern and effectiveness of online word-of-mouth (WOM) in social media. Specifically, our results indicate that implementing a paywall (i.e., charging for the content which was earlier available for free) has a disproportionate impact on the WOM for popular and niche articles, creating a longer tail in the WOM (i.e., content sharing) distribution. Further, we find that the impact of WOM on NYT's website traffic weakens significantly after the introduction of NYT's paywall. These results show that information pricing strategy has implications for product and promotion strategies. These results show that information pricing strategy has implications for product and promotion strategies. The study offers novel and important implications for the theory and practice of strategic use of social media and information pricing strategy.

Keywords: Digital goods, information pricing, social media, long tail, word-of-mouth

1. Introduction

Digital technologies and the Internet have significantly upended the business model of the newspaper publishing industry. Declining circulation of print newspapers (Vanacore 2010) and an increasing trend toward digital news consumption by consumers has driven traditional newspapers to adopt the Internet as the medium to offer digital content. However, given the intense competition in the online news market and almost zero marginal cost for providing news online, newspapers find it difficult to charge a fee for accessing their online content. Therefore, most online newspapers have been providing content free of charge while making money from online advertising. However, as more consumers switch from print to online news consumption, the online advertising revenue does not counterbalance the loss of revenue from print newspaper subscribers (Peters 2011).

Given these realities, it is not surprising that print news publishers have been debating the tradeoff of charging a fee versus providing the content for free. The New York Times (NYT) is one of such newspapers that are experimenting with different pricing models for their online content. Since both subscription and advertising revenue are a function of the newspaper's readership, it is important for NYT to ensure that the traffic to its website does not drop significantly. Recognizing the importance of retaining the current NYT website visitors with possibly low willingness to pay (WTP), NYT has implemented a generous access policy allowing non-subscribers to read 20 articles for free. However, merely retaining current customers is not enough, and like other business, a newspaper publisher needs to rely on advertising and word-of-mouth (WOM) to increase awareness and to acquire new customers. In the context of online newspapers, WOM via social media acts as a dominant channel for raising awareness about a newspaper's articles and acquiring new visitors to newspaper's website.

The WOM about news content is generated as individuals share the content in the social media by posting a link to the news article. Since content sharing is implicit WOM about an article's importance or relevance, we refer to content sharing as WOM in this paper. Recognizing the importance of social media such as Twitter and Facebook in creating WOM about online content and a resulting increase in awareness and traffic, NYT designed the paywall with special access policy for social media users. Specifically, NYT allowed individuals to bypass the paywall when they access NYT content by following NYT links shared in social media.

Intuition would suggest that the introduction of subscription fee model would lower the NYT's website traffic. However, it is not clear how this new pricing strategy will impact the WOM dynamics of the NYT's content in online social media. Using NYT's recent adoption of a paywall as a natural experiment, our study examines the impact of a firm's information pricing strategy on the pattern of WOM and the effectiveness of such WOM in social media.

The results show that a firm's information pricing strategy not only decreases the volume of WOM, but it also has a disproportionate impact on WOM distribution. Specifically, the results suggest that the WOM for popular content drops more significantly, resulting in less concentration of popular content in WOM distribution. Our theory-driven explanation of the long-tail demonstrates that light users who consume few articles have a tendency to consume popular articles. After a paywall introduction, light users have higher attrition likelihoods due to their low WTP for paid content. The relatively larger attrition of light users who prefer popular content reshapes the WOM distribution. To examine the effectiveness of WOM, combined Twitter content sharing data with NYT's website traffic data. Our results suggest that there is a positive impact of the volume of social media WOM on website traffic. However, the contribution of social media WOM to the website traffic weakens after the paywall introduction.

This study makes substantial contributions to the literature. First, we complement the extant research in the information pricing literature by proposing and empirically showing the significant interplay between a firm's pricing strategy (i.e., paywall) and its WOM pattern as well as WOM effectiveness. The unique natural experiment created by NYT's paywall rollout helps us control for confounding factors. Second, our study contributes to the WOM literature that measures the impact of online WOM on various performance measures by highlighting the role of WOM on newspaper website's traffic and proposing a moderating role of information pricing policy on the relationship between WOM and performance outcome i.e., website traffic. Third, we extend the literature examining various aspects of long tail phenomenon, especially in the online context, by examining the impact of information pricing strategy on "long tail" in the distribution of online WOM for information goods such as newspaper articles.

2. Theory and Hypotheses

2.1. The Impact of Paywall on WOM Pattern

Volume of WOM: Research has suggested the difficulty of charging a fee for content because of consumers' low WTP (Picard 2000; Chyi 2005). Consequently, a paywall is expected to lower the online readership of NYT. To the extent that a significant proportion of readers who stop accessing NYT (or reduce their consumption of NYT content) were actively sharing NYT articles earlier, the seeds (i.e., initial sharing) for NYT content will be decreased. Further, given that the volume of WOM in a social media also depends on the viral spread of the content through relationships among members of the social network, the lack of seeds that can be retransmitted by other members in the social media will weaken the potential viral effect. As a result, a paywall will lead to a reduced volume of WOM by lowering both the amount of links to news content shared through seeds and retransmission of seeds.

H1a: The introduction of a paywall will decrease the volume of WOM about NYT content over social media.

Distribution of WOM: Building upon the prediction in H1a, we develop our hypothesis about the role of paywall in shifting the distribution of WOM. The readers of an online newspaper exhibit heterogeneity in terms of the WTP for accessing the digital content. Though their WTP may depend on a variety of factors (Chyi 2005), we focus on usage intensity (i.e., the number of articles read and/or shared), which is one of the important factors that influence the value of the product/service in the context of subscription for accessing content. Consequently, usage intensity is expected to determine whether a consumer will subscribe for the content at a given price or not (Danaher 2002). The NYT paywall pricing strategy essentially creates two versions of the product – a "*free version*" that allows access to fewer than 20 articles and the "*subscription version*" that allows unlimited access to NYT. The consumers who are "light users" tend to be sensitive to price. Therefore, they are more likely to select a free version or switch to alternative news sources. On the other hand, those consumers who are "heavy users" will find the cost of articles lower due to their usage intensity. Thus, they will be more willing to subscribe and select paid content. As a result, after the paywall implementation, the distribution of users in terms of their usage will shift toward heavy users as the paywall will disproportionately affect light users who will have a higher attrition rate.

Next, we argue that this shift in the distribution of consumers due to attrition after the paywall will affect the distribution of the NYT articles read and shared. Recognizing that consumers differ in terms of their preference for popular versus niche products, prior research suggests that the frequency of usage is associated with consumption patterns in terms of content popularity. According to the theory of exposure (McPhee 1963), popular products monopolize the consumption of light consumers, whereas heavy consumers choose a mix of hit and niche products. The exposure theory suggests that consumers who choose niche products tend to be familiar with many alternatives while those who know of few alternatives tend to stick with popular products. Likewise, the variety-seeking literature suggests that users' preference for niche products is positively associated with their quantity of consumption. Simonson (1990), for instance, finds that those customers who buy larger quantities per purchase tend to select a greater variety of items. Elberse (2008) also finds evidence that customers with a higher frequency of usage have a tendency to consume niche products.

To summarize, as illustrated in Figure 1, we expect that the light user segment is more likely to consume popular articles, whereas the heavy user segment is more likely to consume niche articles. Juxtaposing this statement with the earlier argument that light user segments are more likely to discontinue or reduce their contribution to the WOM diffusion about NYT's content after the paywall implementation, we expect that light users who read more popular articles will exhibit stronger attrition vis-à-vis heavy users who read a mix of niche and popular articles.

As a consequence of the attrition of light users who are expected to disproportionately consume more popular articles, there would also be a change in the distribution of the WOM about NYT's content in social media after a paywall implementation. Specifically, as the number of consumers generating WOM about popular content will decrease, the distribution of WOM will shift towards the articles in the tail of the content popularity distribution. Therefore, we expect a longer tail (less concentrated) in the WOM distribution after the paywall implementation.

H1b: The introduction of a paywall will lead to a longer tail of WOM distribution (i.e., the volume of WOM for popular content will drop more than the volume of WOM for niche content).

2.2. The Impact of Paywall on WOM Effectiveness

Next we focus on WOM effectiveness. In our context, WOM effectiveness refers to the ability of WOM to increase the number of individuals who consume the information good. Given that the consumption of newspaper articles occurs at the newspaper's website, an increase in consumption is synonymous with the increase in website traffic. Therefore, we examine the association between the volume of online WOM. The WOM generated as a result of content spreading over an online social network through transmission, consumption and retransmission of information (Stephen and Lehmann 2012; Berger and Milkman 2011) create awareness about the content and increase the set of potential users who can consume the content. The initial group of users who post a message containing a link to an article, also referred to as seeders, broadcast their opinion about the importance or relevance of an article to their followers. We can assume that the seeders are regular visitors to NYT's website. To the extent the WOM by these initial seeders reaches those network members who are not frequent visitors to NYT website and a significant proportion of them decide to visit the newspaper's website, the increase in WOM would lead to higher website traffic. To summarize, in aggregate, the WOM about a firm's content in online social networks is likely to bring new readers (who directly or indirectly follow the seed reader) to the firm's website.

H2a: The volume of WOM will have a positive relationship with website traffic.

After establishing the relationship between WOM and website traffic, we examine how this association differs before and after a paywall implementation. In this highly connected digital age, firms are aware of the value of social media and actively engage them in their business strategy. For example, NYT allows visitors who come from links on social media to bypass its paywall. If this bypass effect is dominant in website traffic generation, the relative strength of the relationship between social media WOM and website traffic may increase after a paywall implementation. On the other hand, a proportion of NYT readers who did not subscribe to NYT after the paywall would have frustrating reading experiences when they get constrained by the NYT paywall's free article limit. To the extent that a significant number of these readers reduce their likelihood for reading NYT article due to their negative experience with NYT's website, the WOM effect on website traffic will weaken as these users will ignore the WOM about NYT on social media. In other words, the exposure to WOM would have weaker impact on website traffic as the likelihood of clicking on NYT content will decrease for a segment of readers who have unpleasant experience with NYT website.

Though the net impact of the paywall on the relationship between WOM and website traffic would be an empirical question, we suggest that the paywall will weaken the strength of relationship between social media WOM and website traffic due to the changes in the distribution of WOM after the paywall implementation. As we argued earlier, the paywall leads to disproportionate decrease in the WOM about popular content. Given that popular articles are the content that is appreciated more by a larger audience (Zentner et al. 2012), we expect that a decrease in the proportion of WOM about popular content will in turn lead to lower the average clicks per link shared through social WOM. We expect that the bypass effect (as mentioned above) will not be able to mitigate the negative effect of change in WOM distribution.

H2b: The association of the volume of social WOM with website traffic will be weakened after a paywall implementation.

3. Data

Our goal in the natural experimental setting is to analyze whether WOM pattern on Twitter and WOM dynamics related to website traffic significantly differ before and after NYT's paywall rollout on March 28, 2011. To test our hypotheses, we combine data from two sources: NYT link sharing on Twitter and website traffic for 21 days before and after NYT's paywall rollout. One potential concern here is whether the treatment effect of a paywall might be biased because of a greater proportion of interesting news events in the pre-paywall period, or vice-versa. In order to isolate this time trend effect and the influence of other extraneous factors, we employ a difference-in-difference approach (Chevalier and Mayzlin 2006) and chose the LA Times (LAT) as a control group as LAT is a major national newspaper in a different geographic region. We collected tweets that contain newspaper link sharing for NYT and LAT articles between February 26, 2011 and March 18, 2011, and between April 4, 2011 and April 24, 2011, respectively. In the time periods, we collected the exhaustive sets of NYT and LAT link sharing. Our final data sets contain 1,287,570 tweets embedding NYT links and 226,911 tweets containing LAT links. The population-size of the data allows us to create and analyze the distributions of WOM, which is elaborated in more detail below. We were also able to obtain daily website traffic data for the two newspapers in the same time period from HitWise[1].

3.1. Measures for WOM (i.e., News Link Sharing) Pattern Analysis

To test the impact of NYT's paywall on the pattern of WOM, we conduct a content-level analysis. We compute the counts of NYT WOM and LAT WOM, which are used as the dependent variables for examining the impact of paywall on the volume of WOM. We examine the distribution of WOM by computing WOM Rank of Content from both the NYT and LAT datasets. Consistent with prior empirical studies (Brynjolfsson et al. 2003; Brynjolfsson et al. 2011), NYT content with the highest WOM is assigned the lowest value for WOM Rank of Contenti. We follow prior long tail studies (Brynjolfsson et al. 2003; Brynjolfsson et al. 2011; Zentner et al. 2012) and measure popularity of content in a relative sense using an ordial ranking

[1] Hitwise data is based on an extensive sample size collected from diverse range of ISP networks and opt-in panels, which ensures that the data is accurate and representative.

of the frequency of content sharing. In our sample, a significant proportion of tweets contained NYT and LAT links to past articles, which may create positive inflation of the paywall effect because the articles in the pre-paywall block simply had a longer duration in our sample allowing greater opportunities for these articles to be shared. To prevent this bias, we use a 1-day horizon (i.e., only focus on WOM for an article that was generated on the day the article was published) in measuring the dependent variables. Table 1 presents the summary statistics of our data sets.

3.2. Measures for WOM Effectiveness Analysis

To examine the impact of a paywall on WOM effectiveness, we conduct website-level analysis. We created the measures of daily WOM volume (aggregated over all the articles in a day) and website traffic. Website traffic metric is widely used in online advertising to determine where and how to advertise. Website visitst is defined as the number of unique visitors to a website on a day. In addition, we counted the total number of Tweets that contain either NYT or LAT link shared in day t. This variable represents the daily WOM volume. Table 2 presents the summary statistics of our site-level data sets. As our data shows that news link sharing has a significant variation depending on the week-days, we create two weekend dummy variables. Initial investigation showed that the independent variable, Tweets, was not normally distributed. As suggested by Gelman and Hill (2007), we take the logarithm on Tweets volume to control for its left-skewed nature. Since the inclusion of paywall dummy variable and the interaction term of paywall and the lagged variable of Tweets volume may lead to high multicollinearity, we mean-centered these variables before generating their interaction term.

4. Model and Empirical Findings

We now describe our empirical testing approach and the findings for the hypotheses.

4.1. The Impact of Paywall on WOM Pattern

4.1.1. Volume of WOM: Difference-in-Differences (DID) Approach

We compare the WOM for 21 days before and after the paywall rollout for NYT content (i.e., the treatment group) and LAT content (i.e., the control group). Because the dependent variable, $\mathrm{WOM_i}$, is the count of link sharing for either NYT or LAT content i, we employ a count model. Hence, we estimate a negative binomial regression (NBD). For the sake of exposition, in Equation 1, we present the specification of our *difference-in-differences* (DID) model for content *i* in a simple linear form. However, the NBD model has a nonlinear functional form, with the expected count of the dependent variable modeled as $\mathrm{E}(\mathrm{y_i}|X_i) = \mu_i = \exp(X_i\beta + \varepsilon_i)$.

$$\mathrm{WOM_i} = \beta_0 + \beta_1 \mathrm{Paywall_i} + \beta_2 \mathrm{NYT_i} + \beta_3 \mathrm{Paywall_i} \times \mathrm{NYT_i} + \varepsilon_i, \ \ \varepsilon_i \sim \mathrm{N}(0, \sigma^2), \quad (1)$$

where $\mathrm{Paywall}_i$ is a dummy that equals one if the time period is after the paywall rollout period (see Table 3 for the variable description). It captures the aggregate factors that would cause changes in the dependent variable in both the treatment and control groups. The dummy variable, NYT_i, captures the possible differences between the treatment and control groups prior to the paywall rollout. The coefficient of interest, β3, is an estimate of the interaction term, $Paywall_i \times NYT_i$, which equals one if an observation is in the treatment group in the post-paywall period.

4.1.2. Pattern of WOM: DDD Approach

To test H1b, we follow prior long tail studies (Brynjolfsson et al. 2003; Brynsjolfsson 2011) and employ log-linear regression to examine how a paywall reshapes the WOM distribution with respect to content popularity. Specifically, we fit WOM and WOM Rank of Content to log-linear regressions that model the power-law distributions of WOM at the content level.

$$\ln(\mathrm{WOM_i}) = \beta_0 + \beta_1 \ln(\mathrm{WOM\ Rank\ of\ Content_i}) + \beta_2\ \mathrm{Paywall_i} + \beta_3 \mathrm{NYT_i} + \beta_4 \mathrm{Paywall_i} \times \ln(\mathrm{WOM\ Rank\ of\ Content_i}) + \beta_5 \mathrm{Paywall_i} \times \mathrm{NYT_i} + \beta_6 \mathrm{NYT_i} \times \ln(\mathrm{WOM\ Rank\ of\ Content_i}) + \beta_7\ \mathrm{Paywall_i} \times \ln(\mathrm{WOM\ Rank\ of\ Content_i}) \times \mathrm{NYT_i} + \varepsilon_i. \quad (2)$$

Models 3 and 4 in Table 4 present the results of the Pareto curve estimation. We focus on the results of the DDD model. The β_7 coefficient on the three-way interaction of the paywall, NYT and WOM rank terms is statistically significant. The interaction plots in Figure 4 plots the two-way interactions between Paywall and WOM rank. As shown in Figure 4A, the absolute value of the slope coefficient for NYT decreases after the paywall implementation. The results suggest that the paywall rollout leads to a shift of NYT's WOM distribution such that the decrease in WOM for articles with higher popularity (i.e., lower WOM rank) is much larger than the decrease in WOM for articles with lower popularity (i.e., higher WOM rank). Thus, we find support for Hypothesis 1B. The slope coefficient for LAT does not exhibit any significant changes before and after NYT's paywall rollout, suggesting that the shift of NYT's WOM distribution is not caused by extraneous factors such as time trends. Therefore, we conclude that NYT's WOM distribution becomes less concentrated (a longer tail) after the paywall rollout.

4.1.2. Long-tail Explanation

To support the underlying mechanism that we proposed for the long-tail effect, we perform a direct test of the inverse relationships between consumers' content sharing intensity and the rank of the content (in terms of WOM popularity) shared by them. Then, we identify two distinct segments of users in terms of content sharing intensity and the rank of the content shared. Finally, we relate these user segments to the likelihood of attrition after the paywall. The results of our latent segmentation analysis using a finite mixture model show that users who have lower consumption levels and higher tendency to share popular content are more likely to leave or discontinue after the paywall than users in the other segment, thus supporting our logic of long tail explanation (see Appendix A for the detailed model specification and estimation results).

4.2. The Impact of Paywall on WOM Effectiveness

4.2.1. Simultaneous Model of Paywall, WOM diffusion and Website Traffic

In the second part of our analysis, we examine the relationship between the volume of daily social WOM and daily website traffic and the dynamics of this relationship after a paywall introduction. Given the interdependence between WOM diffusion and website traffic, we develop the following three-stage least-square (3SLS) model at the site-level: one equation with daily website visits as the dependent variable (the website traffic equation) and one with daily WOM volume as the dependent variable (the WOM equation). Given the interdependence between WOM volume and website traffic, we include the lagged dependent variables.

$$\text{Website visits}_{\mathrm{t}} = \theta_{\mathrm{t}} + \alpha_1 \ln(\mathrm{WOM}_{\mathrm{t}-1}) + \alpha_2 \text{Website visits}_{\mathrm{t}-1} + \alpha_3 \text{Paywall}_{\mathrm{t}} + \alpha_4 \text{Paywall}_{\mathrm{t}} \times \ln(\mathrm{WOM}_{\mathrm{t}-1}) + \alpha_5 \text{Controls}_{\mathrm{t}} + u_{\mathrm{t}}. \quad (4)$$

$$\ln(\mathrm{WOM}_t) = \eta_t + \beta_1 \text{Website visits}_{\mathrm{t}-1} + \beta_2 \, ln(\mathrm{WOM}_{t-1}) + \beta_4 \text{ Paywall}_t + \beta_5 \text{Controls}_t + v_t. \quad (5)$$

For the website traffic equation, Website visits_t denotes the daily gross website traffic at day *t*, and its one-day lagged variable is defined as Website visits_{t-1}. Similarly, $\ln(\mathrm{WOM}_t)$ represents the total number of tweets that contain NYT link at day *t*. To control for the skewed nature of the WOM volume, we took the log transformation on the variable. We include the lagged dependent variables as instruments for endogenous relationships between Website traffic and WOM volume. The dummy variables, Saturday_t and Sunday_t, are included in all equations to control for variations in the content and user reading behavior during weekends.

Table 5 presents the estimation results for the 3SLS model. We find that there is a positive and statistically significant relationship between $\ln(\mathrm{WOM}_{t-1})$ and Website visits_t in the website traffic equation (α_1 = 381,603 ; $p < 0.05$), confirming H2a. We now turn to examine the impact of paywall on such interdependence. The results also show that both Paywall ($\alpha_2 = -160{,}813$; $p < 0.01$) and the interaction of Paywall and $\ln(\mathrm{WOM}_{t-1})$ ($\alpha_3 = -244{,}136$; $p < 0.1$) are negative and statistically significant. We plot the interaction between paywall and the lagged WOM volume. The graph shows that the magnitude of WOM's effect on website traffic generation decreases after NYT's paywall, supporting H2b. The results provide evidence that social WOM is a significant predictor of website traffic; however, a firm's information pricing changes these dynamics. We now turn to discuss the effect of paywall on the counterfactual in our control group (i.e., LAT) that we used to alleviate any confounding factors that are associated with the timing of the NYT payawall implementation. As expected, our results for the same analysis using the LAT sample indicate that NYT's paywall rollout has no significant effect on LAT's website traffic. Furthermore, consistent with our main results from the NYT sample, the volume of WOM on Twitter has a positive and significant effect on website traffic generation ($\alpha_1 = 0.219$; $p < 0.1$), providing additional empirical support for H2a.

4.2.2. Economic Significance

We examine the economic impact of WOM on website traffic and the impact of paywall on this relationship using their marginal effects. As reported in Table 2, the average of log of WOM volume is 1,709,428 in the pre-paywall period. A one-standard deviation (14.4 percent) increase in WOM volume in the previous period increase the volume of website traffic in the current period by 3.4 percent when evaluated at the mean before paywall rollout. However, after NYT's paywall implementation, the same one-standard deviation increase in WOM volume in the previous period increase of website traffic in the current period by only 2.3 percent.

5. Contributions, Limitations and Future Research

The findings of this study offer novel and important implications for the theory and practice of strategic use of social media and information pricing policy.

Prior work in the information pricing literature has primarily examined, mostly analytically, the interaction between information pricing and product strategies (Wu et al. 2008; Bakos and Brynjolfsson 1999; Chen and Seshadri 2007; Choudhary 2010). This study extends this literature by examining the interconnection between information pricing and promotion (i.e., WOM) strategies. We conduct an empirical study that shows the impact of pricing strategy on WOM pattern, which is a part of firm's promotion strategy and has implications for firm's product strategy. Incorporating the effects of different pricing policies on WOM pattern and effectiveness in analytical models will provide a more realistic assessment of the impact of pricing policies on firm performance.

Our study also contributes to the WOM literature. With the emergence of social media, researchers have examined the role of WOM in content diffusion (Susarla et al. 2012) as well as its impact on various aspects of performance (Chevalier and Mayzlin 2006). Our study extends the WOM literature by highlighting that the distribution of WOM for hit/niche content may change as a result of a change in information pricing policy. Further, complementing the studies that measure the impact of online WOM on various performance measures such as product adoption and sales (Godes and Mayzlin 2004; Liu 2006), we suggest that WOM has positive impact on website's traffic, which is a critical factor influencing financial performance.
Finally, our study extends the literature examining various aspects of long tail phenomenon, especially in the online context. Our study examines the impact of information pricing strategy on "long tail" in the distribution of online WOM for information goods such as newspaper articles. Our work enrich this stream of research by highlighting that a firm's information pricing can be a demand side driver of long tail outcomes (Brynjolfsson et al. 2010).

This study also provides several managerial implications for information goods providers. First, our results suggest that a content publisher that has implemented a paywall should be aware of its impact on the changes in the content consumption and sharing distribution. Second, our results also provide support to the practitioners' belief that online WOM plays an important role in the firm's performance. More importantly, we also suggest that pricing policy shift may not only lower the WOM but also weaken WOM's power to bring more website traffic. These results suggest that companies need to measure and manage the impact of firm's strategy on WOM. To achieve this objective, a company may need to invest in information systems that capture, monitor, and analyze the clickstream data from their webservers as well as WOM data from social media platforms like Twitter. Further, managers may need to actively manage the WOM dynamics, especially during the periods of shift in firm strategies.

One caution warranted in generalizing our finding is a possible first mover effect of NYT's paywall. It is possible that consumers may become used to paid subscription models over time. Thus, the behavior of consumers in response to similar pricing strategies by late movers may be different, leading to different WOM dynamics. However, we believe that this first mover effect can only impact the magnitude of our findings. Based on our theory, if consumers become more comfortable with the idea of paying for information goods, the proportion of light users that drop after the paywall will be smaller and consequently the drop in the volume of WOM for popular content will also be relatively low, weakening the long tail pattern of WOM that we observed in our NYT data. In such case, we expect that our theoretical explanation of user preference for product variety and the resulting long tail pattern would hold, through the magnitude of effects will be smaller. Future research on later adoption of paywall would provide more salient implications for firms considering pricing strategy.

References

Bakos, Y., and Brynjolfsson, E. 1999. "Bundling information goods: Pricing, profits, and efficiency," *Management Science* (45:12), pp 1613-1630.

Berger, J., and Milkman, K. L. 2012. "What makes online content viral?," *Journal of Marketing Research* (49:2), pp 192-205.

Brynjolfsson, E., Hu, Y. J., 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.

Brynjolfsson, E., Hu, Y. J., and Simester, D. 2011. "Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales," *Management Science* (57:8), pp 1373-1386.

Brynjolfsson, E., Hu, Y. J., and Smith, M. D. 2010. "Research commentary—long tails vs. superstars: The effect of information technology on product variety and sales concentration patterns," *Information Systems Research* (21:4), pp 736-747.

Chen, Y.-J., and Seshadri, S. 2007. "Product development and pricing strategy for information goods under heterogeneous outside opportunities," *Information Systems Research* (18:2), pp 150-172.

Chevalier, J. A., and Mayzlin, D. 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research* (43), pp. 345-54.

Choudhary, V. 2010. "Use of pricing schemes for differentiating information goods," *Information Systems Research* (21:1), pp 78-92.

Chyi, H. I. 2005. "Willingness to pay for online news: An empirical study on the viability of the subscription model," *Journal of Media Economics* (18:2), pp 131-142.

Danaher, P. J. 2002. "Optimal pricing of new subscription services: Analysis of a market experiment," *Marketing Science* (21:2), pp 119-138.

Dewan, R. M., Freimer, M. L., and Zhang, J. 2003. "Management and valuation of advertisement-supported web sites," *Journal of Management Information Systems* (19:3), pp 87-98.

Dewan, S., and Ramaprasad, J. 2012. "Research Note—Music Blogging, Online Sampling, and the Long Tail," *Information Systems Research* (23:3-Part-2), pp 1056-1067.

Elberse, A. 2008. "Should you invest in the long tail?," *Harvard business review* (86:7/8), p 88.

Gelman, A., and Hill, J. 2007. *Data analysis using regression and multilevel/hierarchical models*, New York: Cambridge University Press.

Godes, D., and Mayzlin, D. 2004. "Using online conversations to study word-of-mouth communication," *Marketing Science* (23:4), pp 545-560.

Liu, Y. 2006. "Word of mouth for movies: Its dynamics and impact on box office revenue," *Journal of marketing* (70), pp 74-89.

McLachlan, G., and Peel, D. 2004. *Finite mixture models*, New York: Wiley.

McPhee, W. N. 1963. *Formal theories of mass behavior.* Free Press of Glencoe, New York, NY.

Peters, J. 2011. The Times Announces Digital Subscription Plan. Available at http://www.nytimes.com/2011/03/18/business/media/18times.html.

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.

Simonson, I. 1990. "The effect of purchase quantity and timing on variety-seeking behavior," *Journal of Marketing Research* (27:2), pp 150-162.

Stephen, A., and Lehmann, D. 2012. "Using Incentives to Encourage Word-of-Mouth Transmissions that Lead to Fast Information Diffusion," Working paper.

Susarla, A., Oh, J.-h., and Tan, Y. 2012. " Influentials or Susceptibles? Analyzing Cascades of Word-of-Mouth Conversations in Online Social Networks," Working paper.

Wu, S.-y., Hitt, L. M., Chen, P.-y., and Anandalingam, G. 2008. "Customized bundle pricing for information goods: A nonlinear mixed-integer programming approach," *Management Science* (54:3), pp 608-622.

Vanacore. 2010. US newspaper circulation falls 8.7 percent. *Associated Press*, Available at http://finance.yahoo.com/news/US-newspaper-circulation-apf-436809869.html?x=0.

Zentner, A., Smith, M., and Kaya, C. 2012. "How Video Rental Patterns Change as Consumers Move Online," Working paper.

Figures and Tables

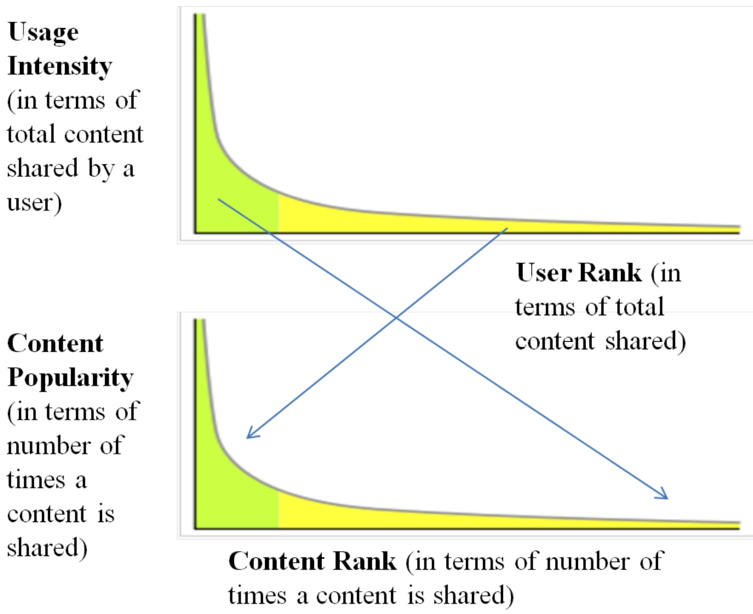


Figure 1. The Inverse Relationship between Content Popularity and Consumer Usage

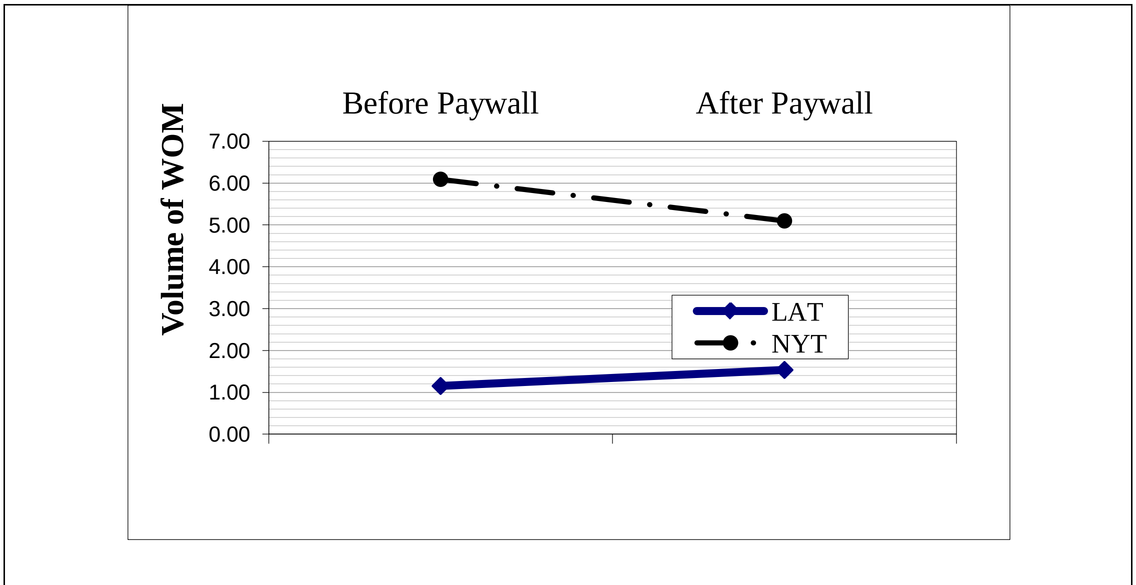
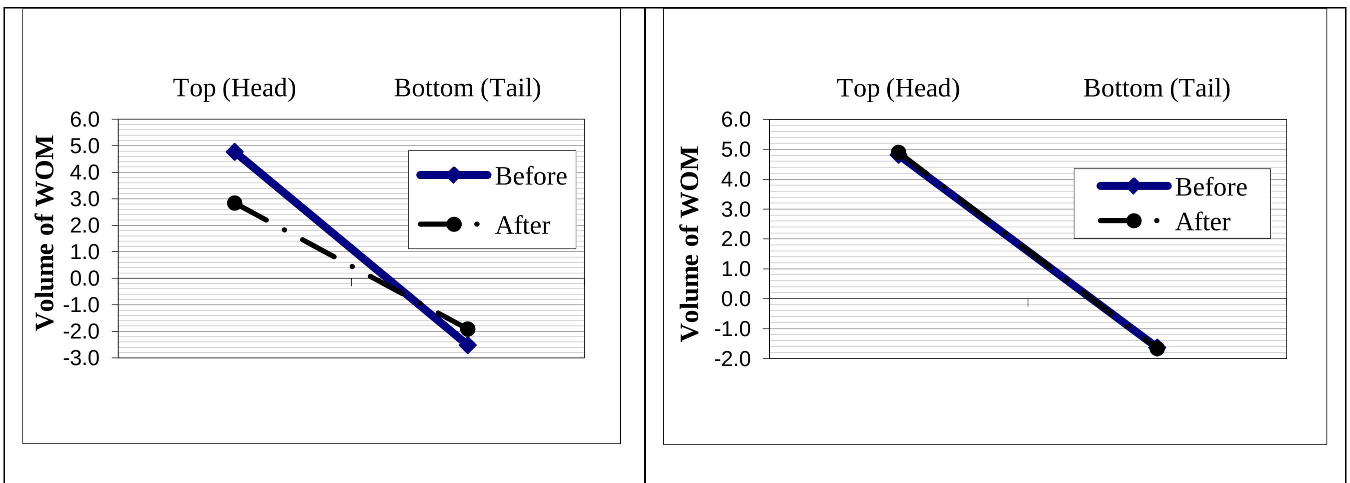


Figure 2. Interaction between NYT Paywall and NYT's vs. LAT's WOM



(a) New York Times

(b) LA Times

Figure 3. Interaction between NYT Paywall and Content's Popularity Rank (in terms of WOM)

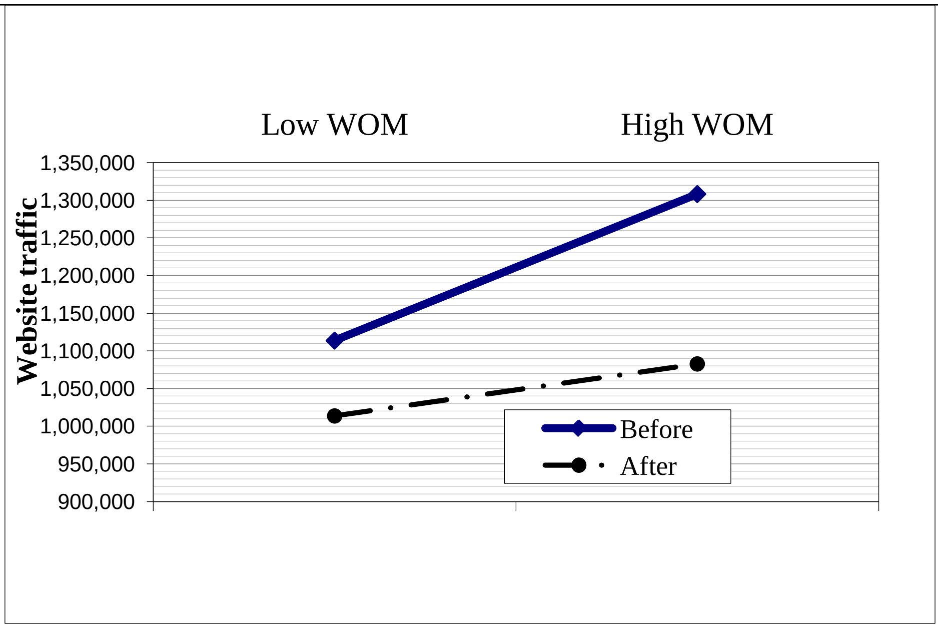


Figure 4. Interaction between NYT Paywall and Website Traffic (3SLS)

Table 1. Summary Statistics: Content-level Data

	Before Paywall Roll-out				After Paywall Roll-out			
	Mean	SD	Min	Max	Mean	SD	Min	Max
NYT WOM	55.20	265.87	1	8563	47.41	140.66	1	4517
	N = 10604				N = 10578			
LAT WOM	10.84	33.66	1	1563	11.00	23.57	1	586
	N = 4618				N = 4512			

Table 2. Summary Statistics: Site-level Data

	Before Paywall Roll-out				After Paywall Roll-out			
	Mean	SD	Min	Max	Mean	SD	Min	Max
NYT Website Visits	1,709,428	247,586.2	1,436,089	2,268,483	1,428,147	106,697.7	1,257,954	1,577,606
NYT WOM	33,772.67	9,595.23	20,260	54,100	27,540.19	5,313.42	16,934	34,513
LAT Website Visits	294,845	73,609.88	224,487	536,155	263,798.7	17,724.68	231,990	295,065
LAT WOM	5,308.286	1,583.748	2,800	9,817	5,497	1,103.353	2,752	7,052

Table 3. Variable Definitions

Site-level Variables	
$Paywall_t$	A dummy variable indicating if day *t* is in the post-paywall period
Website $visits_t$	Total volume of daily website traffic in day *t*
WOM_t	Total volume of NYT link sharing on Twitter in day *t*
$Saturday_t$	A dummy variable coded as 1 if the day is Saturday
$Sunday_t$	A dummy variable coded as 1 if the day is Sunday
Content-level Variables	
$Paywall_i$	A dummy variable coded as 0 if an observation is prior to the paywall rollout, otherwise 1
NYT_i	A dummy variable coded as 1 if an observation is a NYT link, otherwise 0
NYT WOM_i	1-day diffusion count of NYT article link sharing on Twitter
LAT WOM_i	1-day diffusion count of LAT article link sharing on Twitter
WOM Rank of $Content_i$	Rank of an article after sorting articles (in descending order) based on the volume of WOM for that article (i.e., the highest shared articles get the lowest rank)

Table 4. Content-level Analysis

	Model 1	Model 2	Model 3	Model 4
Dependent variable	NYT WOM	WOM[a]	ln(NYT WOM)	ln(WOM)[b]
$Paywall_i$	-0.154***	0.0133	0.702***	0.0636***
	(0.021)	(0.0311)	(0.0825)	(0.00729)
NYT_i		1.619***		1.827***
		(0.0261)		(0.0117)
$Paywall_i \times NYT_i$		-0.166***		0.0834***
		(0.0370)		(0.0167)
ln(WOM Rank of $Content_i$)			-1.586***	-1.443***
			(0.00627)	(0.00476)
$NYT_i \times$ln(WOM Rank of $Content_i$)				-0.475***
				(0.0105)
$Paywall_i \times$ln(WOM Rank of $Content_i$)			-0.0707***	-0.0512***
			(0.00919)	(0.00690)
$Paywall_i \times NYT_i \times$ln(WOM Rank of $Content_i$)				-0.0649***
				(0.0151)
$Saturday_i$	0.132***	0.0602**	-0.0380***	-0.0301***
	(0.032)	(0.0268)	(0.0136)	(0.0107)
$Sunday_i$	0.167***	0.115***	-0.0575***	-0.0512***
	(0.034)	(0.0275)	(0.0144)	(0.0110)
Constant	3.974***	2.370***	16.50***	2.204***
	(0.016)	(0.0221)	(0.0566)	(0.00543)
No. of observations	21,182	30,312	21,182	30,312
R-squared	0.0004	0.0199	0.855	0.866

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.
[a]Model 2: Difference-in-difference model.
[b]Model 4: Difference-in-difference-in-difference model; Because of a concern regarding multicollinearity, ln(Content Popularity) and NYT are mean-centered in Model 4. After such standardization, the VIF scores for all variables are lower than 2.7, and the mean VIF score is 1.94.

Table 5. Estimation Results for WOM and Website Traffic (3SLS Model)

	NYT[a]	LAT[a,b]
Dependent Variable	Website $visits_t$	ln(Website $visits_t$)
ln (WOM_{t-1})	381,603**	0.219*
	(159,887)	(0.123)
Website $visits_{t-1}$	0.298*	1.62e-06***
	(0.172)	(4.82e-07)
$Paywall_t$	-160,813***	-0.0522
	(48,829)	(0.0414)
$Paywall_t$ × ln(WOM_{t-1})	-244,136*	-0.142
	(132,128)	(0.136)
$Saturday_t$	-181,744***	-0.120**
	(55,948)	(0.0567)
$Sunday_t$	-58,550	-0.0484
	(61,116)	(0.0620)
Constant	1.210e+06***	12.13***
	(283,729)	(0.142)
R-squared	0.764	0.561
Dependent Variable:	ln(WOM_t) :	ln(WOM_t)
Website $visits_{t-1}$	0.0941	0.265*
	(0.174)	(0.144)
ln (WOM_{t-1})	0.000*	7.87e-07
	0.000	0.000
$Paywall_t$	-0.0885	0.0573
	(0.0570)	(0.0520)
$Saturday_t$	-0.367***	-0.483***
	(0.0678)	(0.0708)
$Sunday_t$	-0.318***	-0.432***
	(0.0740)	(0.0777)
Constant	-0.424	-0.112
	(0.328)	(0.173)
R-squared	0.717	0.728

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. N= 40

[a] To avoid a concern of multicollinearity (i.e., the collinearity between main effect of paywall and the interaction effect of paywall and WOM volume), ln(WOM_t) is mean-centered for all the relevant variables.

[b] Because of the skewed nature of our LAT website traffic data (i.e., Website $visits_t$), we took log transformation of the relevant variables (i.e., Website $visits_t$, and Website $visits_{t-1}$) in the LAT sample.

Appendix A. Long tail Explanation

The observations in our sample are at the tweet message-level, and the measures can be created by aggregating either at the content or user-level as per the context of the analysis. We incorporate variables based on both kinds of aggregation in this analysis. Based on *WOM rank* of content *i*, we create a user's *Average Rank of Content Shared* at the user level. Average Rank of Content Shared_j is created by aggregating all the articles shared by a user and then taking an average of the "WOM Rank of Content" measure for each article. The metric captures a user's tendency to consume and share popular versus niche content. A lower mean reflects a user's preference for sharing popular content. Then, we create *User rank* for user *j* by sorting users based on the count of the total articles shared by each user in descending order (i.e., the highest sharer gets the lowest rank). Finally, we create $\textit{Attrition}_j$ as a dependent variable by comparing users in the pre- and post-paywall samples and identifying users who discontinued NYT content sharing after the paywall rollout. Specifically, it is coded as 1 if a user in the pre-paywall sample is not found in the post-paywall sample.

We conjecture that our paywall treatment effect is caused by attrition or reduced WOM behavior of certain user segment that has low WTP. In order to take into account a differential impact of a paywall with respect to user heterogeneity, we conduct a latent user segment analysis using finite mixture modeling. The purpose of our latent segment analysis is to identify latent user segments in terms of content sharing level and preference for content popularity. These segments can then be compared for difference in attrition likelihoods[2].

[2] In order to reduce the computational burden of estimating a finite mixture model with large-scale data, we took a subsample approach and used 22 percent of our NYT data. In this dataset, we identified that 23,310 (63.7 percent) users in the pre-paywall sample discontinued NYT content sharing after the paywall rollout.

The finite mixture of a normal regression model captures the differences between light and heavy user segments in the consumption of head and tail content in the content sharing distribution. The S-segment finite mixture model is specified as follows:

$$f(\mathrm{y_j}|X;\ \theta_{1,}\ \theta_2, \dots, \theta_S;\ \pi_{1\ ,}\pi_2, \dots, \pi_S) = \sum_{k=1}^{S} \mathrm{\pi_k f_k}(\mathrm{y_j}|\mathrm{X}; \theta_k), \quad 0 < \theta_s < 1 \text{ and } \textstyle\sum_{k=1}^{S} \mathrm{\pi_k} \tag{3}$$

where $\mathrm{y_j}$ denotes average content popularity as a dependent variable, vector θ_k $= \left(1, X_j^{(1)}, \dots, X_j^{(k)}\right)$ contains all of the parameters of *k* predictor variables that could explain the user-content dimensions, and π_S is the mixing proportions (weights) of the predictor variables $\mathrm{X_j}$. $f_k(y|X; \theta_s)$ are the segment densities of the mixture.

To select the number of segments, we computed model fit indices including the Bayesian Information Criteria (BIC) presented in Table A1 (McLachlan and Peel 2000). For the sake of brevity, we choose the two-segment solution. Table A2 reports the parameter estimates for the 2-segment model[3]. Based on the estimated coefficients and segment weights in Table A2, we conducted post-estimation to determine each subject's segment membership. Tables A3 indicates that Segment 1 is comprised of the light users who read less are more likely to read popular content, whereas Segment 2 is composed of the heavy users who read more and less popular (a mix of niche and popular) content.

[3] To conserve space, we omit the estimation results for the 3-segment model. However, the results are qualitatively similar to the 2-segment model.

Table A1. Model Fit for Alternative Numbers of Segments

Number of latent segments	LL	AIC	BIC	R^2
1	-78369.01	156744.03	156769.58	0.0072
2	-75868.29	151750.59	151810.21	0.6602
3	-75069.08	150160.16	150253.84	0.7887
4	-75030.70	150091.40	150219.14	0.7930

Note. In terms of the BIC criterion, the models with greater numbers of segments improve model performance. However, an interpretation with a two-segment model is more suitable for our hypothesis testing, and adding an additional segment marginally improves the model fit indices after the three-segment model. We therefore opt to report the model estimates for both the two- and three-segment models.

Table A2. Parameter Estimates of Finite Mixture Models

Dependent variable: ln (Average Rank of Content Shared$_j$)		
	Segment 1	Segment 2
Intercept	7.608(.144)***	4.946(.098)***
ln (User rank$_j$)	-0.652(.015)***	0.016(.011)***
Proportion	42.60%	57.40%

Note. Standard errors are in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A3. Postestimation of Finite Mixture Model

	Segment 1		Segment 2	
	Mean	[95% conf. interval]	Mean	[95% conf. interval]
ln (Average Rank of Content Shared$_j$)	1.670(.008)	[1.633, 1.687]	5.166(.007)	[5.152, 5.181]
ln (User rank$_j$)	9.070(.005)	[9.059, 9.081]	8.971(.006)	[8.958, 8.984]

Note. Standard errors are in parentheses. Segment 1 and Segment 2 represent Light & Popular (read less and popular articles) and Heavy & Mix of niche and popular (read more and niche articles) user segments, respectively.

Further, we conduct two-sample test of proportion to examine the differences among these two segments in terms of attrition likelihoods after paywall implementation. The results presented in Table A4 show that users who have lower consumption levels and higher tendency to share popular content are more likely to leave or discontinue after the paywall than users in the other segment, thus supporting our logic of long tail explanation.

Table A4. Two-sample test of proportion

Segment	Mean	[95% of conf. interval]	Test statistic
Light &Popular	0.688	[0.681, 0.695]	z = 19.4303***
Heavy & Mix of niche and popular	0.589	[0.583, 0.596]	

Note. The attrition likelihood in Segment 1 – Light & Popular (0.688) is significantly greater than the attribution likelihood in Segment 2 – Heavy & Mix of Niche and Popular (0.589) at $p < 0.01$.