AN EMPIRICAL INVESTIGATION OF ONLINE WORD-OF-MOUTH DYNAMICS IN DIFFERENT ONLINE COMMUNITIES

Li Qin
Department of Decision Sciences and Information Systems
Silberman College of Business
Fairleigh Dickinson University
1000 River Road, Teaneck, NJ 07666

ABSTRACT

This paper makes an empirical investigation of the dynamics of online word-of-mouth (WOM) for the context of the movie industry by focusing on the interplay between different online communities as well as their predictive power for revenue. Empirical results found that the volume of word-of-mouth in different online communities helps predict each other, and has similar predictive power for revenue. Implications of the findings were discussed.

Keywords: word-of-mouth, WOM, social media, online communities

1. INTRODUCTION

Research shows that Word of Mouth (WOM) influences consumers' purchase decisions (Bynerjee, 1992; Brown and Reingen, 1987). The Internet and web 2.0 have created online communities where consumers interact and influence each other as well as businesses and consumers interact and influence each other. Such online WOM can take place in many forms of online communities such as online reviews, discussion boards, video sites, blogs, microblogs, social networks, and so on. A study on social media engagement of Business Week top 100 global brands showed that on average these 100 businesses engage their customers in 5.6 different social media channels (Wetpaint and Altimeter Group, 2009). Not only an increasingly large number of customers and businesses are engaged in online communities, but also a very large percentage of customers actively conduct word-of-mouth. For instance, according to Technorati 2008 State of Blogsphere Report, "Four in five bloggers post brand or product reviews, with 37% posting them frequently. 90% of bloggers say they post about the brands, music, movies and books that they love (or hate)."

Evidences for the interrelationship between WOM and sales have been discovered in prior studies. Findings indicate that WOM dispersion (Godes and Mayzlin, 2004), valence (Chevalier and Mayzlin, 2006), and volume (Liu, 2006) have significant impact upon product sales. Studies by Duan, Gu, and Whinston (2008) and Qin (2011) investigated the explanatory power between consumer purchase and WOM volume, and found evidence for the effect of online WOM as both a precursor to and an outcome of retail sales. From the managerial perspective, the relationship between online WOM and sales suggests that businesses should effectively manage online WOM to promote sales.

Each online community has its uniqueness in the dynamics of WOM distribution. For instance, for online discussion boards, consumers visit and post in a third-party website (e.g. http://movies.yahoo.com) and their WOM is distributed to those who visit the same website; blogs are dispersed in the cyberspace, yet they join an inter-connected community in which conversations are conducted between blogs as well as between blog authors and readers. As a result, businesses who are engaged in multiple online communities may have different strategies for managing different forms of WOM.

In order for businesses to make informed decisions on their strategies to manage different forms of online WOM, some questions are still to be answered. Such questions include: Do different types of online WOM/communities influence each other? Are different types of online WOM/communities complements to or substitutes of each other? Are different types of online WOM/communities equal? For example, do they have the same predictive power for sales? No prior work has addressed these questions which involve different online communities, and the focus of this paper is to empirically investigate the WOM dynamics in different online communities.

2. RESEARCH DESIGN

To examine the dynamics of different types of online WOM, this paper makes use of a time-series dataset containing daily WOM volume in two different online communities as well as daily revenue for movies. The movie industry has by far received the most attention in the WOM literature, and it is selected as the research context because WOM strongly influences people's movie selection (Bayus 1985; Faber and O'Guinn 1984; Neelamegham and Chintagunta 1999). Based on the study by Liu (2006) which involves both volume and valence of WOM, it was found that most of the explanatory power of WOM is from the volume. Therefore, WOM volume is the focus of this paper. It is the intent of the study to examine the WOM dynamics during the time window when the sales are high and it was found that the revenue for the movies in the sample during the first four weeks after the movie release constitutes 91% of their total revenue in the U.S. Therefore, this study focuses on the WOM dynamics during the first four weeks after the movie release.

In this study, two types of online communities were examined for their WOM dynamics: Online discussion board, and blogs. Note that online discussion board is a form of web 1.0 and blogs a form of web 2.0. The specific research goals are as follows: First, find out how active each form of WOM is on a daily basis after a new product is introduced. For this goal, we collected data on how many blogs (i.e. the volume of blog WOM) discuss each movie and how many posts in an online discussion board (i.e. the volume of discussion board WOM) discuss the same movie on a daily basis after the movie release; Second, investigate whether WOM volume in one online community helps predict that in another online community. That is, whether WOM volume in an online discussion board helps predict WOM volume in blogs, and whether WOM volume in blogs helps predict WOM volume in an online discussion board; Third, quantify the extent to which each type of online WOM volume helps explain the sales of a new product. Findings on the predictive power between WOM volumes in different online communities and between WOM volumes in different online communities and sales can provide insight for managers on developing their strategies on managing WOM in different online communities.

2.1. Data

The WOM data on blog posts in this study are collected from BlogPulse which provides the number of blog buzz by topic (i.e. the movie name in our context) on a daily basis. BlogPulse is an automated trend discovery system for blogs by Nielsen (http://nielsen.com). According to Nielsen's website, it "applies machine-learning and natural-language processing techniques to discover trends in the highly dynamic world of blogs". It is a blog search engine which creates a full-text search index of all of the blog entries it finds every day, analyzes and reports daily activity in the blogosphere. For each movie in the sample, the number of blog buzz collected from this website indicates the number of blogs which discuss the movie on a daily basis. Figure 1 shows the percent of all blog posts (The number of blog posts can be obtained by clicking on the chart) on the movie "Transformers: Revenge of the Fallen" (released on June 24, 2009) on a daily basis.

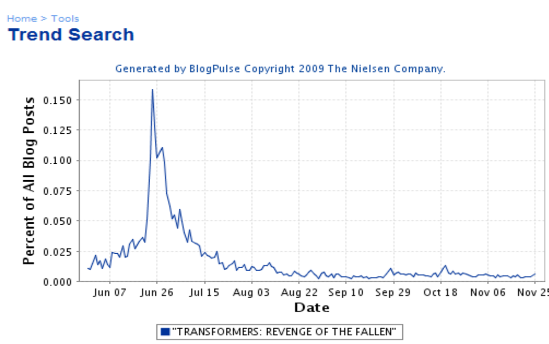


Figure 1. Trend Chart from Blogpulse.com for blog posts on movie "Transformers: Revenge of the Fallen"

The WOM data on discussion board posts are collected from Yahoo!Movies (http://www.movies.yahoo.com). The number of discussion board posts on a movie indicates the number of posts which discuss the movie on a daily basis on this movie discussion board.

Other information on movies including genre, MPAA rating, production budget, release date/age, number of screens, can affect WOM and movie revenue. These information along with daily box office revenue, were collected from publicly available sources: BoxOfficeMojo.com (http://www.boxofficemojo.com) and The Numbers (http://www.the-numbers.com/). Movies were chosen based on their box office rank in the U.S. market and all the movies in our sample were among the top 13 box office based on their opening weekend revenue.

Our data set included 49 movies with the release time in U.S. theaters between March 2009 and October 2009. Our final sample included a time-series cross-sectional dataset of the 49 movies for 4 weeks (28 days) after being released. Table 1 presents the summary statistics for the movie sample. Table 2 presents the description of the key variables used in our empirical model. Table 3 presents the summary statistics for the 4-week daily data.

Table 1 Summary statistics of the movie sample

Variable	N	Mean	Median	SD	Min	Max
Production budget*	36	71.98	40	66.94	7.5	250
U.S. gross*	37	83.04	55.25	91.16	5.21	402.11
Total # of blogs	49	4600	2679	4446.78	570	17151
Total # of discussion board posts	49	441	243	562	23	2774

*million $

Table 2 Key variables and descriptions

Variable	Description
$SCREENS_{it}$	Daily number of screens for movie i on day t
$DAILYGROSS_{it}$	Daily revenue for movie i on day t
$DAILYBLOG_{it}$	Number of blogs for movie i on day t
$DAILYBOARD_{it}$	Number of posts for movie i on day t
AGE_{it}	Number of days movie i has been released on day t
$WEEKEND_{t}$	A dummy variable indicating if day t is a weekend (Friday/Saturday/Sunday)

Table 3 Summary statistics of the daily data

Variable	N	Mean	Median	SD	Min	Max
SCREENS	1372	2796.49	2895.5	849.96	27	4325
DAILYGROSS*	1372	2.49	0.94	4.65	0.0022	62.02
DAILYBLOG	1372	164.31	91.5	202.29	1	2064
DAILYBOARD	1372	16	5	38.91	0	564
AGE	1372	14.5	14.5	8.08	1	28

*million $

2.2. Methodology

Granger (1969) defined causality based on the idea that if a variable affects another variable, the former should help improving the predictions of the latter variable. In other words, a process X_t is said to Granger cause another process Y_t if future values of Y_t can be predicted better using past values of X_t and Y_t than using the past of Y_t alone. Conway et al. (1984) found Granger causality to be a useful tool when the knowledge of Y_t increases one's ability to forecast X_{t+1} in a least squares sense. Granger causality is intended for use in establishing the direction of influence in time series data. Note that Granger causality does not establish causation in the real sense, but measures whether one variable precedes and helps predict another variable.

To test whether x Granger causes y, the following standard model is used:

$$y_t = a_0 + \sum J_{j=1}\alpha_j y_{t-j} + \sum J_{j=1}\delta_j x_{t-j} + u_t$$

where j represents the number of lagged values of each of the variables, also called the order of vector autoregressive (VAR).

For instance, the following equation is used to test whether the volume of discussion board posts Granger-causes the volume of blog posts:

$$\log(DAILYBLOG)_{it} = a_0 + \sum J_{j=1}\log(DAILYBLOG)_{i-t-j} + \sum J_{j=1}\delta_j \log(DAILYBOARD)_{i,t-j} + u_t \quad (1)$$

To find out the predictive power for sales by each form of online WOM, Ordinary Least Squares (OLS) will be used to quantify the explanatory power of each form of online WOM volume. The following equation 1 is to use blog buzz volume to predict daily box office revenue:

$$\log(DAILYGROSS)_{it} = \alpha_0 + \sum J_{j=1}\alpha_j \log(DAILYBLOG)_{i,t-j} + \beta_1 \log(SCREEN)_{it} + \gamma_1 \log(AGE)_{it} +$$

$$\delta_1 WEEKEND_t + \rho_1 genre_i + \sigma_1 MPAArating_i + \varepsilon_{it}, \quad (2)$$

for each day separately (t = 1, 2, …, 28), where i indexes the movies (i = 1, 2, …, 49).

Similarly, the following equation is to use discussion board volume to predict daily box office revenue:

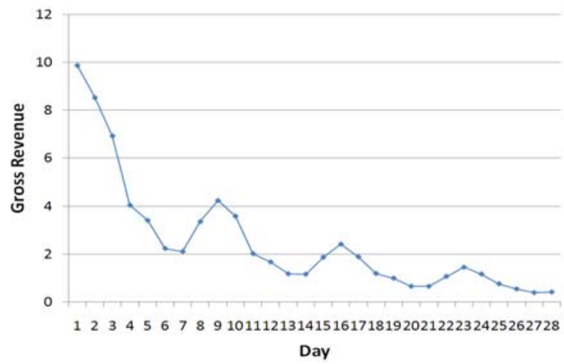
$$\log(DAILYGROSS)_{it} = \zeta_0 + \sum J_{j=1}\zeta_j \log(DAILYBOARD)_{i,t-j} + £_1 \log(SCREEN)_{it} + \varrho_1 \log(AGE)_{it} +$$

$$\chi_1 WEEKEND_t + €_1 genre_i + \eta_1 MPAArating_i + \ell_{it}, \quad (3)$$

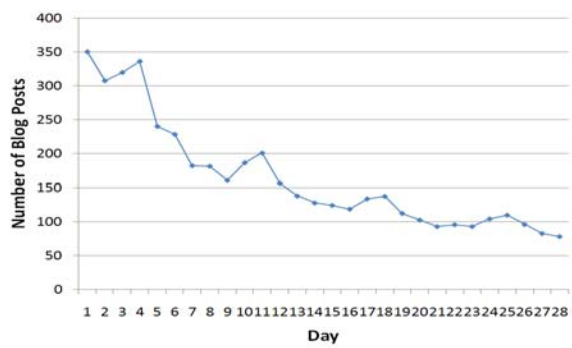
for each day separately (t = 1, 2, …, 28), where i indexes the movies (i = 1, 2, …, 49).

3. FINDINGS

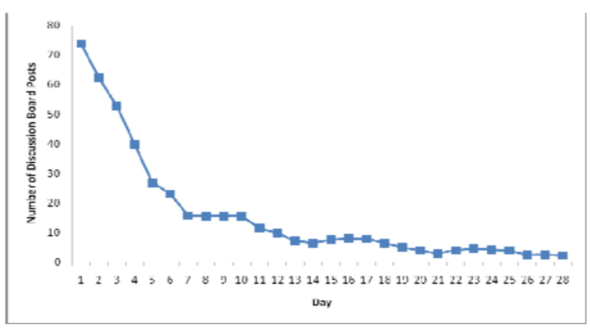
Figure 2 plots how average box office revenue, average blog buzz volume, and average discussion board buzz volume change over time on a daily basis, respectively. The overall trend is that box office revenue, blog buzz volume, and discussion board buzz volume decrease over time after movie release.



(a) Average box office revenue on a daily basis



(b) Average number of blog posts on a daily basis



(c) Average number of discussion board posts on a daily basis

Figure 2. Dynamics of blog WOM, discussion board WOM, and box office revenue for the movie sample

Granger Causality tests were conducted with two lagged values of word-of-mouth volume for both blogs and discussion board posts included. Results show that blog buzz volume granger causes discussion board buzz volume, and discussion board buzz volume granger causes blog buzz volume. That is, the volume of blog posts helps predict the volume of discussion board posts, and the volume of discussion board posts helps predict the volume of blog posts.

For the predictive power of word-of-mouth volume for revenue, the OLS results for equation 1 and 2 are shown in Table 4 and 5, respectively. Table 6 shows the results for predictive power for revenue when both blog WOM and discussion board WOM are included as independent variables.

Table 4 Results to use blog WOM volume to predict revenue

Variable	Coefficient (Standard Error)
$\log(DAILYBLOG)_{it}$	0.50(0.02)*
$\log(SCREENS)_{it}$	1.00(0.01)*
$\log(AGE)_{i}$	-0.43(0.03)*
$WEEKEND_{t}$	0.96(0.04)*
$Genre_{i}$	-0.05(0.01)*
$MPAARating_{i}$	0.09(0.02)*
CONSTANT	3.56(0.22)*
R^2	0.84

*p<0.01

Table 5 Results to use discussion board WOM volume to predict revenue

Variable	Coefficient (Standard Error)
$\log(DAILYBOARD)_{it}$	0.51(0.02)*
$\log(SCREENS)_{it}$	0.95(0.03)*
$\log(AGE)_{i}$	-0.28(0.03)*
$WEEKEND_{t}$	0.81(0.04)*
$Genre_{i}$	-0.03(0.01)*
$MPAARating_{i}$	0.17(0.03)*
CONSTANT	4.72(0.23)*
R^2	0.82

*p<0.01

Table 6 Results to use both blog WOM volume and discussion board WOM volume to predict revenue

Variable	Coefficient (Standard Error)
$\log(\text{DAILYBOARD})_{it}$	0.29(0.03)*
$\log(\text{DAILYBLOG})_{it}$	0.38(0.02)*
$\log(\text{SCREENS})_{it}$	0.90(0.02)*
$\log(\text{AGE})_{i}$	-0.27(0.03)*
WEEKEND_{t}	0.88(0.04)*
Genre_{i}	-0.03(0.01)*
MPAARating_{i}	0.14(0.02)*
CONSTANT	3.84(0.21)*
R^2	0.92

The OLS results indicate that blog buzz volume and discussion board buzz volume have similar predictive power for box office revenue, though using blog buzz volume leads to slightly higher R^2 in predicting the revenue. When including both blog WOM and discussion board WOM as independent variables to predict the revenue, the predictive power has improved R^2 from 0.82/0.84 to 0.92.

4. IMPLICATIONS, LIMITATIONS AND FUTURE WORK

This paper makes an empirical investigation of the dynamics of online word-of-mouth for the context of the movie industry, by focusing on the predictive power between different online communities as well as between different online communities and revenue. It was found that word-of-mouth in different online communities has predictive power for each other, and has similar predictive power for revenue.

The finding that blog buzz volume and discussion board buzz volume granger cause each other presents evidence that buzzes in different online communities interact with each other. The WOM volume on one online community can be used for predicting the WOM volume on another online community. Such finding can help managers to plan their efforts on creating or managing buzz in different online communities. Since different forms of online word-of-mouth display similar predictive power for revenue, managers can make more informed decisions regarding the selection of online WOM to predict revenue.

Similar to the limitation of many studies, the blog data used in this study are restricted to blogs indexed by BlogPulse and the discussion board data are restricted to posts on Yahoo movie website. The results should be interpreted with this limitation.

For future research, the current study can be extended by examining the other forms of online buzz such as online social network, or the other characteristics of online WOM such as valence, or the other aspects of the interplay between different online communities.

References

Bayus, B. (1985) Word of Mouth: The Indirect Effect of Marketing Efforts. Journal *of Advertising Research*, 23(3), pp. 31-39.

Baker, B. (2009) Your customers are talking to everyone. *Information Management*, May.

Brown, J. and Reingen P. (1987) Social Ties and Word-of-Mouth Referral Behavior. *Journal of Consumer Research*, 14 (December), pp. 350-62.

Bynerjee, A. (1992) A Simple Model of Herd Behavior. *Quarterly Journal of Economics*, 107 (3), pp. 797-817.

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

Duan, W., Gu, B., and Whinston, A. B. (2008) The dynamics of online world-of-mouth and product sales-An empirical investigation of the movie industry. *Journal of Retailing*, 84 (2), pp. 233–242.

Faber, R. and O'Guinn, T. (1984) Effect of Media Advertising and Other Sources on Movie Selection. *Journalism Quarterly*, 61, pp. 371-77.

Godes, D. and Mayzlin, D. (2004) Using Online Conversations to Study Word of Mouth Communication. *Marketing Science*, 23 (Fall (4)), pp. 545–60.

Granger, C.W.J. (1969) Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37, pp. 424-438.

Liu, Y. (2006) Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue. *Journal of Marketing*, 70, pp. 74–89.

Neelamegham, R. and Chintagunta, P. (1999) A Bayesian Model to Forecast New Product Performance in Domestic and International Markets. *Marketing Science*, 18(2), pp.115-36.

Qin, L. (2011) Word-of-Blog for Movies: A Predictor and an Outcome of Box Office Revenue? *Journal of Electronic Commerce Research*, Vol. 12, No. 3, pp.187-198.

Wetpaint and Altimeter Group (2009) The world’s most valuable brands. Who’s most engaged? July.