Association for Information Systems
AIS Electronic Library (AISeL)

ICIS 2010 Proceedings

International Conference on Information Systems (ICIS)

2010

CHATTER MATTERS: HOW TWITTER CAN OPEN THE BLACK BOX OF ONLINE WORD-OF-MOUTH

Huaxia Rui
University of Texas at Austin, ruihuaxia@mail.utexas.edu

Yizao Liu
University of Texas at Austin, yizaoliu403@mail.utexas.edu

Andrew B. Whinston
University of Texas at Austin, abw@uts.cc.utexas.edu

CHATTER MATTERS: HOW TWITTER CAN OPEN THE BLACK BOX OF ONLINE WORD-OF-MOUTH

Completed Research Paper

Huaxia Rui
University of Texas at Austin
CBA 6.426, 2100 Speedway
Austin, TX 78712
ruihuaxia@mail.utexas.edu

Yizao Liu
University of Texas at Austin
1 University Station, C3100
Austin, TX 78712
yizaoliu403@mail.utexas.edu

Andrew B. Whinston
University of Texas at Austin
CBA 6.426, 2100 Speedway
Austin, TX 78712
abw@uts.cc.utexas.edu

Abstract

We collect Word-of-mouth (WOM) data on movies from Twitter and employ both a time-series model and a dynamic panel data model to study the influence of WOM on movie box office revenues. Compared with most previous literature that measures WOM through its volume or dispersion, we directly measure the number of recipients of each WOM message using the unique social structural information on Twitter. Thereby we offer a more direct study of WOM and provides a powerful evidence of the causal effect of WOM on product sale which is rarely dealt with in the literature. We also disentangle the different roles of pre-consumption WOM and post-consumption WOM for the first time in the literature and we find that the percentage of pre-consumption WOM has significant explanatory power. Although previous studies conclude that the valence of WOM does not have any explanatory power for movie box office revenue, our time-series based analysis suggests that valence of WOM does play an important role. Our conceptual model and empirical study shed lights on how WOM actually influences product sales and this paper also reveals the value of social networking sites like Twitter to both marketing researchers and practitioners.

Keywords: word-of-mouth, social networks, twitter

Introduction

Word-of-mouth (WOM) is the oldest and probably one of the most important channels of information diffusion among people. For the purchase of a new product or new service, WOM is often considered to be the most credible information source to consumers (Katz and Lazarsfeld 1955). While practitioners are experimenting with strategies such as buzz management, viral marketing, and referral programs to harness the power of WOM, researchers have also been actively studying the influence and management of WOM (Van den Bulte and Lilien 2001, Godes and Mayzlin 2009, Sonnier, McAlister, and Rutz 2010). However, direct measurement of WOM influence has always been challenging because information transmission often is not observable to researchers. The advent of the Internet era has significantly changed this situation by offering researchers the opportunity to study online WOM, which is an easily obtainable subset of WOM. For example, some researchers have used WOM conversations from Usenet to study its influence on TV ratings (Godes and Mayzlin 2004), and others have used WOM posts from Yahoo!Movies to study its influence on movie box office revenues (Liu 2006, Dellarocas, Zhang, and Awad 2007, Duan, Gu, and Whinston 2008, etc). These studies have yielded interesting results and important insights on how WOM influences product sales. However, they all suffer from the lack of structural information about the underlying social networks, which obviously play an important role in the function of WOM. Consider a simple scenario where two users post the same message online about a movie. If the first user's message is received by 10 people while the second user's message is received by 10,000 people, then the effects of the two messages are quite different. If we treat these two messages with equal weight, as most previous studies have done, we miss some important information. We believe part of the reason why such a critical factor has not been taken into account in previous studies is because of the lack of social network data in online forums like Yahoo!Movies. This situation has now been changed by the explosive growth of online social networking services like Twitter and Facebook. In particular, the openness of Twitter allows us to extract rich WOM information to study the influence of WOM on product sales.

Launched publicly in July 2006 as an open social networking and micro-blogging service, Twitter is one of the fastest-growing social network sites in 2009 and has 105 million registered users by April 2010.[1] Users can use Twitter to post and read messages known as tweets (also known as updates), which are text-based posts of up to 140 characters. According to Twitter, users typically write about 65 million tweets a day. A user's followers are those users who subscribe to receive the user's tweets. Twitter is a great venue for the purpose of letting people express themselves and exchange information online. The potential value of Twitter as a marketing tool is increasingly being recognized (Rui, Whinston, and Winkler 2009). In particular, Twitter offers an excellent opportunity for researchers to study WOM for several reasons. First, compared with online forums like Yahoo!Movies, Twitter provides a more natural environment to study the awareness effect of WOM. The awareness effect of WOM on product sales refers to its function of spreading basic information about the product among the population. As the name suggests, the awareness effect influences people's behavior only by informing them and thereby putting the product in their choice set. This influence is in contrast to the so-called persuasive effect which refers to WOM's function of altering people's preferences toward the product and eventually influencing their purchase decisions. Because people who visit online forums like Yahoo!Movies to find out movie review information are most likely already aware of these movies, the awareness effect of WOM there is quite limited. On the other hand, WOM generators on Twitter are actually pushing their tweets to their followers. The differences between the “pull” mode on Yahoo!Movies and the “push” mode on Twitter make Twitter a better environment for researchers to study the awareness effect of WOM. Second, unlike many online forums where no social structural information is available, Twitter provides an Application Program Interface (API) structure with which we can extract the number of followers each author has. This seemingly simple information is actually very important for the study of WOM because it allows us to know the exact number of recipients of each message. The number of followers a Twitter user has is like the size of her audience. The more followers she has, the more people she can reach and the larger the influence of her WOM. This unique feature of Twitter WOM data not only enables us to calibrate the effect of each WOM message but also allows us to go beyond statistical correlation and say more about the causality between variables. We elaborate on this point in the theory background section. Third, while most previous literature focuses on the study of post-consumption WOM (i.e., WOM generated by people who have consumed the product), we deliberately disentangle the different effects of post-consumption WOM and pre-consumption WOM (i.e., WOM generated by people who have not consumed the product). Previous literature seems to suggest that all the WOM after the release of a movie

[1] http://mashable.com/2010/04/14/twitter-registered-users/

is post-consumption WOM. However, our data from Twitter suggests otherwise. [2] People on Twitter frequently talk about their plans or intentions of taking certain actions, like watching a movie or eating a certain breakfast. Interestingly, the triviality of information on Twitter, which many people have criticized, could be extremely valuable information for companies, as well as for researchers. Fourth, because of its simplicity and popularity, there is a huge number of tweets on a vast number of topics. For example, on March 4, 2010, one day before the release of the movie “Alice in Wonderland”, there were 14,738 tweets about this movie. On February 18, 2010, two months after the release of the movie “Avatar”, there were still 12,729 tweets about it. In our empirical study, we use a total of 5,980,586 tweets about 63 movies, which is significantly more than the 12,136 posts used in Liu (2006) and the 95,867 posts used in Duan, Gu, and Whinston (2008). The large number of WOM messages means that we may have less bias in our sample than in the samples used in previous literature.

With the huge collection of WOM data from Twitter, we use both a vector autoregressive model with exogenous variables (VARX) and a dynamic panel data model to study the influence of WOM on movie box office revenues. VARX is an extended version of the well-known VAR model which is particularly powerful when it is used to study the interdependencies between several time series. There are many VAR applications in the marketing literature. [3] Recently, a VAR model has been used by researchers to study the effect of WOM on stock prices in the U.S. airline industry (Luo 2009) and on member sign-up for Internet social networking sites (Trusov, Bucklin, and Pauwels 2009). Since movie sales both drive the WOM and are influenced by WOM, VARX can fully capture the dynamic interactions between the movie box office revenues and other endogenous variables characterizing WOM.

Overall, our study adds several important contributions to the literature. First, by measuring the number of recipients of WOM using social structural information on Twitter, our paper offers a more direct study of WOM [4] and provides powerful evidence of the causal effect of WOM on product sales which is rarely dealt with in the literature. Second, unlike the previous literature, which does not disentangle the pre-consumption WOM and the post-consumption WOM after a movie is released, we use a model that explicitly accounts for their different roles in explaining movie box office revenues, and find that the percentage of pre-consumption WOM of total WOM has significant explanatory power. Third, our research offers data- and methodology-related advances. For example, we demonstrate through the use of Twitter data how researchers can make use of WOM data from social networking sites. We also propose the use of VARX model to study the dynamic interactions between WOM and movie box office revenues. Several previous studies have recognized the importance of the endogeneity problem while studying WOM influence on movie box office revenues (Godes and Mayzlin 2004, Duan, Gu, and Whinston 2008), but none of them has used the VARX model, which is well suited to capture the feedback loops that affect movie box office revenues over time.
The paper is organized as follows. We briefly review relevant literature in the next section and then develop the theoretical background. After that, we describe our data in detail and introduces the methodology. Then we present our empirical results with some discussion. Finally, we conclude our paper and point out future research directions.

Literature Review

The literature on WOM in general is vast. More relevant to this paper is the literature on the influence of online WOM on product sales. Among this stream of literature, Godes and Mayzlin (2004), Liu (2006), and Duan, Gu and Whinston (2008) are most relevant to this paper.

Godes and Mayzlin (2004) is one of the first papers studying online WOM. They collected WOM information on 44 TV shows during the 1999 to 2000 season from the Usenet newsgroup. The WOM information is then used in a

[2] To be specific, we find about 8% of the tweets in our sample (i.e., 516,168 tweets out of 5,980,586 total tweets) talking all about people's intention to watch certain movies.

[3] See Dekimpe and Hanssens (1999), Bronnenberg, Mahajan, and Vanhonacker (2000), Nijs, Srinivasan, and Pauwels (2007) for examples of VAR applications in marketing. Pauwels et al (2004) provides an overview of time-series econometrics in marketing.

[4] Studying WOM through the use of WOM volume or WOM dispersion is an indirect approach for reasons explained before.

panel data model to explain the ratings of those TV shows. They identified the explanatory power of the entropy of conversations across newsgroups, which is a measure of the dispersion of conversations across newsgroups. The study concluded that the volume of conversations does not have any explanatory power. This finding strongly supports the awareness effect of WOM because greater dispersion implies that information is spread to more communities, thereby reaching more people. Because the number of recipients of the WOM could not be observed, dispersion across communities is a reasonable proxy for the number of WOM recipients. Even though the WOM volume does not have a significant effect according to their results, it is an important control variable in their model. However, we should be cautious not to interpret the significance as causality because dispersion might be the cause as well as the outcome of TV ratings, even after the control of WOM volume.

To examine the influence of WOM in the movie industry, Liu (2006) collected 12,136 WOM messages on 40 movies released during May and September in 2002 from Yahoo!Movie. He included WOM volume and WOM valence (measured as a percentage of positive/negative WOM) in a cross-section study, and he found that most of the explanatory power of WOM information comes from the volume of WOM but not from its valence. Duan, Gu, and Whinston (2008) also collected WOM from Yahoo!Movie. Their sample included 95,867 posts on 71 movies released between July 2003 and May 2004. To capture the fact that WOM both influences and is influenced by movie sales, they developed a two-equation system and estimated that using a three-stage least-square procedure. Their results suggested that box office sales are significantly influenced by the volume of online postings but, again, not by the ratings of online postings which measure WOM valence. Both papers identified the explanatory power of WOM volume on movie box office revenues and suggested that WOM might have influence on product sales through the awareness effect.

There are also research works on the effect of online WOM on the sales of products other than TV and movies. For example, Chevalier and Mayzlin (2006) studied the effect of WOM on on book sales. Dhar and Chang (2008) studied the impact of user-generated-content on music sales. Trusov, Bucklin, and Pauwels(2009) studied the effect of WOM on member growth at an Internet social networking site where membership registration could be viewed as a special type of product sales. Sonnier, McAlister, and Rutz (2010) studied the effect of online communications on the sales of some durable goods from some company. Li and Hitt (2008) examined the self-selection biases in online product reviews due to the different preferences of early buyers and later buyers. Through a study of software adoption on the Internet, Duan, Gu and Whinston (2009) found that user reviews have no impact on user adoption of the most popular product, while having an increasingly positive impact on the adoption of lower ranking products. Analogous to the product adoption, Oh and Jeon (2007) studied phenomenon of membership dynamics in the open source community through the lens of Ising theory, which is widely accepted in physics.

Theory Background

There exist plenty of theoretical and empirical works supporting the idea that WOM impacts consumers' decisions which then affect product sales. For example, Banerjee (1992, 1993) and Bikhchandani et al. (1991) suggest that people are easily influenced by others' opinions and actions. Sometimes, rational agents may even ignore their private information and act based only on the information inferred from others' actions, which leads to the "herding" phenomenon. The herding theory provides an interesting perspective of looking at the effect of WOM on product sales. If action does speak louder than words, then the fact that more WOM is generated by the population implies that more people have chosen to purchase the product, which, according to the herding theory, should lead more people to purchase the product even if the WOM may contain lots of negative review. McFadden and Train (1996) developed a model of consumer learning when the benefits and costs of new products are not fully known. One interesting implication of their model is that learning from others diminishes the sales of "niche" products that appeal to a small share of the population and enhances the sales of products that appeal to a larger share. In other words, there is a minimum share of the population that a product must be able to benefit in order for learning from others to increase the sales of the product. For products that are bought only once, such as books and movies for most people, they claim the minimum share is half. Previous empirical research provides support for the positive relationship between volume of the WOM and product sales (Godes and Mayzlin (2004), Liu (2006), Duan, Gu, and Whinston 2008).

Based on these theoretical and empirical results, we propose the following hypothesis on the relationship between Twitter WOM volume and movie box office revenue.

Hypothesis 1 *The volume of Twitter WOM has significant explanatory power for box office revenue in the subsequent period.*

Existing literature uses WOM volume or dispersion to study the awareness effect of WOM. We go one step further by directly measuring the number of recipients, or followers, of each WOM message. We believe the number of recipients of WOM is a more direct and probably more informative explanatory variable than the volume of WOM in explaining the influence of WOM on product sales. More importantly, incorporating both WOM volume and WOM recipient numbers into the model enables us to say more about causality. Generally, correlation does not imply causality, and it is very difficult to draw clean inferences of causality with traditional econometrics (Godes and Mayzlin 2004). The advantage of including the number of WOM recipients into the econometric model is that, after we control for the WOM volume, the correlation between the number of WOM recipients today and the movie box office revenues tomorrow is more likely explained by a causal effect of WOM on movie sales. [5] On the other hand, with only WOM volume, such a causality argument is less convincing because there could be many factors that both drive high box office revenue and more WOM on the movie. In addition to its role as a control variable, WOM volume on Twitter also serves as a proxy of the intensity of the WOM offline or on other online discussion forums. Because the size of WOM recipients is another measure of WOM intensity, the theoretical support for Hypothesis 1 also applies when we use the size of WOM recipients in the study. Hence, we have the following hypothesis:

Hypothesis 2 *The number of Twitter WOM recipients has significant explanatory power for box office revenue in the subsequent period.*

We classify WOM as pre-consumption WOM and post-consumption WOM based on whether the author of WOM has consumed the product or not. [6] Pre-consumption WOM is generally about people's intention or plans to purchase the product, while post-consumption WOM is usually about people's experience and/or attitude towards the product after consumption. While both types of WOM are very important, they should be treated differently when they are used to explain movie box office revenues. By our definition, pre-consumption WOM are generated by potential customers who have explicitly expressed their willingness to purchase the product; the volume of these WOM has a direct effect on future product sales because these authors are more likely than average population to consume the product in the near future. On the other hand, the authors of post-consumption WOM have already consumed the product and are less likely than the average population to purchase the product again in the near future. [7] Hence, post-consumption WOM only has an indirect effect on future product sales while pre-consumption WOM has both direct and indirect effect on future product sales. Therefore, we would expect the percentage of pre-consumption WOM on Twitter to have a positive and significant effect on box office revenues after controlling the total WOM volume. We refer to the percentage of pre-consumption Twitter WOM as the intention tweets ratio and our third hypothesis links the intention tweets ratio and the movie box office revenue.

Hypothesis 3 *The intention tweets ratio has a positive and significant effect on box office revenue in the subsequent period.*

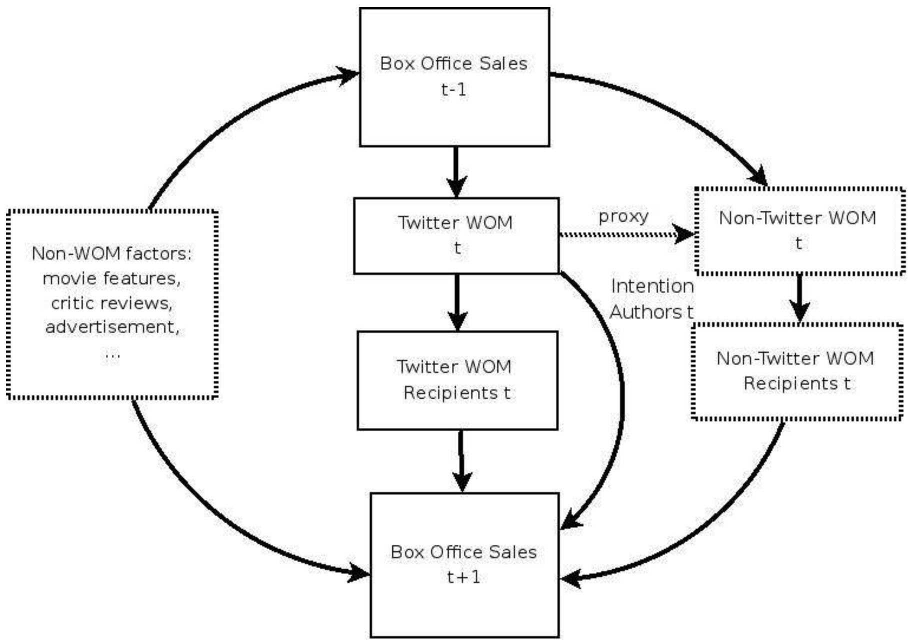
We summarize our conceptual framework in Figure 1 below. As a comparison, the conceptual framework used in most previous literature as we understand it is also shown in the figure. Notice that in the previous literature, WOM is treated more like a black box for its function of affecting future box office revenues while in our model we are able to separate the effects of WOM volume and the number of WOM recipients.

---

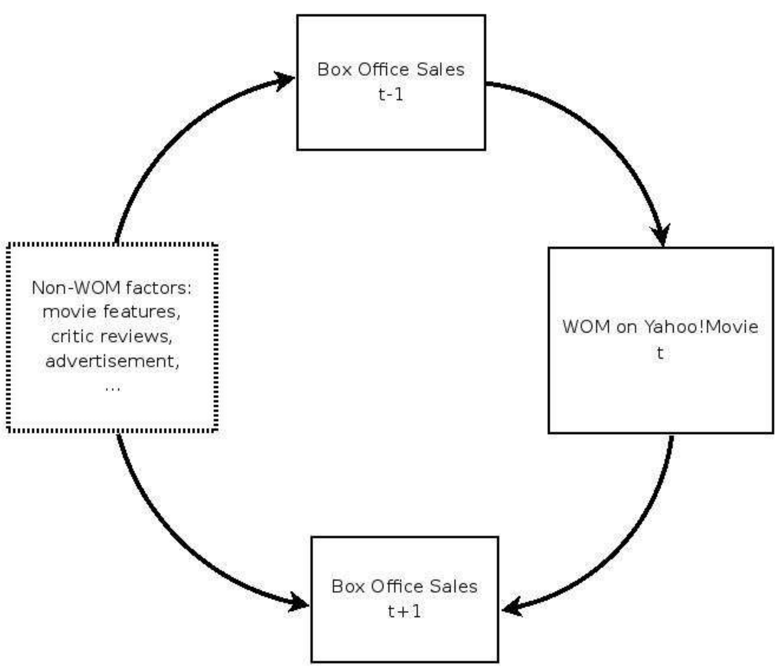
[5] An alternative explanation for the correlation is that higher box office revenue might cause or be related to something that causes people who have more followers on Twitter to tweet about the movie. Although we can not exclude this either theoretically or empirically, we regard this as unlikely

[6] There is actually a third category of online WOM, i.e., WOM generated during consumption. However, it is largely irrelevant to the approach in this paper. We thank an anonymous reviewer for pointing this out.

[7]This is true for many products like movies, books, and other durable goods.



Conceptual Framework of This Paper



Conceptual Framework of Some Previous Literature

Figure 1 Conceptual Framework

Model Specification

Data Description

We collect movie revenue information from BoxOfficeMojo.com [8] and tweet information from Twitter [9]. From BoxOfficeMojo.com, we collect daily box office revenue of movies that are widely released between June 2009 and March 2010. After excluding movies with incomplete data during this period, we use 63 movies in our final analysis.

Table 1: Description and Measures for Variables Used in Time Series Analysis

Variable	Description and Measure
$Revenue_{it}$	Gross Revenue of Movie i in Day t
$Total\,Author_{it}$	Total Number of Tweet Authors of Movie i in Day t
$Intention\,Author_{it}$	Total Number of Tweet Authors Whose Tweets Containing Intention of Seeing Movie i in Day t
$Intention\,Ratio_{it}$	Ratio of Intension Author among All Tweet Authors of Movie i in Day t
$Follower_{it}$	Total Number of Followers of All authors for All Tweets of Movie i in Day t
$Friday_{it}$	A Dummy Variable equals to 1 if Day t is Friday
$Saturday_{it}$	A Dummy Variable equals to 1 if Day t is Saturday
$Sunday_{it}$	A Dummy Variable equals to 1 if Day t is Sunday
$ReleaseWeek_{it}$	Number of Release Week of Movie i in Day t

Although obtaining daily movie revenue data is straightforward, collecting tweets (and author information) on those movies is tricky because of the real-time nature of the data and certain restrictions on API usage. Twitter provides a searching API that returns the most recent 1500 tweets containing the keyword specified by users. We use the name or part of the name of each movie in our sample as the keyword to query Twitter server. In order not to miss any tweets, we query tweets with each keyword once an hour. There are a total number of 5,980,586 tweets mentioning the above 63 movies in the collection. For each tweet, we observe the content of the tweet, the time when it is posted and the author's account name. From the account name, we can get the number of followers the author has.

After we collected the above information, we further aggregate the tweet information into daily summary statistics including the total number of tweets (or authors [10]) for each movie on each date, and the total number of followers authors have for each movie on each date. We also computed the ratio of intention tweets among all the tweets for each movie on each date. By intention tweets, we mean those tweets where the authors clearly express their willingness to watch the movie in the future. For example, the tweet "Wow! I wanna see the lovely bones!!" is clearly an intention tweet. On the other hand, the tweet "DAMN IT!!!Didn't make it...Sold out tickets for Avatar!!!" is also an intention tweet even though it's not obvious at the first glance. To pick out those intention tweets, we first manually indentified 300 intention tweets from which about 20 rules were extracted to characterize the patterns of those tweets. We then wrote a simple computer program with regular expressions that match the patterns we found and the program effectively picked out most of the intention tweets.

Table 1 lists the description of the key variables we used in the analysis. Table 2 gives the summary statistics for all the key variables. In general, the average daily box office revenue for a movie is $1,476,325 and the average number

---

[8] http://www.boxofficemojo.com

[9] http://www.twitter.com

[10] Almost all authors tweet no more than one tweet per day per movie, the total number of tweets and the total number of authors are almost the same for each movie at each date

of tweets is 1,121 among which 112.4 are intention tweets.

For illustration, we present the results of time series analysis for three movies: “My Sister's Keeper” (Movie 1), “Julie & Julia” (Movie 2) and “The Blind Sid” (Movie 3). Table 3 gives the summary statistics for the three movies, respectively. From Table 3, it is clear that the variation among movies is significant. For example, average daily revenue is $ 713,003 for “My Sister's Keepe”, while for “The Blind Sid” it is $ 2,668,357. Correspondingly, the maximal number of daily tweets is 1962 for “My Sister's Keeper”and 4781 for “The Blind Side”

Table 2: Summary Statistics for Variables of All Movies

Variable	Mean	Min	Max
$Revenue_{it}$	1476325	739	62000000
$Total\ Author_{it}$	1121.01	1	13318
$Intention\ Author_{it}$	112.40	0	2977
$Intention\ Ratio_{it}$	0.09	0	1
$Follower_{it}$	652757.70	6	14000000
$Friday_{it}$	0.15	0	1
$Saturday_{it}$	0.15	0	1
$Sunday_{it}$	0.15	0	1
$Release\ Week_{it}$	5.20	1	15
Days in Theater	3622		

Time Series Approach

Many previous papers studied the effect of WOM on box office revenues in the movie industry using either a pooled OLS model or a panel data model. However, none of them used a time series approach to study specifically how a time series of WOM would affect movie revenues. One possible explanation for this absence would be the limitation of data in the previous papers. For example, the median of WOM volume for all movies in their opening week is only 49 in Liu (2006). The low volume of data and a lack of variation in explanatory variable would cause identification problems in the time series analysis. On the other hand, we have an average of 1,121 tweets per movie per day for 63 movies. This amount of detailed data for WOM allows the time series analysis to capture the dynamic nature of WOM and movie gross revenues.

Vector Autoregressive Model

Based on the daily data we have for movie box office revenue and tweet information, we first do a time series analysis for each movie. We model the joint determination of daily box office revenue ($Revenue_t$), intention author ratio ($Intention\ Ratio_t$), number of tweet authors ($Total\ Author_t$) and the number of followers for each tweet author ($Follower_t$) as a multivariate vector autoregressive model with exogenous variables.

Table 3: Summary Statistics of Key Variables for Selected Movies			
	My Sister's Keeper	Julie & Julia	The Blind Side
Variable	Mean	Mean	Mean
$Revenue_{it}$	713003.1	1402902	2668357
$Total\ Author_{it}$	357.5	699.52	981.86
$Intention\ Author_{it}$	63.28	86.67	133.97
$Intention\ Ratio_{it}$	0.13	0.10	0.11
$Follower_{it}$	157331.60	570528.40	780583.20
Days in Theater	68	61	93

$$Y_t = \begin{pmatrix} Revenue_t \\ IntentionRatio_t \\ Total\ Author_t \\ Follower_t \end{pmatrix} : \ VARX(1)$$

We estimate the following VAR(1) model for four variables jointly for each movie:

$$\begin{pmatrix} Revenue_t \\ IntentionRatio_t \\ TotalAuthor_t \\ Follower_t \end{pmatrix} = \begin{pmatrix} \alpha_r \\ \alpha_{ia} \\ \alpha_{ta} \\ \alpha_f \end{pmatrix} Releaseweek_t + \begin{pmatrix} \varphi_{11}^1 & \varphi_{12}^1 & \varphi_{13}^1 & \varphi_{14}^1 \\ \varphi_{21}^1 & \varphi_{22}^1 & \varphi_{23}^1 & \varphi_{24}^1 \\ \varphi_{31}^1 & \varphi_{32}^1 & \varphi_{33}^1 & \varphi_{34}^1 \\ \varphi_{41}^1 & \varphi_{42}^1 & \varphi_{43}^1 & \varphi_{44}^1 \end{pmatrix} \begin{pmatrix} Revenue_{t-1} \\ IntentionRatio_{t-1} \\ TotalAuthor_{t-1} \\ Follower_{t-1} \end{pmatrix}$$
$$+ \begin{pmatrix} \sum_{k=1}^{k} \beta_r^k X_{tk} \\ \sum_{k=1}^{k} \beta_{ia}^k X_{tk} \\ \sum_{k=1}^{k} \beta_{ta}^k X_{tk} \\ \sum_{k=1}^{k} \beta_f^k X_{tk} \end{pmatrix} + \varepsilon_t \qquad (1)$$

The descriptions of $Revenue_t$, $Total\ Author_t$, $Intention\ Ratio_t$ and $Follower_t$ are in Table 1. $Revenue_{t-1}$, $Intention\ Ratio_{t-1}$, $TotalAutho\ r_{t-1}$ and $Follower_{t-1}$ are the one period lag of those variables. Φ is (4×4) matrix of autoregressive coefficients. $ReleaseWeek_t$ is an exogenous variable capturing the time trend in the model because, in general, movie revenue experiences a declining trend over its life cycle and α is the coefficient for the time trend. We also include a vector of exogenous variables X_{tk} into the VARX model. These variables include the dummy variables indicating whether day t is Friday, Saturday or Sunday, and β^k is the corresponding coefficient. The (4×1) vector ε_t is a vector generalization of white noise:

$$E(\varepsilon_t) = 0 \qquad (2)$$

$$E(\varepsilon_t, \varepsilon_{\tau'}) = \begin{cases} \Omega & \text{for } t = \tau \\ 0 & \text{Otherwise} \end{cases} \qquad (3)$$

with Ω (4×4) symmetric positive definite matrix.

We estimate the four equations of the above model simultaneously using a Maximum Likelihood Estimation for each movie and then perform the Granger Causality Test after estimation.

Test of Granger Causality

Many previous studies treat WOM as exogenous, while in fact, WOM is often not only the driving force of consumer purchase but also the outcome of product sales. The causality between product sales and WOM works in both directions. Therefore, to capture the dual nature of WOM and to clearly understand the effect of WOM on product sales (movie revenue in this analysis), we formulate a multivariate Granger causality test of WOM and movie revenue after the previous VARX model estimation.

The Granger causality test is a technique for determining whether one time series is useful in forecasting another. Although ordinary regressions reflect only correlations, there is an interpretation of a set of tests that can also reveal something about causality.

Technically, y_t fails to Granger-cause x_t if

$$MSE[\hat{E}(x_{t+s} \mid x_t, x_{t-1}, ...)] = MSE[\hat{E}(x_{t+s} \mid x_t, x_{t-1}, ..., y_t, y_{t-1}, ...)] \qquad (4)$$

which means that y is not linearly informative about the future value of x.

To implement a Granger causality test, we assume a particular autoregressive lag length p and estimate:

$$x_t = c_1 + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + ... + \alpha_p x_{t-p} + \beta_1 y_{t-1} + \beta_1 y_{t-2} + ... + \beta_p y_{t-p} + u_t \qquad (5)$$

by OLS. We then conduct an F-test of the null hypothesis:

$$H_0 : \beta_1 = \beta_1 = ... = \beta_p = 0 \qquad (6)$$

We first calculate the sum of squared residues from (5): $RSS_1 = \sum_{t=1}^{T} \hat{u}_t^2$ and then compare this sum with the sum of squared residuals of a univariate autoregression for x_t: $RSS_0 = \sum_{t=1}^{T} \hat{e}_t^2$, . The test statistic is then written as:

$$S_1 = \frac{(RSS_0 - RSS_1)/p}{RSS_1/(T - 2p - 1)} \qquad (7)$$

We could then test whether WOM (y_{t-1}), the number of tweets in day $t-1$ for a movie, Granger-causes the total box office revenue of day t (x_t), and can determine whether the volume of WOM can help forecast box office revenues for a movie.

Dynamic Panel Data Approach

To capture the dynamic nature of the data, as well as the cross-sectional effect, and to make full use of the richness of our data, we further formulate and estimate a dynamic panel data model using the method of Arellano and Bond (1991).

We write dynamic panel data model with strictly exogenous variables and autoregressive specification of the form:

$$y_{it} = \alpha y_{i,t-1} + \beta'(L) x_{it}^{*} + \eta_i + \nu_{it} = \delta_{i'} x_{it} + \eta_i + \nu_{it} \qquad (8)$$

where the dependent variable y_{it} is the movie gross revenue for movie i at week t and $y_{i,t-1}$ is its own one-period lag value. x_{it}^{*} is a set of explanatory variable, including $Total\,Author_t$, $Intention\,Ratio_t$ and $Follower_t$. $\beta'(L)$ is a vector of polynomials in the lag operator. η_i is the unobserved movie-specific effects that capture the idiosyncratic characteristics for each movie, such as genre, production budget, marketing cost, and quality. By using the non-time-varying movie-specific effects, we would be able to control the unobserved heterogeneity across movies. The v_{it} are assumed to have finite moments and in particular $E(v_{it}) = E(v_{it}v_{is}) = 0$ for $t \neq s$. That is, we assume lack of serial correlation but not necessarily independence over time.

$$x_{it} = \begin{pmatrix} y_{i,t-1} \\ x_{it}^{*} \end{pmatrix}$$

is a ($k \times 1$) vector and the η_i are individual specific effects.

Following Arellano and Bond (1991), we estimate the above problem using an optimal GMM method. The GMM estimator of the ($k \times 1$) coefficient vector δ is

$$\hat{\delta} = (\overline{X}'ZA_N Z'\overline{X})^{-1}\overline{X}'ZA_N Z'\overline{y} \quad (9)$$

where $\overline{X}$ is a stacked $(T-2)$N × k matrix of observations on $\overline{x}_{it}$ and $\overline{y}$. Z is a (T-2) × (T-2)[(k-1)(T+1)+(T-1)]/2 block diagonal matrix whose *s*th block is given by $(y_{i1}...y_{is}x_{i1}^{*'}...x_{iT}^{*'})$, $(s = 1,...T-2)$. The alternative choice of A_N would produce one-step or two-step estimators (Arellano and Bond 1991).

Estimation Results

In this section, we discuss the estimation results using the data and models described in the previous section. Regression results for the time series VARX model using daily data are provided in the first subsection. Results for the dynamic panel data model using weekly cross-sectional data are provided at the end of this section.

VARX Model Estimation and Test Results

VARX Model

Before estimation, we first assess the model by a unit root test to check the stationarity of the four variables. An Augmented Dickey-Fuller (ADF) test is performed with the null hypothesis that the data generating process is random walk with unit root. In particular, we tested the following equation with one lag and time trend, where y_t is the time series to be tested and x_t is the time trend.

$$\Delta y_t = \alpha + \rho y_{t-1} + \sum_{i=1}^{k}\beta_i \Delta y_{t-i} + \gamma x_t + \varepsilon \quad (10)$$

The unit-root test results are presented in Table 4 for three movies for all four variables. Column 2, 4, 6 from Table 4 report the ADF test statistics for each variable. These values range from -4.265 to -9.28, which are less than the 5% critical value. Therefore, we can reject the null hypothesis of a unit root with a 95% confidence interval and the test results suggest that all four time series of the variables are stationary and do not cointegrate in equilibrium.

The optimal lag length of the VARX model is 1, according to Schwartz's Bayesian Information Criterion (SBIC). This result is not surprising considering the evanescent feature of Twitter. As stated on its homepage, Twitter is the place to "discover what's happening right now". A tweet posted several days ago might be considered old and its influence is probably minimal already.

We then preform the time series analysis for all 63 movies in our data set and present the estimation results for three of them for illustration only. From Table 5, the 1-period lag value of revenue has a positive and significant coefficient for all three movies, implying positive autocorrelations for daily box office revenues.

The coefficient for the total number of tweet authors is significant for all movies, supporting Hypothesis 1. However, the sign varies among movies. The total number of tweet authors has a positive effect on movie revenue for "My Sister's Keeper" while the effect for the other two movies is negative. This mixed effect of total volume of WOM is also true for all other movies in our sample: Among all movie revenue equations with significant effects for total tweet authors, 47% of them are positive and 53% are negative.

Table 4: Unit Root ADF Test Result

Variable	My Sister's Keeper		Julie & Julia		The Blind Side		Results
	ADF Test Stat	5 % C. V.	ADF Test Stat	5 % C. V.	ADF Test Stat	5 % C. V.	
Revenue	-6.286	-3.484	-7.525	-3.491	-8.08	-3.459	No Unit Root
Total Author	-9.263	-3.484	-9.28	-3.491	-5.286	-3.459	No Unit Root
Intention Ratio	-5.233	-3.484	-6.543	-3.491	-4.281	-3.459	No Unit Root
Total Follower	-6.24	-3.484	-4.265	-3.491	-6.948	-3.459	No Unit Root

The intention author ratio turns out to be a significant predictor of movie revenue in the subsequent period, which supports Hypothesis 3. This result is a strong indication of the value of recognizing people's intention through the analysis of Twitter data. It also suggests the potential opportunities of targeted advertising on Twitter.

Table 5: Estimation Results from Time Series Model for Selected Movies

Revenue Equation						
	My Sister's Keeper		Julie & Julia		The Blind Side	
Variable	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
$Revenue_{i,t-1}$	0.40	3.88	0.67	5.88	0.61	5.68
$Total\ Author_{i,t-1}$	749.26	3.16	-442.35	-2.57	-1009.92	-2.53
$Intention\ Ratio_{i,t-1}$	1240221.00	2.32	5977933.00	2.50	16900000.00	5.17
$Follower_{i,t-1}$	0.11	1.69	0.38	3.32	0.35	2.47
$Friday_{it}$	218696.70	2.79	752965.80	4.30	2138529.00	5.39
$Saturday_{it}$	112410.30	1.34	1058341.00	5.70	1877482.00	4.26
$Sunday_{it}$	117800.10	1.55	-61984.34	-0.29	-713222.30	-1.54
$ReleaseWeek_{it}$	-31093.26	-3.51	-87047.70	-3.08	-84587.79	-2.86

Hypothesis 2 is supported by the positive and significant result of $Follower_t$ for the three movies. This result is very important. If the number of followers associated with the tweets affects revenue positively, this factor is more like a one-way influence of WOM because it is very unlikely that the movie revenue would affect the total number of followers in any way after we control for the total number of tweet authors. This finding clearly offers strong support to the causality conclusion that WOM has a causal effect on product sales.

Another interesting result not shown in the table, is the estimation results of the movie "Year One". We find the coefficient of $Follower_t$ to be significant and negative, which is quite surprising because it means that the more people who are informed about the movie in this period, the less revenue the movie has in the next period. To discover the reason for this negative coefficient, we analyzed the tweets on this movie and found significantly higher percentage of negative WOM for this movie than most other movies. Hence, we conclude that the negative coefficient of $Follower_t$ is probably because of the large amount of negative WOM which is a strong support for the persuasive effect of WOM. Even though previous studies fail to find any explanatory power of WOM valence for movie sales, the persuasive effect of WOM should not be ignored. This finding poses both challenge and opportunities for managers who want to harness the power of social media.

The time trend indicator $ReleaseWeek_{it}$, as we expected, is negative and significant, suggesting the declining trend in daily box office revenues for each week. Also, there is a positive relationship between weekend dummy variables, Friday and Saturday, and daily box office revenues.

Since the VARX model is based on the covariance stationary assumption, we test the covariance stationarity of the model by testing the eigenvalue stability condition after estimation. For all movies, all the eigenvalues lie inside the unit circle; therefore, our VARX model satisfies the stability condition.

Granger Causality Test

Table 6 reports estimated Granger Causality test statistics, as described in the previous section, for revenue equation for three movies, respectively. In particular, we test whether the past values of the number of tweet authors, intention ratio, the number of followers and all exogenous variables as a whole would provide statistically significant information about future value of movie gross box office revenues. The null hypothesis is that the tested variable does not Granger cause movie revenue. For movie "My Sister's Keeper", the p-value is smaller than 5% for $Total\ Author_{i,t-1}$, $Intention\ Ratio_{i,t-1}$ and All, suggesting that we can reject the null hypothesis of no Granger Causality at 5% confidence level. The p-value for variable $Follower_{i,t-1}$ is 0.09, leading to a rejection of null hypothesis of no Granger Causality at 10% confidence level. Overall, for movie "My Sister's Keeper", the past value of the number of tweet authors, the number of tweets which express an intention of seeing the movie and the number of followers have causal effects on next day's box office revenue and are useful in forecasting movie revenue for "My Sister's Keeper". Similarly, for movie "Julie & Julia" and "The Blind Side", the p-value for testing variables are all smaller than 5%, suggesting a rejection of null hypothesis at 5% level and thus a existence of causal relationship between tested variables and movie box office revenues.

Table 6: Granger Causality Test for Revenue Equation

Revenue Equation						
	My Sister's Keeper		Julie & Julia		The Blind Side	
Variable	χ^2	P-Value	χ^2	P-Value	χ^2	P-Value
$Total\ Author_{i,t-1}$	9.98	0.00	12.17	0.00	6.38	0.01
$Intention\ Ratio_{i,t-1}$	5.38	0.02	6.15	0.01	26.67	0.00
$Follower_{i,t-1}$	2.85	0.09	13.17	0.00	6.08	0.01
All	23.62	0.00	32.67	0.00	28.51	0.00

Fit of the Model

In order to measure the optimality of time-series forecasts for our VARX model, we further perform a post-estimation comparison of forecast errors (measured by mean squared error) for different models, in order to assess the fit of our model. In particular, we estimate and compare the VAR models using the following specifications:

Alternative Model I:

$$Y_t = \begin{pmatrix} Revenue_t \\ Intention\,Author_t \\ Total\,Author_t \end{pmatrix} : \quad VARX(1)$$

Alternative Model II:

$$Y_t = \left(Revenue_t\right): \quad AR(1)$$

Alternative Model III:

$$Y_t = \left(Revenue_t\right): \quad ARX(1)$$

Here Model I is a three variable VARX(1) model without total follower; Model II is autoregression model AR(1) using movie gross revenue and its own one period lag alone and Model III is ARX(1) model for movie gross revenue with exogenous variables, such as weekend dummies and time trend.

Table 7 reports and compare the forecast error for all four models for three movie. The forecast error using original VARX(1) model are smallest among all models for all movies, which suggests that our choice of VARX(1) models perform best in terms of model fit.

Results from Dynamic Panel Data Model

Using the weekly cross-sectional data for 63 movies, we estimated the unbalanced dynamic panel data model. The explanatory variables we used are $Revenue_{t-1}$, $Total\,Author_t$, $Total\,Author_{t-1}$, $Intention\,Ratio_t$, $Intention\,Ratio_{t-1}$, $Follower_t$, $Follower_{t-1}$. The panel we used is unbalanced because some movies are in theaters for a longer time than others. In addition, the use of unbalanced panel may lessen the impact of self selection of movies in the sample.

Table 7: Forecast Error Comparison for Different Model Specifications

Model Specification	Forecast Error ($\times 10^{10}$)		
	My Sister's Keeper	Julie & Julia	The Blind Side
Original Model	0.429	1.84	14.8
Alternative Model I	0.446	2.01	15.9
Alternative Model II	0.636	3.92	31.7
Alternative Model III	0.579	2.06	19.6

Table 8: Estimation Results from Dynamics Panel Data for All Movies

Variable	**Estimate**	**t-stat**
$Revenue_{i,t-1}$	0.47	31.21

$Total\ Author_{i,t}$	437.82	4.12
$Total\ Author_{i,t-1}$	-555.76	-7.38
$Intention\ Ratio_{i,t}$	22900000	2.74
$Intention\ Ratio_{i,t-1}$	25500000	3.28
$Follower_{i,t}$	-0.07	-0.60
$Follower_{i,t-1}$	0.35	4.38
$Constant$	-2679281	-3.65

In Table 8, we report estimates for the dynamic panel data model. The estimation results are quite similar to the time series model for individual movies in terms of the sign. The coefficient for lagged weekly revenue is 0.47, which is positive and significant: the larger the previous week's gross revenue is, the higher the current week's revenue will be. For intention author ratio, we have positive and significant effects for both period t's intention ratio and its one-period lagged term. Therefore both the previous and the current week's percentage of intention tweets contribute positively to the current week's box office revenue, while the previous week's intention tweets ratio has slightly stronger impact on the revenue. The effects of total tweets author is still mixed: positive and significant effect for this period's value, but negative and significant estimate for the lagged variable. It is hard to tell the combined effects of total number of WOM on movie revenue, considering all movies at the same time. This result on $Total\ Author$ is consistant with the mixed effects we find from VARX time series model: among all movie revenue equations with significant effects for total tweet authors, 47% of them are positive and 53% are negative.

For the total number of follower, the influence is positive and significant for the number of followers during the last period, meaning the more recipients the WOM has for previous week, the higher box office revenue the movie has for this week, while for the number recipents of WOM for the current does have siginificant inflence on current week's movie revenue. In terms of the absolute effect, similar to the result for the intention author ratio, we have a much stronger influence of the lagged variable of followers than the contemporaneous one. This is very surprising considering the real-time feature of Twitter. One possible explanation for this is the fact that most of the sales for a movie in a week occurs during the weekend rather than during the middle of the week. Most of the movies are released during the weekend which is the start of a week in our data. Hence previous week's impact on this week's variables tends to be strong.

Conclusion

The goal of this paper is to use the extremely rich data on Twitter to study the influence of WOM on box office revenues. Our studies suggest that WOM does matter, and information retrieved from WOM on Twitter has significant explanatory power for box office revenues. Moreover, the data and model chosen in this study enabled us to conclude the causal effect of WOM on movie sales more convincingly, which significantly differentiates this study from previous literature. We also find additional explanatory power from pre-consumption WOM, which is particularly meaningful for managers. An important question that firms face in advertising is how to develop effective targeting strategies. The success of Google's AdWords is an example of how important targeted advertising is to firms. Our results suggest that Twitter is a natural environment where people express their intention to purchase certain products, and firms could potentially make use of this information to run targeted advertising on Twitter. This research also has other important managerial implications. For example, our time-series model could be easily turned into a forecasting model that managers could use to predict future product sales. Another important although not surprising implication of our study is that managers should identify influential people on Twitter to harness the power of WOM. Managers could use the number of followers a Twitter user has as a simple measure of influence, but it is quite possible to construct more sophisticated and more accurate measures of influence based on information available on Twitter.

Compared with the tremendous amount of data on Twitter, our paper only exploits a very small portion of it. With Twitter's easy-to-use API structure and its ever-growing popularity, marketing researchers and practitioners are really sitting on a goldmine, and digging into it could be particularly rewarding. We believe the following issues, which are also the limitations in this paper, could be promising directions to pursue in the future.

First, one implicit assumption we used in this paper is that people have the same influence on their followers. This is obviously a rather simplistic assumption. For the study of awareness effect of WOM, this assumption is valid. However, when studying the persuasive effect of WOM, this is inappropriate. Is it true that those who have a large number of followers are more influential among the followers because the number of followers itself signals the authority? Or maybe people who have fewer followers have greater influence among the followers because they might be closer? Furthermore, a person might have different capacities to influence each of his/her followers. The challenge of measuring people's influence is probably one reason why the persuasive effect of WOM is so difficult to study.

Second, sentiment analysis is another challenge in studying the persuasive effect of WOM and is also an important future research direction. On the one hand, we are happy to see large volumes of WOM data because it reduces the sample bias; on the other hand, analyzing people's attitudes becomes a challenge because manually checking each WOM message is obviously not feasible. We used a set of very effective rules to identify tweets that express intention of purchasing a product, but identifying more subtle attitudes remains a challenge. Currently, sentiment analysis is an active research field in computational linguistics and could be particularly useful in the next few years in marketing. We performed some simple sentiment analysis by examining people's intention, but more delicate sentiment analysis remains a challenge.

References

Arellano, M., and Bond, S. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies* (58:2), pp. 277-297.

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

Banerjee, A. 1993. “The Economics of Rumours,” *Review of Economic Studies* (60:2), pp. 309-327.

Bikhchandani, S.D., Hirshleifer, D., and Welch, I. 1992. “A Theory of Fads, Fashions, Custom and Cultural Change as Information Cascades,” *Journal of Political Economy* (100:5), pp. 992-1026.

Bronnenberg, B. J., Mahajan, V., and Vanhonacker, W. R. 2000. “The Emergence of Market Structure in New Repeat-purchase Categories: A Dynamic Approach and an Empirical Application,” *Journal of Marketing Research* (37:1), pp. 16-31.

Chevalier, J., and Mayzlin, D. 2006. “The Effect of Word of Mouth on Sales: Online Book Reviews,” *Journal of Marketing Research* (43:3), pp. 345-354.

Dekimpe, M. G., and Hanssens, D. M. 1999. “Sustained Spending and Persistent Response: A New Look at Long-term Marketing Profitability,” *Journal of Marketing Research* (36:4), pp. 397-412.

Dellarocas, C., Zhang, X., and Awad, N. F. 2007. “Exploring the Value of Online Product Reviews in Forecasting Sales: The Case of Motion Pictures,” *Journal of Interactive Marketing* (21:4), pp. 23-45.

Dhar, V., and Chang, E. A. 2008. “Does Chatter Matter? The Impact of User-Generated Content on Music Sales,” *working paper*.

Duan, W., Gu, B., and Whinston, A. B. 2008. “Do Online Reviews Matter? - An Investigation of Panel Data,” *Decision Support Systems* (45:4), pp. 1007-1016.

Duan, W., Gu, B., and Whinston, A.B. 2009. “Informational Cascades and Software Adoption on the Internet,” *MIS*

Quarterly (33:1), pp. 23-48.

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

Godes, D., and Mayzlin, D. 2009. "Firm-Created Word-of-Mouth Communication: Evidence from a Field Test," *Marketing Science* (28:4), pp. 721-739.

Katz, Elihu, and Paul F. Lazarsfeld (1955): Personal Influence, Glencoe, IL: Free Press.

Li, X., and Hitt, L. 2008. "Self Selection and Information Role of Online Product Reviews," *Information Systems Research* (19:4), pp. 456-474.

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

Luo, X. 2009. "Quantifying the Long-Term Impact of Negative Word of Mouth on Cash Flows and Stock Prices," *Marketing Science* (28:1), pp. 148-165.

McFadden, D.L. and Train, K.E. 1996. "Consumers' Evaluation of New Products: Learning from Self and Others," *Journal of Political Economy* (104:4), pp. 683-703.

Nijs, V. R., Srinivasan S., and Pauwels, K. 2007. "Retail-Price Drivers and Retailer Profits," *Marketing Science* (26:4), pp. 473-487.

Oh, W. and Jeon, S. 2007. "Membership Herding and Network Stability in the Open Source Community: The Ising Perspective," *Management Science* (53:7), pp. 1086-1101.

Pauwels, K., Currim, I., Dekipe, M.G., Ghysels, E., Hanssens, D.M., Mizik, N., and Naik P. 2004. "Modeling Marketing Dynamics by Time Series Econometrics," *Marketing Letters* (15:4), pp. 167-183.

Rui, H., Whinston, A.B., and Winkler, E. 2009. "Follow the Tweets," *Wall Street Journal*, November 30.

Sonnier, G. P., McAlister, L., and Rutz, O. J., 2010. "A Dynamic Model of the Effect of Online Communications on Firm Sales," *Working Paper*.

Trusov, M., Bucklin, R. E., and Pauwels, K. 2009. "Effects of Word-of-Mouth Versus Traditional Marketing: Findings from an Internet Social Networking Site," *Journal of Marketing* (73), pp. 90-102.

Van den Bulte, C. and Joshi, Y. V. 2007. "`New Product Diffusion with Influentials and Imitators," *Marketing Science* (26:3), pp. 400-421.