NET Institute*

www.NETinst.org

Working Paper #11-27

October 2011

Whose and What Chatter Matters? The Impact of Tweets on Movie Sales

Huaxia Rui	Yizao Liu	Andrew Whinston
University of Texas at Austin	University of Connecticut	University of Texas at Austin

* The Networks, Electronic Commerce, and Telecommunications ("NET") Institute, http://www.NETinst.org, is a non-profit institution devoted to research on network industries, electronic commerce, telecommunications, the Internet, "virtual networks" comprised of computers that share the same technical standard or operating system, and on network issues in general.

Whose and What Chatter Matters? The Impact of Tweets on Movie Sales *

Huaxia Rui †, Yizao Liu ‡, and Andrew Whinston §

Abstract

Social broadcasting networks like Twitter in the U.S. and "Weibo" in China are transforming the way online word-of-mouth is disseminated and consumed in our society. We investigate whether and how Twitter WOM affects movie sales using publicly available Twitter data and common text mining algorithms. We find that more Twitter WOM messages always leads to more movie sales, however, the magnitude of the impact crucially depends of whom the WOM messages are from and what the WOM messages are about. Measuring Twitter users' influence by their number of followers, we find that WOM messages from a more influential person is significantly larger than WOM messages from a less influential person. In support to some recent findings that valence does matter to the impact of WOM on product sales, we also find that the impact of positive Twitter WOM is larger than negative WOM. However, the most powerful Twitter WOM are those tweets where the authors express their intention to watch a certain movie. We attribute this to the dual effects of intention tweets on movie sales: the direct effect on movie sales through the WOM author's own purchase behavior, and the indirect effect on movie sales through either the awareness effect or the persuasive effect of the WOM on its recipients. These findings provide different perspectives of understanding WOM compared with earlier literature and have important managerial implications.

JEL Classification: M3, C2
Keywords: Twitter, word-of-mouth, dynamic panel data

---

*The authors thank the NET Institute (www.NETinst.org) for financial support.
†University of Texas at Austin, E-mail: ruihuaxia@mail.utexas.edu
‡University of Connecticut, E-mail: yizao.liu@uconn.edu
§University of Texas at Austin, E-mail: abw@uts.cc.utexas.edu

1 Introduction

Social broadcasting services/networks like Twitter in the U.S. and "Weibo" in China[1] are transforming the way online word-of-mouth is disseminated and consumed in our society. As a leading example of social broadcasting network, Twitter has witnessed explosive growth in the last few years. As of January 2011, there were nearly 200 million registered users on Twitter who post 110 million tweets per day [2]. While the rise of social broadcasting networks may have significant social and political impact, of great interest to the marketing professionals is whether the huge amount of messages (a.k.a. tweets) generated and consumed by the vibrant Twitter community have any effect on product sales. The purpose of this paper is to take an initial step into answering this important question.

Tweets can be viewed as a type of online word of mouth (WOM) with brevity and immediacy as their distinguishing features. 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) and online WOM is an important subset of WOM in the Internet era. 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. For example, Godes and Mayzlin used WOM conversations from Usenet to study its influence on TV ratings (Godes and Mayzlin 2004) and they found a measure of the dispersion of conversations across communities has explanatory power but the volume of WOM does not. Many researchers used posts from Yahoo!Movies to study the effect of WOM on movie box office revenues (Liu 2006, Duan, Gu, and Whinston 2008, Chintagunta, Gopinath, and Venkataraman 2010, etc) and results are mixed. For example, both Liu (2006) and Duan et al. (2008) found that the volume of

[1] We call these sites *social broadcasting* technologies because they each are simultaneously a social network and a broadcasting service/network.

[2] http://www.forbes.com/sites/oliverchiang/2011/01/19/twitter-hits-nearly-200m-users-110m-tweets-per-day-focuses-on-global-expansion/

reviews matters but the valence does not. This view is also partially supported by Dhar and Chang (2009) where it is found that future sales of a music album are positively correlated with the volume of blog posts about that album. On the other hand, Chintagunta et al. measured the impact of national online user reviews on designated market area-level local geographic box office performance of movies and their findings suggest that it is the valence that drives box office performance, not the volume. Recent study by Sonnier, McAlister, and Rutz (2011) seems to support this view although their results are based on the analysis of a different product category[3]. Adding to this debate is another study by Onishi and Manchanda (2010) where they found mixed results for volume and valence using Japanese blog and sales data of three different products. Researchers have also examined the effect of online WOM on the sales of products from other perspectives. For example, Chevalier and Mayzlin (2006) examined the effect of consumer reviews on relative sales of books at Amazon.com and Barnesandnoble.com and they found that an improvement in a book's reviews leads to an increase in relative sales at that site. 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 sale, and they found that WOM referrals have substantially longer carryover effects than traditional marketing actions and produce substantially higher response elasticities. Dellarocas, Zhang, and Awad (2007) focused on forecasting movie sales using movie review data from Yahoo!Movies.

Our study differs from previous WOM literatures in several important ways, which are closely related to the unique features of Twitter WOM. 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,

[3]Product names are not mentioned in the paper, but the products are characterized as durable search goods.

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 generated on Twitter are actually pushed to the followers of authors. The differences between the "pull" mode on Yahoo!Movies and the "push" mode on Twitter makes 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 may be important for the study of WOM for two reasons. First, 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 effect of her WOM. Second, the number of followers a user could be a coarse proxy of the user's social influence. The very same WOM message may have quite different impact on the recipients depending on whom the message is from. There is little debate as to whether chatter matters to firms and earlier literature has studied extensively the question of what kind of chatter matters, but whose chatter matters? The two step flow theory in sociology suggests that some people (opinion leaders) are more influential than others (imitators) and information often moves first from mass media to opinion leaders and then from opinion leaders to imitators (Katz and Lazarsfeld 1955, Gladwell 2000, Slywotzky and Shapiro 1993). Applying this theory to WOM leads to the hypothesis that WOM messages from certain people may have disproportionate influence on firm product sales. Surprisingly, there is little research in marketing science addressing this question with a notable exception of the paper by Van den Bulte and Lilien (2001) where they developed a model of innovation diffusion in markets with two segments. Inspired by the two step flow theory, we divide the WOM messages from Twitter into two types based on the influence of the authors, measured by the number of followers, and estimate their respective effects on movie sales. To the best of our knowledge, we are the first among the marketing literature on WOM to account for the personal influence of WOM messages. We find that WOM messages from a more influential person is significantly larger than WOM messages from a less influential person.

Third, while most previous literature focuses on the study of review-type WOM, we deliberately disentangle the different effects of post-consumption WOM (i.e., WOM generated by people who have consumed the product) and pre-consumption WOM (i.e., WOM generated by people who have not consumed the product) [4]. Pre-consumption WOM is generally about people's intentions or plans to purchase a product [5], while post-consumption WOM is usually about people's experience and/or attitude toward a product after consumption. While previous literature seems to suggest that all the WOM after the release of a movie is post-consumption WOM, this is not true for Twitter WOM. People on Twitter frequently talk about their plans or intentions of taking certain actions, like watching a movie or having breakfast. The intention may be expressed directly or indirectly. For example, when people expresses their intention to watch a movie, they may explicitly say that, or through complaints about tickets, traffic, etc. We believe the prevalence of pre-consumption WOM on Twitter poses new challenge but also offers new opportunities for managers and researchers. The main challenge comes from automatic identification of these tweets while an obvious advantage is the addition of this new dimension of WOM measurement besides valence. Intuitively, pre-consumption WOM should be treated differently when they are used to explain

[4] Liu (2006) considered pre-release WOM which is a subset of pre-consumption WOM, and found that it has significant explanatory power for aggregate box office revenue. However, pre-consumption WOM is not limited to pre-release WOM in the case of movie.

[5] For this reason, we use pre-consumption tweets and intention tweets interchangeably in this paper.

movie box office revenues. the authors of post-consumption WOM have already consumed the product and are less likely to purchase the product again in the near future [6], hence, post-consumption WOM affects future product sales indirectly through its awareness effect and/or persuasive effect. On the other hand, because the authors of pre-consumption WOM have not consumed the product and are more likely than the average population to consume the product in the near future, we would expect pre-consumption WOM to have both a direct and indirect effect on future product sales. However, it is hard to predict whether pre-consumption WOM would have a larger or smaller impact on movie sales compared with post-consumption WOM. On the one hand, the direct effect of pre-consumption WOM on movie sales seems to suggest a larger impact on movie sales, but on the other hand, pre-consumption WOM contains less information about product quality, which may results in smaller impact. Our empirical results suggest that the effect of pre-consumption WOM on movie sales is larger than post-consumption WOM, whether the post-consumption WOM is positive or not.

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 in Section 3, we use a total of 4,166,623 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. However, the cost of the large data size is that we have to rely heavily on computer programs to classify WOM messages into pre-consumption WOM and post-consumption WOM and further do sentiment analysis on post-consumption WOM. Although the application of sentiment analysis (or

[6]This is true for many products, like movies, and durable goods.

computational linguistics more generally) in marketing science is still nascent, we believe there is a big role for these technique in the future because of the rapid growth of user-generated-content.

Our WOM data is collected from Twitter.com and movie sales data is collected from boxofficemojo.com, both of which are publicly available. We use a dynamic panel data model to study the influence of WOM on movie sales to handle the endogeneity problem typical in the study of WOM impact on firm product sales.

The rest of the paper is organized as follows. We first describe in detail how we collect and process our data in 2. We then introduce our methodology in Section 3 and present our empirical results in Section 4. Finally, we conclude our paper and point out some future research in Section 5.

2 Data

2.1 Data Collection

We collect movie revenue information from BoxOfficeMojo.com[7] and tweet information from Twitter [8]. From BoxOfficeMojo.com, we collect daily box office revenues of movies that were widely released between June 2009 and February 2010 [9]. After excluding movies with incomplete data during this period, we use 63 movies in our final analysis.

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 [10] .We use multiple computers to query for tweets once an

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

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

[9]We excluded movie titles for which it is difficult to correctly identify tweets that were related to those movies. For example, it is very hard to distinguish tweets talking about the movie 2012 from tweets talking about the year 2012.

[10]Twitter streaming API would be more suitable for this purpose but they were not available at the time we started the data collection.

hour. There are a total number of 4,166,623 tweets mentioning the above 63 movies in the collection. For each tweet, which is a text-based message of up to 140 characters, we collect the content of the tweet, the time when it is posted, and the author's account name. From the author's account name, we can get the number of followers the author has [11].

2.2 Data Processing

After the tweets are collected into the system, a simple filtering program is periodically executed to filter out advertising tweets. While we used several rules to determine whether a tweet is an advertising tweet, the most effective one is simply by checking whether the tweet contains a URL. There are also some irrelevant tweets containing the search keyword but are not about the movies. This is particularly a problem if the movie name is a single word or a commonly used phrase. We first randomly select 200 tweets for each movie and manually check if there are irrelevant tweets. For some movies like "Ninja Assassin" and "Shutter Island", there is almost no irrelevant tweet because these two phrases are rarely used on Twitter in contexts other than those movies. However, for some movies like "Wolfman", "Hangover", and "It's Complicated", there are irrelevant tweets. To reduce those irrelevant tweets, we adopted two approaches. First, we used a movie lexicon containing words or phrases like *movie, cinema, film, theater, ticket* to pick out relevant tweets that contain words or phrases in the lexicon; Second, for each movie, we used a customized lexicon for those irrelevant tweets and eliminated tweets containing words or phrases in that lexicon. For example, for the movie "Hangover", if a tweet contains the phrase "suffering from a hangover" or the phrase "drunk", then that tweet is classified as an irrelevant tweet. If there are still tweets undetermined after these two procedures, we manually classify them.

After filtering out the advertising tweets, we classify a tweet into one of the four mutually exclusive categories: intention, positive, negative, and neutral. An intention tweet is a tweet

[11] A user's followers are the people who subscribe to receive the user's tweets.

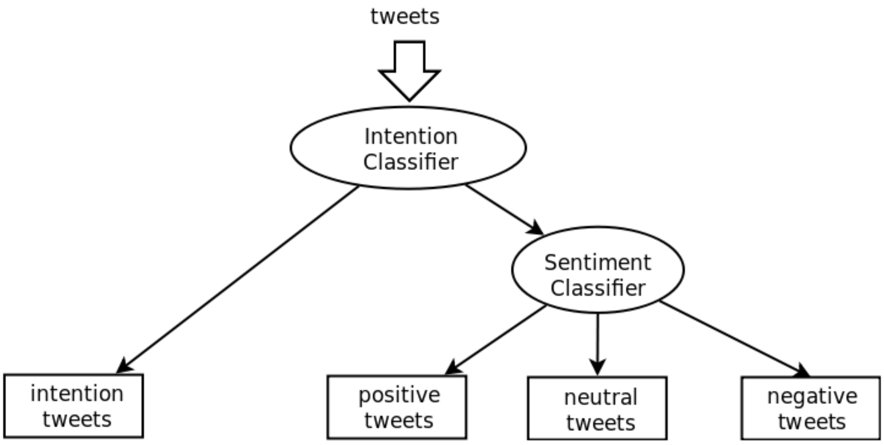

Figure 1: Tweet Classification

where the author expresses his/her intention to watch a certain movie in the future. A positive tweet is a tweet where the author expresses positive sentiment towards the movie. Similarly, a negative tweet is a tweet where the author expresses negative sentiment towards the movie. Neutral tweets are all other tweets that do no belong to any of the above three categories. Figure 1 illustrates the classification scheme.

We use an intention lexicon to extract features from tweets and then use a support vector machine (SVM) to construct the intention classifier. The intention lexicon is built from the movie tweets in our sample. For the sentiment analysis of tweets, we construct a Naive Bayesian classifier which draws upon a lexicon of positive words/phrases and negative words/phrases. Naive Bayesian classifiers are often used in the literature for text mining because of its simplicity. Of course there are more sophisticated classifiers for sentiment analysis in general which might yield higher accuracy. An in-depth study of these methods is not the focus of this paper. Both classifiers are trained and tested on a corpus of about 3,000 tweets that are manually labeled. The precisions and recalls for the intention classifier and the sentiment classifier are reported in Table 1 and Table 2 respectively.

Table 1: Precisions and Recalls of the Intention Classifier

	Precision	Recall
Intention tweets	94.4%	78.8%
Non-intention tweets	77.6%	86.2%

Table 2: Precisions and Recalls of the Sentiment Classifier

	Precision	Recall
positive	78%	85%
negative	62%	70%
neutral	82%	73%

2.3 Variables

Table 3 lists the description and measurements of the key explanatory variables we used and Table 4 provides the summary statistics for all the key variables. Gross Revenue is the movie gross revenues in a week. Here one week is defined from this Friday to next Thursday. On average, a movie's weekly box office revenue is around 9.5 million dollars. Weekend Gross Revenue is movie box office revenues for weekends only: Friday, Saturday, and Sunday. The average weekend gross box office revenue is around 6.3 million. The lagged Weekend Gross Revenue indicates the movie revenue of Friday, Saturday, and Sunday from the previous week. Since there is a declining trend for movie box office revenues in general, the average lagged Weekend Gross Revenue is higher than that of the present week. In our dynamic panel data analysis, we use Weekend Gross Revenue as our dependent variable, and use variables from the previous week as explanatory variables. There are several reasons why we use weekend gross revenues. First, most movie revenues are generated in the weekend. In fact, in our data sample, over 70% gross revenue for a movie comes from the weekend gross revenues only. Second, the time sensitivity of tweets. Tweet is more time sensitive than other types of WOM. When a tweet is posted, it will usually be seen by its follower immediately. If we include gross revenue from Monday to Thursday, we are actually trying to explain the gross revenue using WOM from over one week ago. It is less likely that a tweets from one week ago will still be seen because of the large volume of tweets everyday. Therefore, in our model, we focus on the analysis of weekend gross revenue only.

To capture the effects of WOM on movie box office revenue, we further include total number of tweets mentioning the movie's name as explanatory variables. In particular, we separate the tweets for each movie into two groups according to the power of their influences. On Twitter, the number of followers of each tweets is different among people. Celebrities like Oprah has more than six million followers while an ordinary user may only have two hundred followers. So 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? To test this, we use two groups to measure WOM's difference persuasive influence, which is never been studied in the literature. Specifically, the first group, Type 1 Tweets, includes tweets with total number of followers less than 650 in a week from This Friday to next Thursday. In our sample of 4,166,623 tweets, 90% of the tweets authors have less than 650 followers. The rest 10% of tweets are included into the second group, Type 2 Tweets. Those are tweets authors who has much more followers, thus are possibly more influential. On average, each movie has 6,049 Type 1 tweets and 578 Type 2 Tweets each week. The lagged Type 1 and Type 2 tweets represents tweets from last Friday to this Thursday.

We also computed the ratio of intention tweets among all the tweets for each movie in a week. 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

Table 3: Key Variables Used in Dynamic Panel Data Estimation

Gross Revenues	Movie gross box office revenues from Friday to next Thursday
Weekend Revenues	Movie gross box office revenues for Friday, Saturday, and Sunday only
Type 1 Tweets	Total number of tweets with followers less than 650 (fewer audiences) from Friday to next Thursday
Type 2 Tweets	Total number of tweets with followers more than 650 (more audiences) from Friday to next Thursday
Intention Tweets Ratio (%)	Total number of tweets showing intention of seeing movie i from Friday to next Thursday
Positive Tweets Ratio (%)	Ratio of tweets with positive comments in a week
Neutral Tweets Ratio (%)	Ratio of tweets with neutral comments in a week
Negative Tweets Ratio (%)	Ratio of tweets with negative comments in a week

it's not obvious at first glance. In our sample, almost 30% of the total tweets in a week express the intention of going to a movie, which compose a significant part among all tweets.

The effect of WOM's valence on sales has long been discussed in the literature, while the empirical evidences are still inconclusive till now. For example, Duan, Gu and Whinston (2008) find that the rating of online user reviews has no significant impact on movies' box office revenues. Chintagunta, Gopinath and Venkataraman (2010) suggest that it is the valence that seems to matter and not the volume when they measures the impact of national online user reviews on designated markets areas level local geographic box office performance of movies. In this analysis, we also make attempts to identify whether tweets valence will influence movie box office revenues. As explained in Section 2.2, three ratios are included: the ratio of positive tweets, neutral tweets, and negative tweets. Among all tweets for a movie in a week, 26.69% of them are positive, 38.03% of them are neutral and only 4.86% of those are negative.

Table 4: Summary Statistics of Key Variables for All Movies

Variable	Estimate	SD
Gross Revenue	9,435,003	21,000,000
Weekend Gross Revenue	6,285,750	11,900,000
Lag Weekend Gross Revenue	7,083,236	12,500,000
Type 1 Tweets	6,049.36	10,142.01
Lag Type 1 Tweets	5,474.57	9,682.17
Type 2 Tweets	577.91	1,081.56
Lag Type 2 Tweets	522.46	1,031.97
Intention Tweets Ratio (%)	30.42	9.61
Lag Intention Tweets Ratio (%)	29.89	9.80
Positive Tweets Ratio(%)	26.69	7.13
Lag Positive Tweets Ratio(%)	26.54	7.48
Negative Tweets Ratio(%)	4.86	3.84
Lag Negative Tweets Ratio(%)	4.81	3.72
Neutral Tweets Ratio(%)	38.03	10.35
Lag Neutral Tweets Ratio(%)	38.76	10.90
No. of Weekly Observation	572	

3 Model

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

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

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 variables, including Type 1 Tweets, Type 2 Tweets, Intention Tweets Ratio, Positive Tweets Ratio, and Negative Tweets Ratio. $\beta'(L)$ is a vector of polynomials in the lag operator to allow lagged value of explanatory variables to be included in the model. 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 ν_{it} are assumed to have finite moments, and, in particular, $E(\nu_{it}) = E(\nu_{it}\nu_{is}) = 0$ for $t \neq s$. That is, we assume a 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

$$\hat{\delta} = (\bar{X}' Z A_N Z' \bar{X})^{-1} \bar{X}' Z A_N Z' \bar{y}, \tag{2}$$

where $\bar{X}$ is a stacked $(T-2)$N $\times$ k matrix of observations on $\bar{x}_{it}$ and $\bar{y}$. Z is a (T-2) $\times$ (T-2)[(k-1)(T+1)+(T-1)]/2 block diagonal matrix whose sth 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).

4 Empirical Results

4.1 Results from Dynamic Panel Data Model

Using the weekly cross-sectional data for 63 movies, we estimate the unbalanced dynamic panel data model. The panel we used is unbalanced because some movies are in theaters longer than others. In addition, the use of the unbalanced panel may lessen the effect of self selection of movies in the sample.

In this panel data estimation, we also aggregate daily data into weekly data to shrink the number of time period (T). The focus of a typical dynamic panel data model is on panels where a large number of individuals are observed for a small number of time periods. However, in our analysis, we have a total number of 63 movies (N) while many of them are in theatre for over 90 days (T), which makes T far larger than N. This makes the estimation of the dynamic panel data model impossible. Therefore, we have to do this aggregation to shorten the number of time period. After the weekly aggregation, the longest time period for a movie is fifteen weeks.

In Table 5, we report estimates from the dynamic panel data model of weekend movie gross revenues. One week lag value of movie weekend revenue has positive effects of the current week's revenue, implying positive autocorrelations for daily box office revenues.

As acknowledged by many previous works, the relationship between movie box office revenue and WOM is intertwined. Higher volume of tweets will have awareness effect of a movie and thus will push up its revenue, while higher revenue of a movie may also in turn induce people to talk more about it. Therefore, there is a problem of potential endogeneity. To account for this, we use lagged value of tweets volume and tweets valence from the previous week as our explanatory variables. It is reasonable to believe that the previous week's WOM will have effects on this weekend's movie revenues but will not be affected by the future movie revenues. By doing this, we expect to lessen the endogeneity problem of the dynamic panel data model.

As expected, both types of tweets have positive and significant impacts on movie box office revenue. This suggests that WOM do have positive effects on movie box office revenue: the more people talk about the movie, the higher movie sales. This results is consistent with results from several previous works.

One advantage of our data is the ability to identify the distinct impacts of tweets authors with different number of followers, which is never been studied in the literature because of the data limitation. In this analysis, we find that for Type 1 tweets with fewer audience, 1 tweet will increase the weekend movie revenues by only $75.74 every week. However, for tweets with many audience, 1 tweet increase from the previous seven days will lead to almost $2,021 increase in weekend movie revenue every week. These results confirm the hypothesis that those who have a large number of followers (audiences) are more influential among the followers and have more impacts in pushing up movie box office revenues.

The intention tweets ratio turns out to be a significant predictor of movie revenues in the subsequent period. One percent increase in intention tweets will increase the weekend movie box office revenues by $57,137. 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 and marketing on Twitter.

In the literature, many works has tried to identify the effects of WOM valence, but most of their results lead to a conclusion that valence of WOM is indifferent and will not affect movie revenues. Different from literature, we found valence do have non-negligible impacts on movie box office revenues. Specifically, WOM with positive comments on the movie will increase movie sales significantly. One percent increase in positive tweets will increase the movie's weekend revenue by $ 17,841. Surprisingly, the negative tweet from the previous seven days also turns out to have significant and positive effects on movies' weekend box office revenue. However, the impact is smaller in magnitude, comparing to the effects of positive tweets. One percent increase in the volume of tweets with negative commons only increase the weekend gross revenues by $ 13,052. Though surprisingly, this results is not very hard to understand. Movies is experience good whose consumption is a one-time behavior most of the time. Most of the negative comments come from those have already seen the movie in theater. Although they are not speaking high of the movie, they still create awareness effects for their follower who will read the tweets. The higher proportion of consumers who know about the movie, the higher probability that they will go and see the movie. This is consistent with the saying in marketing from a long time ago: any publicity is good publicity.

4.2 Robust Check for Different Tweets Classifications

In the benchmark panel data model discussed in section **??**, we classify the total tweets into two types according to their total number of followers. The associated cut-off point we choose is 650, which is the 90% quantile of all tweets followers numbers. This means each Type 1 tweets has less than 650 followers and each Type 2 tweets has more than 650 followers. In this section, in order to check whether the estimation results are robust to different methods of classification, we run the same model six times using different tweets

Table 5: Estimation Results from Dynamics Panel Data for All Movies

Variable	Estimate	SD	P > $\|z\|$
Lag Weekend Gross	0.40	0.01	0.000
Type 1 Tweets	75.74	8.32	0.000
Type 2 Tweets	2,020.29	99.43	0.000
Intention Tweets Ratio (%)	57,136.67	2,196.51	0.000
Positive Tweets Ratio (%)	17,840.72	3,214.99	0.000
Negative Tweets Ratio (%)	13,051.69	1,650.80	0.000
Constant	-2,357,401	172,555	0.000

classifications. The cut-off point are 200, 300, 400, and 500. A cut-off point of 200 indicates each Type 1 tweets has less than 200 followers and each Type 2 tweets has more than 200 followers, et cetera. In order to run these models, we re-group the each raw tweet according to different followers cut-off points and then aggregate the data into daily and further weekly data for panel data estimation.

In Table 6, we present estimation results for six different robust models. In general, the estimation results do not vary significantly from the benchmark model results in Table 5, except for the coefficients of Type 1 Tweets. The top panel in Table 6 displays the estimates of model using 200 followers and 300 followers as cut-off points, respectively. The estimates of model using 400 followers and 500 followers as cut-off points are presented in the lower panel in the Table. As in the benchmark model, Intention Tweets Ratio, Positive Tweets Ratio and Negative Tweets Ratio all have positive and significant impacts of movies' weekend box office revenues for all models. For all models, the ratio of negative tweets has smaller effects on movie revenues than that of positive tweets. Type 2 tweets with relatively larger number of followers has noticeably higher impacts on movie revenue. However, the effects of Type 1 tweets now become negative and significant for all models with cut-off points 200,300, 400, and 500. This suggests that most movie revenues are driven by WOM with higher impacts and more audiences.

4.3 An Individual Movie Test

The panel data model specified in section **??** use all movies in our data set together to explore the general relationship between movie box office revenue, tweets number with different follower scope, intention tweets and tweets valence. This has also been used widely in the literature. While the panel data model focuses on a big and complete picture, we might also wonder whether the relationship we discovered work on individual movies. In order to confirm this, we perform a test of similar specifications on single movies using an Autoregressive (AR) Model.

Specifically, we formulate an AR(1) model of movie daily revenue:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta X_t + \epsilon_t \tag{3}$$

where the dependent variable Y_t is the dependent movie box office revenue in day t. the Y_{t-1} is the 1 day lagged value of Y_t, which is the movie revenue from previous day. X_t is a vector of exogenous variables that capture other features affecting movie revenue. In this test, we use the same set of variables defined in Table 3, which include number of two type tweets, intention tweets ratio, positive tweets ratio, and negative tweets ratio. ϵ_t is white noise, which satisfies

$$\begin{aligned}
&E(\epsilon_t) = 0\\
&Var(\epsilon_t) = \sigma^2\\
&Cov(\epsilon_t, \epsilon_{t-h}) = 0 \;\forall t, h \neq 0
\end{aligned} \tag{4}$$

Table 6: Robust Result Check from Dynamics Panel Data for All Movies

	Cut-off Point: 200 Followers			Cut-off Point: 300 Followers		
Variable	Estimate	SD	P > $\|z\|$	Estimate	SD	P > $\|z\|$
Lag Weekend Gross	0.57	0.05	0.000	0.57	0.05	0.000
Type 1 Tweets	-618.4802	1.76	0.000	-548.25	1.68	0.000
Type 2 Tweets	1776.06	5.65	0.000	2222.68	6.78	0.000
Intention Tweets Ratio	85917.03	1148.67	0.000	88512.07	1481.32	0.000
Positive Tweets Ratio (%)	55920.59	2889.25	0.000	60071.88	3104.95	0.000
Negative Tweets Ratio (%)	53412.13	1784.78	0.000	56928.66	2422.47	0.000
Constant	-3778658	138626.80	0.000	-3922695	164981.30	0.000

	Cut-off Point: 400 Followers			Cut-off Point: 500 Followers		
Variable	Estimate	SD	P > $\|z\|$	Estimate	SD	P > $\|z\|$
Lag Weekend Gross	0.57	0.04	0.000	0.58	0.04	0.000
Type 1 Tweets	-517.07	2.15	0.000	-502.21	2.14	0.000
Type 2 Tweets	2675.26	9.66	0.000	3140.60	11.55	0.000
Intention Tweets Ratio	89434.27	1312.37	0.000	89531.85	1422.58	0.000
Positive Tweets Ratio (%)	61937.92	3275.45	0.000	63387.07	3262.27	0.000
Negative Tweets Ratio (%)	56416.97	1777.95	0.000	57078.67	1646.61	0.000
Constant	-3888811	147550.20	0.000	-3876744	155112.70	0.000

Table 7 displays the results on individual movies using AR Model for two movies in the sample: “The proposal” and “Inglourious Basterds”. In this estimation, we use daily observation to perform the analysis, instead of using weekly data as in the panel data model. The reason is for a time series model in a single movie, a longer time period is better while a weekly summation will shrink the data.

The results in Table 7 confirm that the relationship we explore from the cross-sectional panel data also stand for individual movies. Lagged value of movie revenue will increase next day's revenue for both movie. The effects of Type 1 and Type 2 tweets turns out to be very different for the two movies in terms of magnitude. For movie “ The Proposal”, One Type 1 tweet with follower less than 650 increase the movie revenue by \$ 2,197.But the impact of Type 2 tweet is quite impressive: the daily revenue will increase by almost \$ 30,000 from only one more tweet increase. On the contrary, the effects of WOM are much smaller for movie “Inglourious Basterds”. The effects of Type 1 and Type 2 tweets on movie revenue are only \$978.45 and \$ 4,259.83, respectively. Same as the results in panel data model, the more tweets express the willingness to go to a movie, the higher movie box office revenue.

One point that is worth mentioning is the insignificance of parameter tweets valence (Positive Tweets Ratio and Negative Tweets Ratio), which is different from the results in cross-sectional data. In cross-sectional data model, we find that, positive WOM will increase movie revenue while the effect of negative WOM is insignificant. In this individual movie test, all valence parameters are insignificant. One possible explanation for this is the lack of variation of valence data for a single movie over its whole life. This is well explained in Figure 2. In Figure 2, we plot the dynamics of gross box office revenue, positive tweets ratio, and negative tweets for movie “ The Proposal”. The solid line is gross daily revenue, which is decreasing periodically over time. The dashed line and the dot dashed line are the positive tweets ratio and negative tweets ratio, respectively. These two line are basically constant over time: positive tweets ratio fluctuates slightly around 37% while negative tweets

Table 7: AR Model Test on Individual Movies

	The Proposal			Inglourious Basterds		
Variable	Estimate	SD	t-stat	Estimate	SD	t-stat
Lag Gross Revenue	0.62	0.15	4.04	0.67	0.14	4.97
Type 1 Tweets	2,197.32	959.25	2.29	978.45	118.74	8.24
Type 2 Tweets	29,905.67	6,190.62	4.83	4,259.83	798.22	5.34
Intention Tweets Ratio (%)	45,283.59	15,549.44	2.91	63,292.10	22,778.18	2.78
Positive Tweets Ratio (%)	-38,718.39	200,00.58	-1.94	-34,257.90	20,326.53	-1.69
Negative Tweets Ratio (%)	-1,150,00	170,000	-0.68	6,431.35	103,000	0.06
Weekend	1,088,340	358,356	3.04	970,989	357,972	2.71
No. of Daily Observation	58			62		

keep quite constant around 1.5%. These relative constant valence ratio over time cause the insignificance of the impact of valence on individual move revenue over time.

5 Conclusion

The goal of this paper is to investigate whether and how Twitter WOM, a recently popular and relatively new form of online WOM, affects movie sales. We collected Twitter WOM data using Twitter API and movie sales data from boxofficemojo.com, both of which are publicly available, we then carry out tweets classification and sentiment analysis using well-known algorithms in text mining. Having extracted variables characterizing Twitter WOM, we used a dynamic panel data model to explore the effect of Twitter WOM on movie sales. Our study adds several important contributions to the literature. We take a first step

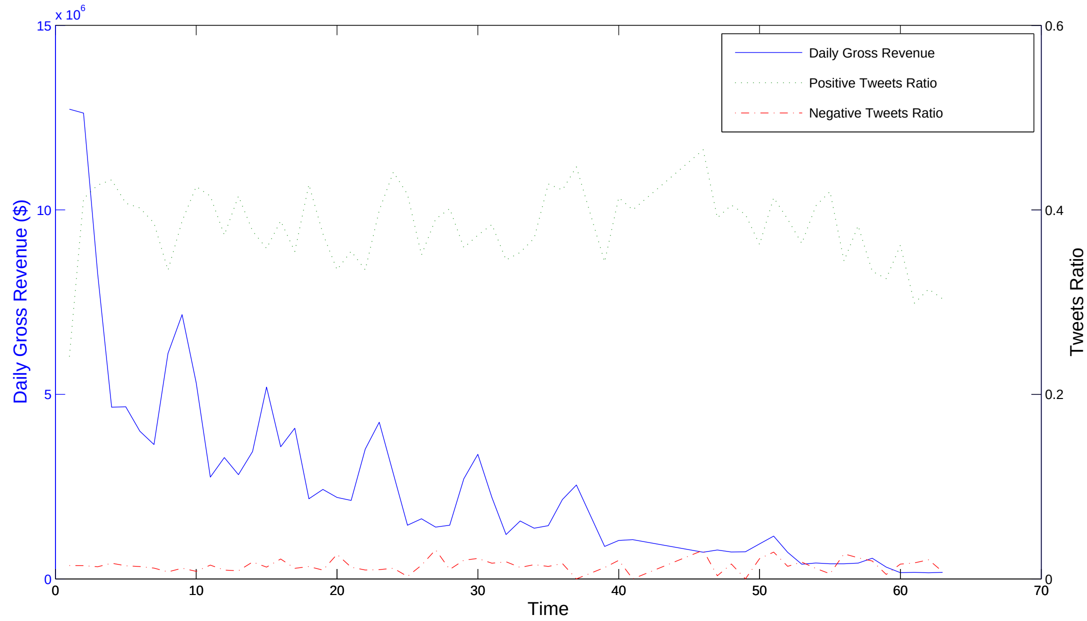

Figure 2: Dynamics of Movie Box Office Revenues and Twitter Valence for "The Proposal"

into measuring the potentially different impacts of WOM on movie sales from people with different levels of influence. Assuming that the number of followers a Twitter user has is a coarse proxy of her influence, our empirical results suggest that indeed, WOM messages from a more influential person is significantly larger than WOM messages from a less influential person. Nevertheless, more WOM messages always lead to more movie sales, regardless of whom the WOM messages are from. Our second contribution is identifying and estimating the impact of a different type of WOM on movie sales, namely, the pre-consumption WOM. The prevalence of pre-consumption WOM is most likely a result of the recent popularity of social broadcasting services, which probably explains why it is largely ignored in the earlier literature on WOM. We find that effect of pre-consumption WOM on movie sales is larger than post-consumption WOM, whether the post-consumption WOM is positive or not. We attribute this to the dual effects of pre-consumption WOM on movie sales: the direct effect on movie sales through the WOM author's own purchase behavior, and the indirect effect on movie sales through either the awareness effect or the persuasive effect of the WOM on its recipients. The third contribution of this study is to support the view that the valence of WOM does matter. Unlike most of the previous literature that uses ratings provided by users, we analyzed the sentiment of each tweet using a classical Naive Bayesian classifier and the estimation results of our econometric model suggest that positive WOM leads to more movie sales than negative WOM.

All data and algorithms used in the paper are readily available to marketing researchers and practitioners. 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, we believe it could be particularly rewarding for marketing researchers and practitioners to dig into this goldmine. The following issues, which are also the limitations in this paper, could be promising directions to pursue in the future.

First, the number of followers is obviously a very coarse measure of a Twitter user's personal influence. The practice of dividing tweets into two groups based the number of followers the author of each tweet has is a compromise between accounting users' influence while evaluating the impact of WOM on movie sales, and accurately measuring users' influence. Future research could refine the measurement of users' influence and incorporate a better influence measurement into the econometric model, which may potentially yield interesting and useful results regarding personal influence, WOM, and firm product sales.

Second, sentiment analysis is another challenge in studying the effect of online WOM on product sales in today's Web 2.0 era. 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. The algorithms we used to classify tweets are very effective but far from being perfect.[12] On the other hand, identifying a tweet as positive, neutral or negative offers only

[12]Currently, sentiment analysis is an active research field in computational linguistics and could be par-

one dimension of WOM sentiment, although this is probably the most important dimension. Still, there might be other dimensions of WOM sentiment that are interesting to explore, like the intention feature that is explored in this study. After all, human language contains far more information than valence.

ticularly useful to marketing researchers.

References

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

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

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

Dhar, V., and E. A. Chang (2009): "Does Chatter Matter? The Impact of User-Generated Content on Music Sales," *Journal of Interactive Marketing*, **23(4)**, 300-307.

Duan, Wenjing, Bin Gu, and Andrew B. Whinston (2008): "Do Online Reviews Matter? An Investigation of Panel Data," *Decision Support Systems*, **45(4)**, 1007-1016.

Gladwell, Malcolm (2000): *The Tipping Point*, New York: Little, Brown and Company.

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

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

Katz, E., and P. F. Lazarsfeld (1955): *Personal Influence*, Glencoe, IL: Free Press.

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

Sonnier, G. P., L. McAlister, and O. J. Rutz (2011): "A Dynamic Model of the Effect of Online Communications on Firm Sales," *Marketing Science*, **30(4)**, 702-716.

Slywotzky, Adrian J., Benson P. Shapiro (1993): "Leveraging to Beat the Odds: The New Marketing Mind-set," *Harvard Business Review*, **71(5)**, 97107.

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

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