Sep 14th, 2015

# Pre-release Word-of-Mouth Dynamics: The Role of Spikes

**Sarah Gelper,** Assistant Professor of Marketing, Eindhoven University of Technology s.gelper@tue.nl

**Renana Peres,** Assistant Professor of Marketing**,** School of Business Administration**,** Hebrew University of Jerusalem renana.peres@mail.huji.ac.il

**Jehoshua Eliashberg** Sebastian S. Kresge Professor of Marketing and Professor of Operations and Information Management**,** Wharton Business School**,** University of Pennsylvania eliashberg@wharton.upenn.edu

## Abstract

Prior to launch, many new products generate buzz through various social media. We study the time dynamics of pre-release word of mouth (WOM) for movies. Such WOM typically increases toward release and contains sudden spikes.

This article provides a first comprehensive treatment of WOM spikes. We introduce a dynamic model for spiky WOM and estimate it using robust Kalman filtering. Combined with extensive content analysis of more than 90,000 pre-release online WOM messages, we study the drivers and content of spikes, as well as how they relate to box office revenues.

Our results indicate that pre-release spikes are an inherent component of WOM and are not random noise. Spikes are ignited in response to firm-created communications, such as the movie trailers, yet they also emerge spontaneously. Relative to regular WOM, WOM in spikes is more opinionated and deals with more specific aspects of the movie. Notably, pre-release spikes are an early indicator of future box office revenues: Controlling for the overall number of WOM messages and other movie characteristics, we observe that movies with spiky pre-release WOM realize, on average, more ticket sales.

**Keywords**: User generated content, pre-release word of mouth, times series, movies, box office revenues, spikes, bursts.

Sep 14th, 2015

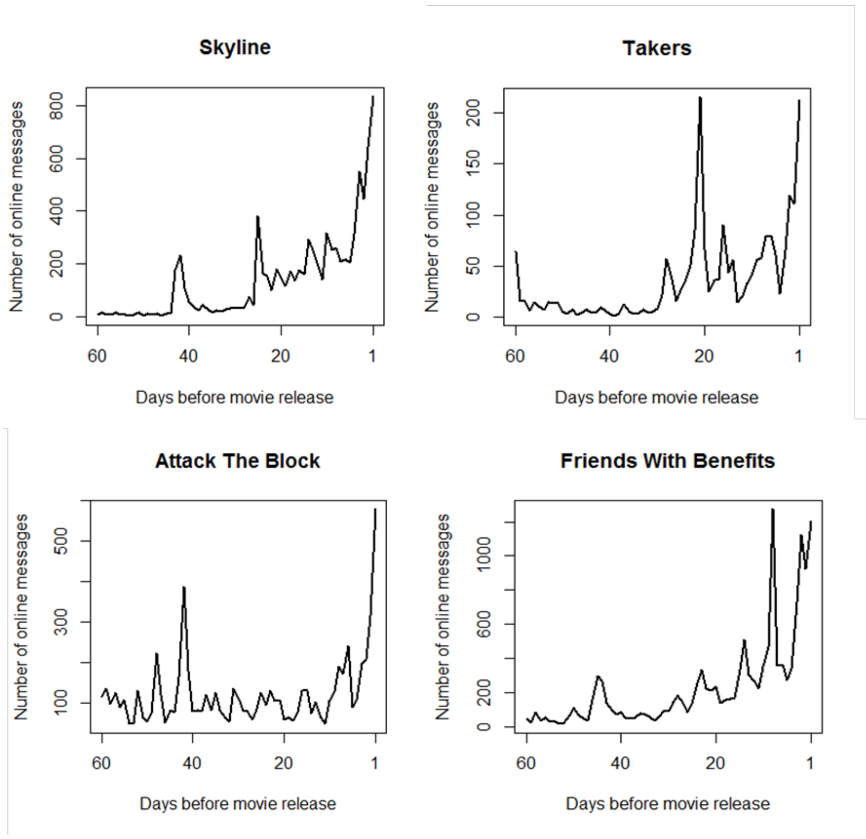
# 1 Introduction

The movie *Friends with Benefits* was released on July 22, 2011. An examination of the time dynamics of the number of online WOM messages for this movie on Twitter, blogs and user forums (Figure 1), indicates that it does not evolve smoothly, but contains sudden "spikes", or bursts of interpersonal communication among consumers. This pattern is not unique to *Friends with Benefits*. Our data, as illustrated in Figure 1, indicate that the pre-release WOM for many movies contains spikes. These WOM spikes are of managerial interest: For an advertiser, they might represent a successful outcome of an advertising campaign. For a marketer, they represent peaks of consumer activity that could be leveraged as part of the product's marketing strategy. Moreover, such spikes potentially carry predictive value: as we show later, spikes in pre-release WOM are an early indicator of future sales.

The spiky nature of social interactions has been recognized in the social network literature (e.g. Biggs, 2002; Barabasi 2005; Crane and Sornette 2008; Leskovec et al., 2009; Myers and Leskovec, 2014). These articles are part of a growing interest in emergent phenomena, a term that refers to large-scale ordered behavior that emerges from interactions between the individual elements of natural systems (Darley 1994). Despite the general interest in spiky patterns in social networks, the marketing literature has largely ignored spikes. WOM studies in marketing mostly focused on cumulative WOM (e.g., Kim and Hanssens 2013; Lovett, Peres, and Shachar 2013). Those studies that do consider the progression of WOM over time, explore the effect of WOM at time *t* on sales at time *t+1* (e.g., Godes and Mayzlin 2004; Liu 2006), disregarding the overall temporal shape of WOM, or else they focus on trends by smoothing out the spikes (O'Connor et al., 2010; Xiong and Bharadwai, 2014).

The goal of this paper is to provide a comprehensive examination of WOM spikes. Our main claim is that spikes are an important and inherent component of WOM. They are not random noise, or outliers that can be ignored, but a manifestation of focused awareness of, attention to, and interest in the product. Specifically, we answer three questions. First, what is the mechanism that *creates* WOM spikes? Using an agent-based model, we demonstrate how mechanisms suggested conceptually by social network theory can create spiky WOM. We empirically link the occurrence of spikes in WOM about a product to the firm's marketing communications. Second, what are the *characteristics* of WOM spikes? To answer this question, we study the ubiquity and content of spikes. We conduct a large-scale classification of the online WOM messages using Amazon Mechanical Turk. This allows us to empirically explore differences between WOM during a spike and regular non-spike WOM. We relate these differences to the spike creation mechanisms. The third question addresses the relationships between WOM *spikes and the product sales*: Do products with spikier WOM generate higher sales levels, controlling for the overall level of WOM?

**Figure 1:** The number of online WOM messages from Twitter, blogs, and user forums, for the movies *Skyline* (released Nov 12, 2010), *Takers* (released Aug 27, 2010), *Attack the Block* (released July 29, 2011), and *Friends with Benefits* (released July 22, 2011).




In our analysis we do not make causal claims. Instead, we seek to measure, explain and characterize a phenomenon that, despite its ubiquity and potentially meaningful implications for marketers, has not yet been addressed in the marketing literature. In this sense, this work is a "phenomenon paper", similar in spirit to the work of Zhang, Bradlow, and Small (2015) who describe the the clumpiness phenomenon, Pham, Lee, and Stephen (2012) work on the emotional oracle effect, and Ehrenberg et al. (1990) work on double jeopardy.

Our focus in this paper is on pre-release WOM on movies. Movies, as well as other entertainment products such as video games, books, and music, receive elaborate WOM prior to release. Notably, while other pre-release activities such as search (Kim and Hanssens 2013), advance purchase orders (Moe and Fader 2009) and participation in prediction markets (Foutz and Jank 2010) have been studied, pre-release WOM has received little attention thus far (Xiong and Bharadwaj 2014). Pre-release WOM provides a unique natural setting for studying spikes: focusing on the pre-release period enables detaching the WOM from the purchase itself, so the WOM it is not influenced by usage patterns. Thus, one can separate the effect of WOM on sales from the opposite effect of sales on WOM. The special characteristics of the movie industry – the frequent releases, short life-cycles, well-defined release dates, and active pre-release social interactions – make it especially suitable for the purpose of this study (see Eliashberg, Elberse, Leenders 2006 for a review). Table 1 summarizes our contribution relative to the existing literature on WOM on movies.

We have compiled and analyzed a large dataset of pre-release online messages of the top 157 movies released in Hollywood during 2010 and 2011, plus data on the movies box office revenues, advertising expenditures, and PR communication events. We find that WOM spikes occur both in response to firm-created marketing communications, such as the release of a movie trailer, but can also as emerge due to spontaneous synchronization.

We ran a large-scale content classification procedure on Amazon Mechanical Turk (MTurk) to analyze the content of more than 90,000 messages along multiple content dimensions. The analysis reveals that WOM in spikes differs from regular non-spike WOM. Relative to non-spike WOM, spike messages are more likely to mention specific movie aspects such as the actors, the director, or the trailer. WOM spike messages are also more opinionated – both more positive and negative – than regular WOM. Our results show that movies with spikier pre-release WOM, on average and controlling for movie characteristics such as the production budget and the star power, realize higher box-office tickets sales. This is especially the case for movies with many spikes that cannot be linked to firm-created marketing communications.

**Table 1:** Literature on dynamic word of mouth on movies

| Paper | Subject | WOM Data | Pre-release WOM | Content analysis | Dynamic analysis | Spikes or other irregularities |
|---|---|---|---|---|---|---|
| Duan, Gu, and Whinston (2008) | The positive feedback of WOM-sales and sales-WOM. | User reviews Yahoo!movies. Jul2003-May2004, 71 movies | No | No (user rating score) | Impact of WOM on day *t-1* on BO revenues on day *t* and *t+1* | No |
| Dellarocas, Zhang and Awad (2007) | Using online product reviews to forecast sales. | User reviews Yahoo!movies 2002 | No | Sentiment | A hazard model for the BO revenues at time *t* as a function of WOM at *t-1* | No |
| Godes and Mayzlin (2004) | How WOM affects ratings, Considering the positive feedback mechanisms, The impact of the level and dispersion of WOM mentions across user forums. | Usenet conversations on new TV shows 1999-2000 44 shows | No | Sentiment analysis on sample data | Impact of WOM for episode *t-1* on rating at *t* | No |
| Liu (2006) | How WOM on movies impacts sales, Comparing pre- and post-release WOM sentiment, Impact of movie characteristics on WOM. | User reviews Yahoo!movies May-Sep2002, 40 movies | 1 week pre-release 2 weeks post release | Sentiment | Impact of WOM at week *t-1* on BO revenues at *t* | No |
| Mahajan, Muller and Kerin (1984) | A diffusion model incorporating negative WOM. | Survey on the WOM and movie viewing of 67 subjects | No | No | A diffusion model based on positive and negative WOM | No |
| Nagle and Riedl (2013) | How disagreement in reviews drives online WOM. | User reviews Yahoo!movies 2007-2009 433 movies | No | Sentiment through rating and word analysis | Impact of disagreement at week *t-1* on propensity to review at *t* | No |
| Onishi and Manchanda (2012) | The mutual influence of blogging activity, TV advertising, and sales. | Blog mentions, 2007, on 12 movies. | Yes | Sentiment | Impact of the factors at *t-1* on outcomes at time t. | No |
| This paper | Spikes, their characteristics, and role in BO sales. | Posts from user forums, Twitter, and Blogs 2009-2010, 157 movies | Yes | Topic, tone, external events, and sentiment analysis | Robust Kalman filter to identify spikes; use spikes and trend from t0 to t days before release to predict BO revenues | Yes |

The remainder of this paper is organized as follows: Section 2 describes a theoretical framework on the creation of spikes and an agent based model illustrating that WOM spikes are the result of individual-level interactions. Section 3 describes the dataset and the process of WOM content analysis conducted via Amazon Mechanical Turk. Section 4 presents the methodology of spike identification. Section 5 presents our empirical findings on the creation and content of spikes, as well as a model linking the dynamic behavior of social systems which include spikes and box office revenues. Section 6 summarizes the main findings and discusses their implications for research and practice.

# 2 Spikes in Word of Mouth

We use social network theory to conceptualize and illustrate how interactions between individuals can create spiky aggregate WOM. We first review the relevant theories from social system and network research. Based on these theories, we then present an agent-based simulation model, demonstrating how and when individual-level interactions lead to spikes.

## 2.1 Theoretical Background: Spikes Result from Social Interactions

Spikiness is not a new concept in social system research. Granovetter (1978) showed that a wide range of behaviors such as riots, strikes, voting, migration waves and diffusion of rumors do not evolve smoothly over time, but rather come in bursts of spikes. Biggs (2003, 2005), studied strike waves and observed that “transgressive contention occurs in waves. People suddenly shift from quiescence to defiance” (Biggs 2005, page 1684). In their work on meme tracking Leskovec, Backstorm, and Kleinberg (2009), explore the spiky time dynamics of memes on news articles showing that the frequency at which a specific meme appears increases sharply and then decays sharply. Barabasi (2010) describes the spiky pattern of social and historical events, such as crusades, crimes, and personal productivity. More generally, these studies reflect growing interest in emergent phenomena – individual level interactions that give rise to large-scale collective behaviors (Darley 1994). Some of these behaviors, such as riots, panics, and bubbles, are of temporary nature (Izumi and Ueda 1998).

Beyond just the documentation of spikes, several attempts were made to explain their emergence and evolvement. While each attempt represents a different perspective, we will try and integrate them into a single framework by discussing the three main spike aspects: ignition, growth, and decay.

An intuitive explanation to the *ignition* of spikes is that a spike is a direct reaction to an external event of interest, (e.g. a new government policy which drives people to go out to protest; an earthquake drives people to communicate about it (Shi et al 2013)). However, external events are not the only possible cause for spikes: Crane and Sornette (2008) and Myers, Zhu, and Leskovec (2012) explore response to YouTube videos and Twitter messages, respectively, to show that spikes can also happen due to a cumulative unplanned spontaneous synchronization of individual responses (see also Strogatz 2004 for a conceptual discussion). Thus, our framework invokes the viewpoint that spikes can be ignited due to either an external event, or internally due to a random synchronization among individuals.

After ignition, what makes the initial activity *grow* into a spike? While a spike can happen simply if multiple people responded simultaneously to an external event, independent of each other, social network literature emphasizes the importance of social interactions to spike growth. Biggs (2003) claims that exogenous variables alone are unable to explain spikes. He argues that the growth of spikes is also contingent on a "positive feedback" mechanism, caused by the tendency of individuals to follow and reinforce popular behaviours. This tendency was documented and modeled in theories on fashion and informational cascades (Bikhchandani, Hirshleifer and Welch 1992). In marketing, advertising theories discuss the repetition effect (Pechmann and Stewart 1988, Nordhielm 2002), stating that repeated exposures to advertising messages will lead consumers to be more aware of the topic, remember it better, and transfer it further (Batra and Ray 1986, Vakrastas and Ambler 1999). While advertising is controlled by an entity very different from a social system, the repetition effect might also apply to repeated exposures to WOM messages.

The extant literature recognized that spike creation has a random component – not all initial ignitions develop to a high level of activity. Granovetter (1978) modeled this randomness by stating that each individual has his personal threshold for action in terms of the number of other people engaged in the behavior. The critical mass for collective action fluctuates randomly due to fluctuations in the individual thresholds, and as a result, some events might generate collective actions while similar event on other times might not.

Barabasi (2005) modeled this randomness from a slightly different direction. His model states that the individual pattern of performing everyday tasks is spiky (e.g., a person does not answer her emails as they arrive but rather concentrates them to certain times). When tasks require interactions among individuals (as in WOM communications), these individual spikes will be even stronger, because people depend on feedback from others (Oliveira and Vazquez 2009). When considering the aggregation of individual behaviors into network dynamics, we can expect a spike in aggregate behavior to reflect a situation in which multiple individuals place a high priority on an interactive task at the same time. As mentioned above this can happen either spontaneously, or in response to an external driver. Since the individual behavior with respect to individual tasks is stochastic, the entire process is of random nature.

Spikes are transient and eventually *decay*. New information, which shifts the convergence to a new equilibrium (Bikhchandani, Hirshleifer and Welch 1992), loss of relevance (Myers, Zhu, and Leskovec 2012), or saturation after a certain number of repetitions (Cacioppo and Petty 1979, Calder and Sternthal 1980), drive people to shift their focus. In WOM communications, this means that topics have a "life expectancy" – recent topics are discussed more, while older topics receive less attention (Leskovec, Backstorm, and Kleinberg 2009). As a result, the spike will decay.

Integrating the basic principles of the above theories, and applying them in the context of word of mouth on movies, we suggest that a WOM spike on a movie could be ignited externally, by an event such as a press conference or a trailer – labelled *event spikes* – or ignited internally, by a spontaneous synchronization of several people talking about the movie – labelled *internal spikes*. After ignition, positive feedback creates a chain reaction of responses, which causes a sharp increase in the overall WOM level. This increase is temporary and the spike decays to the normal WOM level, due to people's tendency to talk about recent and relevant topics.

In this next session we present insights from an agent-based model we developed based on the three principles discussed above: external and internal ignition, positive feedback, and recency. We will show how they combine to create spiky WOM dynamics.

## 2.2 An Agent-Based Simulation for Spike Creation

To get a better understanding of the mechanisms that create spikes, we run an agent based simulation in which we assume that individuals choose which topics to discuss from a consideration set of available topics. Consider $N$ individuals, each choosing probabilistically a topic to mention in each time period $t=1, 2, 3,\ldots, T$. The probability that a given person mentions topic $j$ in period $t$ depends on two factors: positive feedback and recency. The positive feedback is captured by a monotonically increasing function of the cumulative number of times that topic $j$ has been mentioned up to the current time period $t$, denoted by $f(n_{jt})$. Recency is captured by a monotonically decrease function of the time elapsed since topic $j$ was first mentioned, denoted by $r(t- t_j)$ where $t_j$ is the time period at which topic $j$ was first mentioned. Given $f(n_{jt})$ and $r(t-t_j)$, the probability that a given person mentions topic $j$ in period $t$ is proportional to $f(n_{jt})r(t-t_j)$. To set up the simulation, at $t=0$, there are exactly as many topics to choose from as there are individuals $N$ and we let each person mention a unique topic $j$ ($j=1, 2, \ldots N$). In the next periods, $t=1, 2, 3, \ldots T,$ exactly one new topic becomes available in each period. As a result, in each period $t$, each person chooses a topic among $N+t$ available topics, of which $N+t-1$ topics were available at period $t$–1, and one is a new topic. Note, that we follow here the conceptualization of Leskovec, Backstorm and Kleinberg (2009) regarding the inflow of topics. Other models used binomial distribution (Myers, Zhu, and Leskovec 2012). Each individual chooses an existing topic $j$ ($j=1, 2, 3, \ldots N+t-1$), with a probability proportional to $f(n_{jt})r(t-t_j)$.

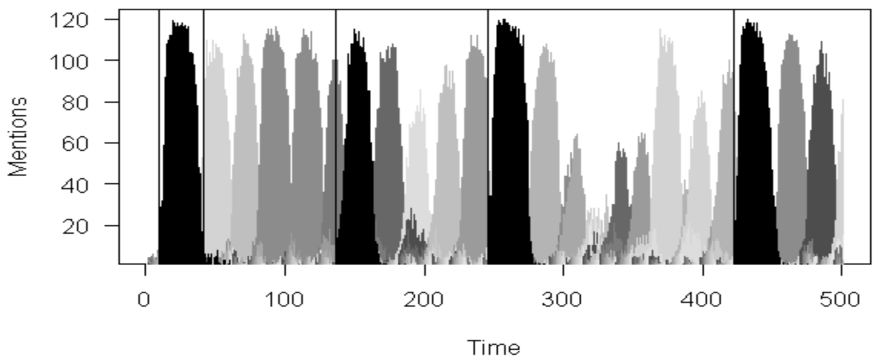
**Table 2:** For time period $t$, summary of probabilities of an individual choosing an existing topic $j$ or the new topic made available in that time period.

| If no external event happens (probability 1- $P_{event}$ ) | If an external event happens (probability $P_{event}$) |
|---|---|
| An individual chooses an existing topic $j$ at time $t$ with probability $$\frac{f(n_{tj})r(t-t_j)}{\sum_{d=1}^{N+t-1} f(n_{td})r(t-t_d) + f(0)r(0)}$$ | An individual chooses an existing topic $j$ at time $t$ with probability $$\frac{f(n_{tj})r(t-t_j)}{\sum_{d=1}^{N+t-1} f(n_{td})r(t-t_d)}\,(1-P_{new})$$ |
| An individual choose the new topic at time $t$ with probability $$\frac{f(0)r(0)}{\sum_{d=1}^{N+t-1} f(n_{td})r(t-t_d) + f(0)r(0)}$$ | An individual choose the new topic at time $t$ with probability $$P_{new}$$ |

The new topic that becomes available is assumed to have emerged either internally or as a result of the occurrence of an external event. The probability that the new topic has been triggered by an external event is denoted $P_{event}$, (1- $P_{event}$ if emerged internally). If the new topic emerges internally, each person chooses the new topic with probability proportional to $f(0)r(0)$. We set $f(0)r(0)>0$ such that new internal topics arise spontaneously. When the external event is the source of the new topic, each individual chooses that new topic with a fixed probability $P_{new}$. See Table 2 for a summary of the choice probabilities.

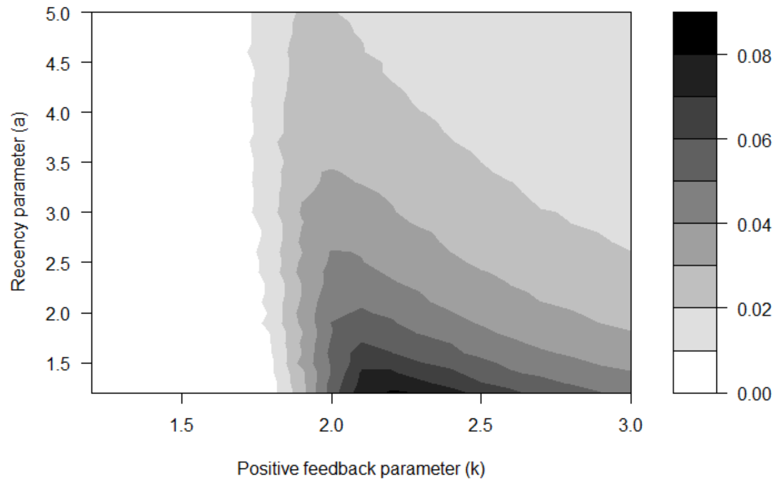
We run the simulation with $N$=120, $T$=500, $P_{event}$=0.01, $P_{new}=0.2$, $f(n)=100+n^k$ and $r(t)=1/exp(t)^{1/a}$. Larger values of $k$ indicate a stronger positive feedback, while smaller values of $a$ indicate a stronger preference for recent topics. With this value of $P_{event}$, the number of external events is small, and most spikes occur spontaneously without the external trigger. Figure 2 plots the number WOM mentions of each topic in each time period, for one simulation of the agent based model with $a$=3, $k$=2. The figure shows that external events, which occur at the times of the solid vertical lines, are not always needed to create a spiky pattern. Many spikes emerge in periods without external events. Moreover, an external event does not necessarily create a spike – while five external events have occurred, four of these have created a spike and one did not.

**Figure 2.** WOM number of mentions of all topics over time. All the shades of grey are topics that emerge internally; those triggered by an event are in black. The vertical lines indicate external events occurrences.




In Figure 3, we vary systematically the parameters *a* and *k* and measure the level of spikiness. As a measure of spikiness, we use the proportion of topics that reached the threshold of being chosen by at least half the population during at least one time period. According to this definition, if during *N* days, the entire social system was busy discussing a single topic, the spikiness is 1/*N*, which indicates a very low level of spikiness. On the extreme opposite case, if the topics change daily so that each day more than half of the population discusses a new topic, the spikiness is at its highest possible level 1. The darker the grey color, the spikier the aggregate WOM is. High spikiness happens when the social system has a strong preference to discuss recent topics (smaller values of *a*), and a moderate level of positive feedback (*k*). If the positive feedback is very low, only recent topics are discussed and no topic can gain enough attention to “catch on” and create a spike. On the other hand, if the positive feedback is very high, a small number of topics dominate, each for a long time, and the interest in each topic fails to decay rapidly i.e. does not follow a spike-like shape.

**Figure 3.** Level of spikiness as a function of the recency and positive feedback parameters, averaged over 20 runs for each parameter combinations.




The agent-based simulation demonstrates that the theoretical arguments, which are based on communication mechanisms discussed in the social networks literature, are capable of explaining the creation of spikes in social systems. The combination of a positive feedback mechanism coupled with a preference for recent topics generates spikes. These spikes can emerge either spontaneously, or in response to external events.

# 3 Data

## 3.1 Word-of-Mouth and Movie Characteristics: Data Sources

Our dataset consists of a total of 157 movies released between August 20, 2010, and August 10, 2011, for which we observe pre-release online WOM and box office revenues. We use opening-weekend box office revenues as key performance metric because these are less affected by post-release WOM than cumulative box-office revenues, allowing us to isolate the connection between pre-release WOM and ticket sales. Sales data are collected from BoxOfficeMojo. We use the Nielsen-McKinsey's Incite Buzzmetrics tool to retrieve the WOM data. This tool (which is no longer commercially available) was a proprietary text-mining engine that crawls continuously through millions of social media websites (e.g., blogs, user forums, discussion boards, Twitter) and archives their content. Brands, as well as any other topic, can be searched for in this archive with the aid of designated keyword-based queries. Because the archive is not movie-specific, in certain cases it was necessary to use complicated queries to distinguish mentions of a particular movie from other topics of interest with similar names.[1] For each movie, we applied the engine to count the daily number of messages from the time the movie was first mentioned online until the release date. In addition, we have the complete text of the messages, which we use for the content analysis.

For each movie, we collected a set of movie characteristics: production budget (from BoxOfficeMojo), advertising expenditures (from Ad$spender dataset of Kantar Media), MPAA rating (from IMDb), star power, and genre (action, animation, comedy, drama, horror, science fiction, thriller, from IMDb). We measure star power using the IMDb popularity measure which is based on the IMDb users' searches of the star. This ranking varies on a weekly basis. Because our analyses are based on data starting 60 days prior to release, we use star power averaged over 60 days prior to the release date[2]. The genre is indicated by seven dummy variables, and one movie can be assigned multiple genres (e.g., the movie *Bridesmaids* is both comedy and drama).

To analyze external events, we collected, for each movie, the list of all press events and firm-created marketing communication events related to the movie and its cast during the period of 60 days prior to release. This was done using IMDb, the movies web sites, and the LexisNexis news archive.

[1] For example, the query for the movie "Another Earth" (released Oct 12, 2011) was: (("Another Earth" and (movie or film or movies or saw or trailer or see or watch or watching or watchin or Review or Reviews or Sundance or festival or Cahill or Brit or Marling or Mapother or Erlbach or fox) and not (tremma or et or quake or "like planet" or "earth-shaking" or "Scientific American" or RP))).
[2] Traditionally, star power was measured using the *Hollywood Reporter* star-power index (Elberse and Eliashberg 2003, who collected data on movie release during 1999). However, this measurement stopped in 2000.

## 3.2 Word-of-Mouth Content

To delve deeper into the characteristics of spikes, and to analyze their content relative to regular WOM that is not part of a spike day, we carried out an extensive content analysis. Analyzing online messages is a complicated task, especially for movies. The vocabulary is large, since people talk about movies in various contexts, and many of the messages contain nonstandard words (e.g., OOOOH, Yayyy), spelling (e.g., viooz instead of views), acronyms (e.g., OMG, LOL), or punctuation (e.g., quick!?!:;;). Also, we sought to evaluate message content dimensions beyond mere mentions of the movie, such as topic, tone, sentiment, and external events. To perform such an analysis, we used the Mechanical Turk (MTurk) platform of Amazon, a crowdsourcing platform that enables labor-intensive tasks to be carried out using the services of a large number of people. MTurk has been increasingly used by academics for behavioral experiments (Winter and Suri 2012), and content analysis (Conley and Tosti-Kharas 2014). Best practices for research using MTurk have also been established (e.g. Chandler, Mueller, and Paolacci 2004).

Each online message was classified along four dimensions:

1. Topic: The main movie aspect to which the message relates. Studies on WOM among moviegoers show that people discuss aspects such as the storyline, acting and actors, director, cinematography, and soundtrack (Peacock 2000, Corrigan and White 2013). Because it is usually necessary to watch a movie in order to discuss its soundtrack and cinematography, we did not include those categories here. Instead, we added several topics that, according to our observations, were mentioned frequently in the data. Our final list of categories for the "topic" dimension includes: actor, director, the movie storyline or film-making, the genre, a trailer, professional critics' reviews, another movie, and movie listing[3]. Note that although a message might cover several topics, we asked the workers to choose the most prominent one.

[3] "Another movie" refers to messages in which the focal movie is mentioned about the main topic of the message is another movie. "Movie listing" refers to messages that simply contain lists of movies; such messages are common in social media (e.g., "Here it is to you from Mojo.com: Cowboys & Aliens. The Smurfs. Captain America: The First Avenger").

2. Tone: The emotional quality, or manner in which the topic is presented. Categories in this dimension include: call for others to watch the movie, watching intentions of the message author, opinion, a non-opinionated description of the movie, gossip, or a mere mention of the movie (e.g. "The Smurfs – next week in theatres"). Note that some categories in this dimension are embedded within others (e.g. every call for others to watch the movie is also an opinion). We instructed the workers to choose the answer which is highest in the hierarchy. For example the tweet "Soul Surfer should be awesome! Go see it!!!" was classified as a call for others to watch the movie.
3. Sentiment: The overall sentiment of the message. Categories in this dimension included the following: positive, negative, neutral – namely, no sentiment (e.g. James Franco is the main character in 127 hours), or mixed – containing both positive and negative statements (e.g. "Danny Boyle is an excellent director, but he did a bad job").
4. External event that may drive the message: On the basis of the message content, the workers had to decide whether the author had clearly identified an external event that was likely to have served as the motive for posting the message. Categories included the following: no external event, a trailer[4], other advertising, press or media event, early release, and movie premiere. Note that the workers' judgment was based solely on the message text. Of course, we do not know the real motive for writing a given message, unless the text explicitly refers to it (e.g., "I just watched the trailer, and…"). Thus, the "no external event" option means that the text did not clearly identify any external event. Note that this data comes on top of our data collection on the true occurrence of external events.

We conducted this classification on each movie for which the text of WOM messages was available to us 60 days prior to the release date (this reduced the sample from 157 to 106 movies in total). We sampled systematically for each movie one of each 10 WOM messages, from Twitter, blogs, and user forums. Altogether, we analyzed 67,740 tweets, 11,655 blog posts, and 12,840 messages from user forums. Each message was classified on the four dimensions discussed above, resulting in a massive total number of 368,940 classifications. The procedure

[4] Note that "Trailer" is a category in both the Topic and External Event dimensions. A message will be classified as Trailer in the topic dimension, if the trailer is the main topic of the message. In the External Event dimension, messages mentioning the release, or viewing the trailer as the ***motive*** for writing the post, will be classified as Trailer. For example: "Just watched the trailer of Friends with Benefits. Mila Kunis is a superstar!!!" will be classified as topic: actor; external event: trailer.

was as follows: We first grouped the messages into HITs (Amazon's jargon for a Human Intelligence Task, namely, a single job for a worker). A HIT consisted of either 20 tweets, or 10 user forum posts, or 5 blog posts. The grouping was done so that the time to complete a HIT was approximately 3 minutes. For convenience, all messages in a HIT were concentrated on one movie, and on a single content dimension. For example, a HIT could focus on classifying the Topic dimension for 20 Tweets for the movie 127 Hours.

Once accepting the HIT, the worker is presented with a message to classify and is asked to classify it into one of the categories. Then he/she clicks the “Continue” button, and continues to the next message in the HIT. Upon completing all the messages, the worker submits the HIT, and it is recorded in the MTurk result log[5]. Then, the worker can choose to accept another HIT. Prior to starting a HIT, the worker was asked to go over a set of examples, and make sure he or she had understood the task. We encouraged workers to do multiple HITs and gain expertise in the task, by offering a 20% bonus for each 10 HITs that the worker completed successfully. To ensure quality and prevent fatigue, we only used high reputation workers (a task owner, termed Requester in MTurk, can define the threshold level for reputation required for workers to participate in the task. The reputation of a worker is determined by his/her historical performance in MTurk, and feedbacks from previous requesters), and workers were limited to a maximum of 50 HITS per day.

In order to encourage workers to gain expertise, we released the HITs to MTurk sequentially, in batches, where each batch contained HITs of a single content dimension, in either Twitter, blogs, or user forums. When all HITs in a batch were completed (this took ~24 hours), we released the next batch. Altogether 1,953 workers participated in the task. The average time per HIT was 3 minutes, and the payment was 0.15$ per HIT plus the bonus.

[5] Examples of the user interface for all the content dimensions can be found in:
http://bschool.huji.ac.il/bs/MTurk/

**Figure 4**: Log histogram of the number of HITs per worker

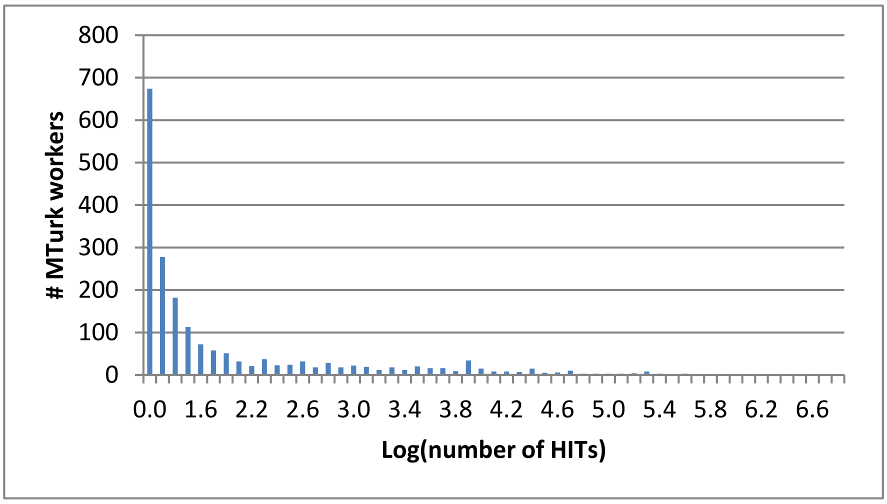



Figure 4 presents a log histogram of the number of HITs per worker for the entire content analysis. About 80% of the HITs were done by 16.7% of the workers (326 people). A third of the workers (674 people) performed just a single HIT each. As mentioned above, even workers who performed multiple HITs were limited to do no more than 50 HITs per day. Thus, we believe the combination of high-reputation workers, daily supervision, and active rejection of unsatisfactory work, together with the balance we attempted to create between enabling workers to gain expertise and preventing fatigue, optimized the quality of the workers' classifications. In addition, a team of four research assistants supervised their work, and served as a support center for questions. The supervision team went over the HITs daily, approved or rejected HITs and rejected the work of workers who were not satisfactory.

Table 3 summarizes the classification results. We see that a considerable percentage of the messages deal with the storyline or film-making of the movie (57.3%). As for the tone, while many messages were opinionated, expressed watching intentions, or called upon others to take action, many messages were simply non-opinionated mentions of the movie (31.85%). Most of the WOM was either positive (43.92%) or neutral (44.79%) (in line with Liu 2006 and other). External drivers were mentioned in about 49% of the messages, while 45.33% of the messages did not mention an external driver.

**Table 3**: Summary statistics of the MTurk classification (n=92,235 messages for each dimension)

| **TOPIC** | | **TONE** | |
|---|---|---|---|
| Actor | 12.36% | Call For Action | 7.84% |
| Director | 2.26% | Watching Intentions | 22.92% |
| Storyline or film-making | 57.35% | Opinion | 20.37% |
| Trailer | 9.41% | Gossip | 7.98% |
| Critics | 4.18% | Non-Opinionated Description | 5.58% |
| Genre | 1.43% | Mere Mention | 31.85% |
| Another Movie | 2.48% | Other or no response | 3.46% |
| Movie Listing | 3.55% | **EXTERNAL DRIVER** | |
| Other or no response | 6.98% | Trailer | 13.82% |
| **SENTIMENT** | | Press Event | 12.89% |
| Positive | 43.92% | Early Release | 3.57% |
| Negative | 6.96% | Movie Premiere | 7.94% |
| Mixed | 4.26% | Another Movie Event | 11.46% |
| Neutral | 44.79% | No External Event | 45.33% |
| No response | 0.07% | Other or no response | 4.98% |

Previous studies suggest that high reputation MTurk workers rarely fail in performing tasks (Peer, Vosgerau, and Acquisti 2013). This is also what we found from testing the workers in our task. However, as an additional quality assurance measure, we conducted ex-post tests of the workers' responses. An additional team of 10 trained research assistants served as experts and performed the tests in three stages. First, they examined the majority classifications of messages that were repeated multiple times throughout the dataset [6]. As experts the research assistants could either accept the majority classification (happened in 95.5% of the cases), or reject it, if the majority was marginal or seemed incorrect. If a research assistant decided to reject the majority classification, she corrected the answer based on her judgment and after a discussion with at least three other members of the team.

In the second stage, the research assistants conducted a semantic test: For each dimension, we created a dictionary of words and phrases corresponding for each category (for example, when people talk about trailers, they use the words *trailer*, *video*, *clip* etc.). Each

[6] Messages are frequently repeated on Twitter due to retweeting, but also in blogs; 17% of our messages were repeated more than once.

message whose classification did not seem to match its content, according to the dictionary (e.g., a message containing the word "trailer" that was not classified as a trailer in the topic dimension), was marked as suspicious, evaluated manually, and corrected if needed. In the third stage, we tracked workers who had completed large numbers of HITs and did random manual evaluations of their work. We also manually evaluated HITs completed by workers who had been flagged in previous stages as more prone to mistakes, as well as HITs with dimensions and categories that seemed more problematic. Overall, 12.3% of the messages went through manual evaluation.

**Table 4**: Performance of the content analysis. Measured for all the workers whose work was manually evaluated in at least one of the three stages of data verification.[7].

| | False Negative | True Positive | False Positive | True Negative | Recall TP/(TP+FN) | Precision TP/(TP+FP) |
|---|---|---|---|---|---|---|
| **Topic** | 6.4% | 93.6% | 0.6% | 99.4% | 93.6% | 90.8% |
| **Tone** | 34.6% | 65.4% | 5.2% | 94.8% | 65.4% | 59.7% |
| **Sentiment** | 21.2% | 78.8% | 4.4% | 95.6% | 78.8% | 60.9% |
| **External Driver** | 7.9% | 92.1% | 1.3% | 98.7% | 92.1% | 90.7% |

Our testing was biased towards the more "suspicious" workers; however, due to the pioneering nature of this classification, it is interesting to highlight some performance statistics: Table 4 presents the performance for each content dimension, pooled over blogs, Twitter and user forums, calculated separately for each category and then averaged across the different categories included in the dimension. To evaluate a worker's performance in classifying a message, we compared the answer given by the worker against the supervised answer (i.e., the answer determined by members of the research team). We carried out this evaluation for all

[7] Note that the Recall and Precision measures were calculated directly from the number of classified messages, and are not a direct mathematical processing of the columns to their left. For example, in the External Driver dimension, the trailer category, the percentage of True Positive (TP) is the number of True Positive messages (i.e., agreement between the worker's answer and the expert answer), divided by all the cases in which the supervised answer indicated a trailer external driver, regardless of the worker's response, which is exactly TP+FN. Thus, Recall rate is identical to the True Positive rate.

workers whose work had been manually checked in at least one of the above three stages. As Table 4 indicates, performance was high for all content dimensions, and was slightly lower for Tone, which is the most abstract dimension in our analysis.

As mentioned above, the numbers in Table 4 were calculated by pooling the data over messages from blogs, Twitter and user forums. Calculating separately for each of these channels, we found that the accuracy for blogs was slightly higher than for the other two.

# 4 Modeling and Identifying Spikes: A Kalman Filter Approach

To study WOM spikes, it is first necessary to identify when a WOM spike occurs. This identification is not trivial, because the general level of WOM increases as release approaches, and we need to distinguish whether an increase in WOM indicates a spike or just a random variation around the trend. We first formulate a model describing the pre-release WOM as a time-series process incorporating both the trend toward release as well as spikes. We then describe the Kalman filter that is used for spike identification in WOM time-series data.

## 4.1 Modelling Dynamic Spiky Word of Mouth

We base our model on Gelper et al. (2010). Denote by $WOM_t$ the observed volume of WOM for a movie on day $t$ (defined as the number of mentions of the movie on that day). It consists of three components: a base level ($Level_t$), sudden unexpected spikes that can occur ($Spike_t$), and a random noise component $e_t$. Hence

$$WOM_t = Level_t + Spike_t + e_t \tag{1}$$

$$e_t \sim N(0, \sigma_t). \tag{2}$$

The standard deviation $\sigma_t$ is a time-varying error volatility given by

$$\sigma_t = \sigma_{t-1} + \eta_t^V, \tag{3}$$

where $\eta_t^V$ is a zero-mean random noise component with finite variance.

We define spikes as large deviations from the base level. However, because the base level is not constant, the base level on day *t* is modeled as the base level on the previous day plus a trend ($Trend_t$) and a zero-mean random noise component ($\eta_t^L$):

$$Level_t = Level_{t-1} + Trend_{t-1} + \eta_t^L. \quad (4)$$

To allow for maximum flexibility, the trend is modeled as *local* and is not constant over time:

$$Trend_t = Trend_{t-1} + \eta_t^{Tr}. \quad (5)$$

where $\eta_t^{Tr}$ is a zero-mean random noise component. The local linear trend model with spikes defined by equations (1) to (5) is a variation of the classical local linear trend model (Hamilton 1994). In the specific case of our dataset, the trend is positive as WOM increases toward release, but the model is flexible and does not assume or require a positive trend.

## 4.2 Identifying Spikes

To identify whether a WOM spike occurs on day *t*, assume we use the estimated the base level, trend, and the volatility of WOM up to the previous day *t*-1. We can then construct the expected base level on day *t* and compare it to the observed WOM on that day. Following equation (4), the expected base level on day *t* given all information up to day *t*-1, $E(Level_t)$, is obtained as

$$E(Level_t) = Level_{t-1} + Trend_{t-1}\,. \quad (6)$$

We say the observed WOM at day *t*, ($WOM_t$) contains a spike if it greatly exceeds the expected level $E(Level_t)$. We define a threshold of five standard deviations[8], such that only unexpectedly upward jumps in WOM are identified as spikes:

$$Spike_t = \begin{cases} 0 & if\ WOM_t \leq\ E(Level_t + 5E(\sigma_t)) \\ WOM_t - E\big(Level_t + 5E(\sigma_t)\big) & if\ WOM_t > E(Level_t + 5E(\sigma_t)) \end{cases} \quad (7)$$

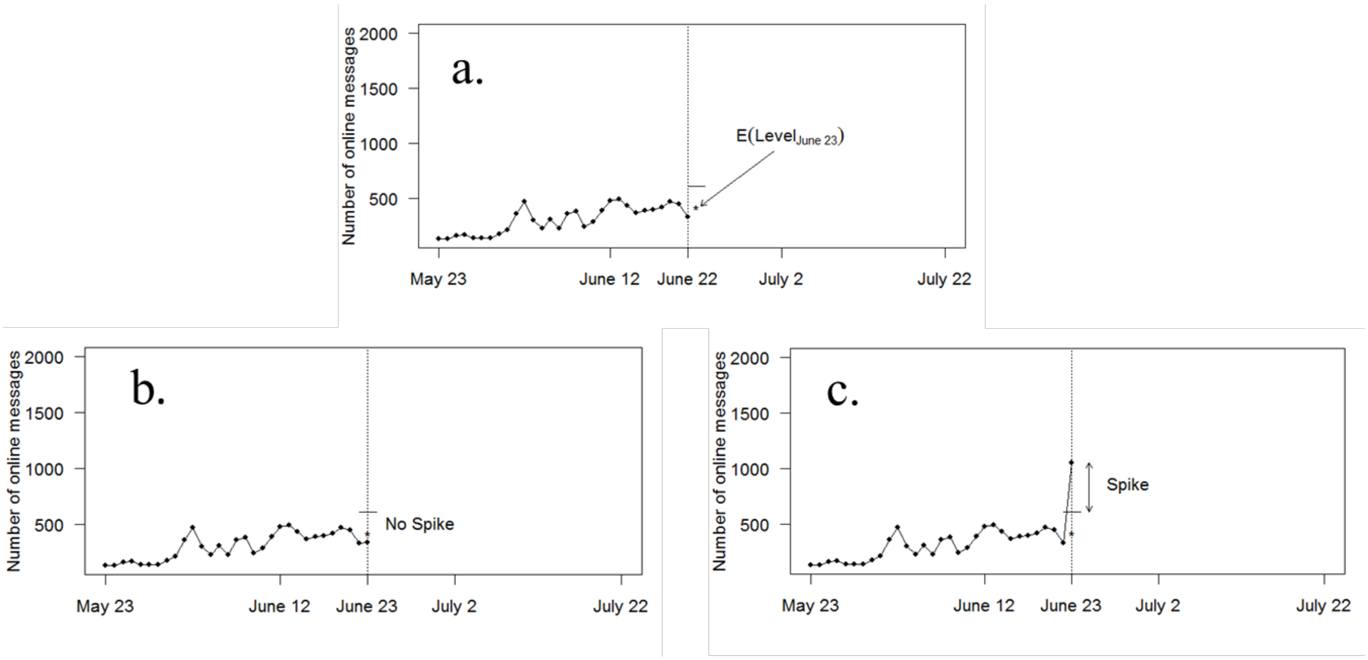
where $E(\sigma_t) = \sigma_{t-1}$.

Figure 5 illustrates the identification of spikes for the movie *Captain America* (released on July 22, 2011). Panel (a) shows the WOM up to June 22, 2011, one month before release.

[8] A robustness analysis for the number of standard deviations can be found in the Web Appendix (part 1).

The expected level on June 23 is given by the * mark. The horizontal line indicates the upper bound that we use as a threshold for a spike. We next observe the actual volume of WOM about the movie on June 23. If the observed WOM is below the threshold (see panel (b)), no spike occurs on June 23. If, however, the observed WOM on June 23 is as in panel (c), exceeding the threshold, a spike is identified. The magnitude of the spike is defined as the number of standard deviations by which the observed WOM exceeds the threshold.

**Figure 5:** Spike identification for the movie *Captain America* (released: July 22, 2011) based on observed pre-release WOM (solid line), expected WOM (*), and threshold (-).




## 4.3 **Using a Kalman Filter for Spikes Identification**

To estimate the error volatility, base level, and trend defined in equations (3), (4) and (5) from period $T_0$ to any date $t$, we use a Kalman filter approach. It is convenient to recast the model in state-space notation. The state vector $x_t$ and transition matrix $A$ are given by

$$x_t = \begin{bmatrix} Level_t \\ Trend_t \\ \sigma_t \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{8}$$

such that $x_t = Ax_{t-1} + \eta_t$ with $\eta_t = (\eta_t^L, \eta_t^{Tr}, \eta_t^V)'$.

The Kalman filter is a recursive estimation method where the estimated state components $x_t$ at day $t$ are obtained as the weighted average of the estimated state components at $t$-$1$ and the new information gained on day $t$. The weights given to current versus previous information are determined by the model parameters $\lambda_1$ for the base level, $\lambda_2$ for the trend, and $\lambda_3$ for the volatility. Thus,

$$\widehat{Level}_t = \widehat{Level}_{t-1} + \widehat{Trend}_{t-1} + \lambda_1(WOM_t^C - (\widehat{Level}_{t-1} + \widehat{Trend}_{t-1})) \tag{9}$$

where $WOM_t^C$ is the WOM baseline level on day $t$:

$$WOM_t^C = \begin{cases} WOM_t & if\ Spike_t = 0 \\ E(Level_t) + 5E(\sigma_t) & if\ Spike_t > 0. \end{cases} \tag{10}$$

Equation (10) is the one-sided Huber psi-function, used in robust statistics (Maronna, Martin, and Yohai 2006). The Kalman filter is a modification of the cleaning procedure proposed in Gelper et al. (2010). In their approach, outliers can occur both in positive and negative deviations from the mean. Since our WOM data, only upward spiked are observed, we consider only positive deviations from the base level.

After estimating the base $\widehat{Level}_t$ as in equation (9), we can also estimate the trend,

$$\widehat{Trend}_t = \widehat{Trend}_{t-1} + \lambda_2[(\widehat{Level}_t - \widehat{Level}_{t-1}) - \widehat{Trend}_{t-1})] \tag{11}$$

and error volatility

$$\hat{\sigma}_t = \hat{\sigma}_{t-1} + \lambda_3[MAD(WOM_s - \widehat{Level}_s; s = t, t-1, t-2) - \hat{\sigma}_{t-1}] \tag{12}$$

Here $MAD(WOM_s - \widehat{Level}_s; s = t, t-1, t-2)$ is a robust estimate of the volatility based on the latest information. The median absolute deviation (MAD) is defined as

$$MAD(x) = c\ median(|x - median(x)|) \tag{13}$$

For $c$ = 1.4826, MAD($x$) is a consistent estimator for the population standard deviation of $x$ (Maronna et al. 2006). Using a robust estimator of volatility is important because the observed WOM contains spikes that would inflate the estimated volatility if we used a non-robust estimator such as the standard deviation.

The advantage of the estimation approach in equations (9) to (13) as compared with a standard Kalman filter approach is twofold. First, a standard approach is not designed for a spiky time series and thus cannot be used to identify spikes. Second, because WOM spikes could be considered outliers, a standard estimation approach would be highly unstable in the presence of spikes. By using the cleaned version of the WOM process and a robust scale estimator in equation (13), the influence of one WOM observation on the estimation procedure is bounded. The procedure we use thus provides a *robust* estimation of the base level, trend, and volatility.

We can apply the robust Kalman filter in equations (9) to (13) only if we know the parameters $\lambda_1$, $\lambda_2$ and $\lambda_3$. We estimate these parameters using maximum likelihood on the cleaned WOM time-series data. On any day *t*, we estimate the parameters using data up to day *t*. Given the assumption of a normally distributed error component in equation (2), the negative log-likelihood (NLL) of the model for a given parameter set $\lambda_1$, $\lambda_2$ and $\lambda_3$, using cleaned WOM, is given by

$$NLL_t(\lambda_1,\lambda_2,\lambda_3) = \sum_{\tau=T_0+1}^{t} \ln(\hat{\sigma}_\tau) + \sum_{t=T_0+1}^{t} \frac{(WOM_\tau^C - (\widehat{Level}_{\tau-1} + \widehat{Trend}_{\tau-1}))^2}{2\hat{\sigma}_\tau^2} \qquad (14)$$

up to a constant. For given values of$\lambda_1$, $\lambda_2$ and $\lambda_3$, all quantities in equation (14) are obtained recursively based on the Kalman filter. The recursive computation requires a startup period of three observations. The starting values are set to

$$\widehat{Level}_{T_0} = WOM_{T_0}$$

$$\widehat{Trend}_{T_0} = WOM_{T_0} - WOM_{T_0-1} \qquad (15)$$

$$\hat{\sigma}_{T_0} = MAD(WOM_{T_0-2}, WOM_{T_0-1}, WOM_{T_0})$$

where $T_0$ is the index of day three after the first day the movie is mentioned online (within the 60 pre-release days considered). The exact value of $T_0$ is specific to each movie because some movies generate interest earlier than other movies. We then identify the spikes, based on the estimation of $\lambda_1$, $\lambda_2$ and $\lambda_3$ and subsequent estimation of the state space components on each day. The negative log-likelihood in equation (14) is numerically minimized using the limited-memory version of the BFGS quasi-Newton method of Byrd et al. (1995). We conducted an extensive simulation test to assess the accuracy of the Kalman Filter in identifying spikes. It is described in the Web Appendix (Part 2).

# 5 Empirical Analyses and Results

## 5.1 The Nature of Spikes: Descriptive Statistics

For each movie in the dataset, we analyze the time period beginning 60 days before its release. Table 5 presents summary statistics on the spikes during this period. On average, we counted 3.25 spikes per movie. Figure 6 gives the distribution of the number of spikes per movie (top left panel) over 60 prior to release day period. A spike lasts for 1.83 days on average, with 10.82 days on average between two consecutive spikes. The spikes vary substantially in their magnitude. The height of a spike is defined as the number of standard deviations by which the highest point of the spike exceeds the threshold of the robust Kalman filter. The average spike magnitude is 12.02 standard deviations, (median of 4.6).

**Table 5:** Spike descriptives

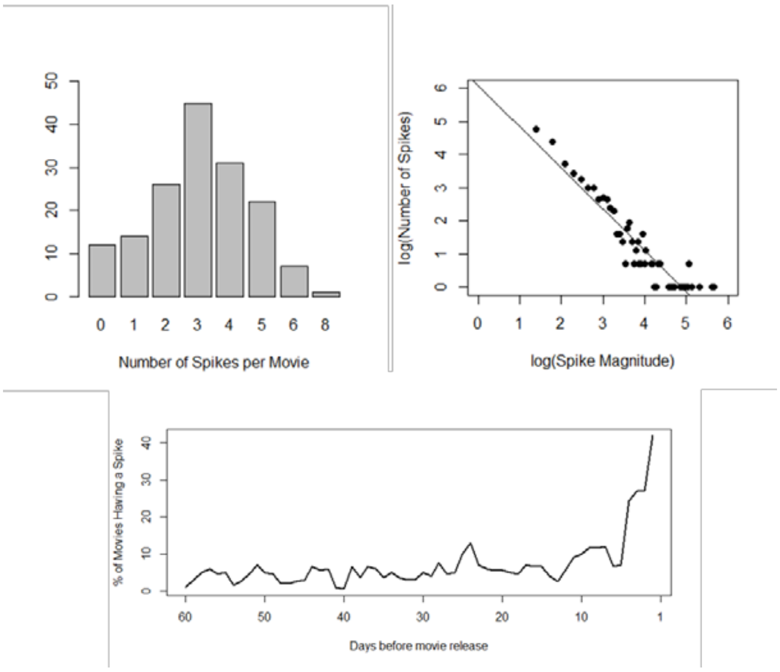
| | min | median | max | mean | s.d. |
|---|---|---|---|---|---|
| Number of spikes per movie | 0 | 3 | 8 | 3.25 | 1.56 |
| Spike duration (days) | 1 | 1 | 8 | 1.83 | 1.28 |
| Number of days between spikes | 1 | 7 | 54 | 10.82 | 10.09 |
| Spike magnitude (number of standard deviations above the 5 stdev threshold) | 0.002 | 4.639 | 304 | 12.02 | 25.37 |

The top right panel in Figure 6 shows a log-log histogram of the spike magnitude[9]. We see the distribution follows a power law, with many small spikes and a small number of large spikes. The bottom panel shows, for each of the 60 days prior to movie release, the percentage of

[9] For simplicity, we categorized the spikes in bins of two standard deviations; that is, we counted the number of spikes whose magnitude was between 0-2 standard deviations, 2–4, etc., above the 5 standard deviation threshold.

movies for which a spike took place on that day. About one week before release, we see a sharp increase in the occurrence of spikes. This will be further discussed in the next section.

**Figure 6:** Spike descriptives for all 157 movies. Top left panel: the distribution of the number of spikes per movie. Top right panel: the distribution of spike magnitude. Bottom panel: the percentage of movies having at least one spike vs. time to release.




## 5.2 The Occurrence of Spikes

The theory on spikes as the result of individual-level interactions suggests that spikes are created by an initial ignition of interest (event-driven or spontaneous), followed by positive feedback, and a decay resulting from a shift of interest. However, as we described above, a more straightforward explanation might be that a WOM spike is simply a large-scale aggregation of individual reactions to an external event of interest, where each member of the population responds independently of the others. While social network theorists (Biggs 2003, Granovetter 1978) claim that external events alone are unable to explain these spikes, this claim should be empirically validated.

In this section, we delve deeper into the creation of spikes. In particular, we investigate the relationships between external events and spikes. We focus on four types of movie-related pre-release events, referred to hereafter as “external events”: trailer releases, early movie releases (e.g. at film festivals), press events and the movie premiere. On average, in our dataset, a movie has 9 such events in the 60 days before release. These events are more likely to take place on days that are closer to movie release (a simple linear probability model with P(external event occurs $t$ days before release) = b*$t$ gives b=-0.0062, $p$ <0.001). This is consistent with Eliashberg et al. (2000). Across our entire dataset, 31% of the spikes emerge on the same day of an external event occurrence. The remaining 69% of the spikes do not co-occur with an external event[10].

Further examining the types of external events that are involved in spikes, we estimate a logit model that identifies which event types are more likely to co-occur with a spikes.  In the model, we control for the time-to-release and include movie fixed effects to control for unobserved movie-specific confounders.

**Table 6:** Results of logit regression of the probability to observe a spike as a function of types of external variables ($n$=9,420: 60 days for each of the 157 movies)

| DV: Probability of a spike occurs *t* days before release | | |
|---|---|---|
| | Coefficient | *p*-value |
| Days before release (*t*)[11] | -0.02 | <0.01 |
| Release of the Trailer | 0.83 | <0.01 |
| Early Release | 1.57 | <0.01 |
| Press Event | 0.70 | <0.01 |
| Premiere | 0.88 | <0.01 |
| Movie Fixed Effects | Included | <0.01 |

The results in Table 6 indicate that each of the four event types is significantly and positively associated with the likelihood that a spike will occur on the event day. Spikes are especially likely to emerge on the early release dates. Controlling for the movie events, spikes are more likely to occur close to release. This observation is in line with our theoretical

[10] We also tried to measure correlations with time delay between the event and the spike, but correlations are significantly smaller than the same-day correlation. Therefore we keep our focus on same day correlations.
[11] Note that $t$ ranges from 1 to 60, such that day $t$ is $t$ days before release. Therefore, for all our models hereafter, a negative coefficient means that probability of a spike increases as release approaches.

framework because as the movie's release approaches, the level of interest in the movie increases, hence the initial ignition is more likely to "catch" and evoke a discussion.

## 5.3 The Characteristics of Spikes

This section describes whether WOM differs from "regular" non-spike WOM, or whether it is just "more of the same". Because of the positive feedback mechanisms of spikes that are not present on non-spike days, we expect that the content of spikes will be different from the content of regular non-spike WOM.

We used the MTurk content data classification to run four multinomial logit models, one for each content dimension (topic, tone, sentiment and external driver). The dependent variable was the probability of the message content to correspond to a specific category within the focal dimension (e.g., that the topic is the Actor), and the explanatory variable was whether the message was part of a spike day. As controls, we used the days before release and external event dummies to note whether there was a trailer release, a press event, an early release, or a premiere on the day the message was posted. We also included movie and channel fixed effects. For each content dimension (e.g., Topic), we used a multinomial logit model for the probability that $C_{ijlt}$, defined as the content of message $i$ on movie $j$ in channel $l$ (Twitter, blogs, user forums) written $t$ days before release, equals answer option $k$ (e.g. Actor) relative to the reference option $k_0$:

$$Log\left(\frac{\Pr(C_{ijlt}=k)}{\Pr(C_{ijlt}=k_0)}\right) = \beta_{kjl} + \beta_{k1}t + \beta_{k2}Spike_{ijt} + \beta_{k3}Trailer_{ijt} + \beta_{k4}PressEvent_{ijt} + \beta_{k5}EarlyRelease_{ijt} + \beta_{k6}Premiere_{ijt} \quad (16)$$

The rationale behind this model is that individuals write their messages taking into account factors such as the information available to them, and interactions in their social system, and their personal attitudes, and all these determine the topic, tone, and sentiment of the resulting WOM message.

Table 7 displays the results of the model on the Topic dimension. The reference category ($k_0$) is "Movie Listing". Controlling for external, firm-created communication events, spike messages more likely deal with movie-specific aspects (such as the actor, director, the trailer, and critics' reviews), as opposed to merely mentioning the movie as part of a list. This result in line with our theoretical framework envisioning the spike as a burst of interest in the movie. Naturally, spike messages on a movie deal less with another movie.

**Table 7:** A multinomial logit model for the probability that a WOM message has a certain topic, reference option is "MovieListing" (n=80857)[12].

| Dependent variable: | Actor | Director | Storyline or film-making | Trailer | Critics | Genre | Another Movie |
|---|---|---|---|---|---|---|---|
| Days before release | -0.008** | -0.013** | -0.013** | 0.031** | -0.052** | -0.005 | -0.003 |
| Message is part of a spike day | 0.277** | 0.420** | 0.061 | 0.334** | 0.549** | -0.035 | -0.137 |
| Trailer launched on the message day | 0.183* | 0.037 | 0.150* | 0.222** | -0.036 | 0.374** | 0.009 |
| Press event on the message day | 0.195 | 0.221 | 0.185 | -0.587** | 0.226 | -0.717* | 0.251 |
| Early release on the message day | 0.142 | 0.593 | 0.894** | 0.925** | 0.889 | 0.792* | 0.713* |
| Premiere on the message day | -0.409* | -0.872* | -0.814** | -0.791** | 0.317 | -0.545 | -1.201** |
| Movie fixed effects | Included | Included | Included | Included | Included | Included | Included |
| Channel fixed effects | Included | Included | Included | Included | Included | Included | Included |

** p-value<0.01; * p-value<0.05.

**Table 8:** A multinomial logit model for the probability that a WOM message has a certain tone; reference category is "Mention" (n=83,918).

| Dependent variable: | Call for Action | Watching Intentions | Opinion | Gossip | Description |
|---|---|---|---|---|---|
| Days before release | -0.007** | -0.012** | -0.005** | -0.012** | 0.001** |
| Message is part of a spike day | 0.025 | -0.013 | 0.157** | 0.058 | 0.027 |
| Trailer launched on the message day | 0.051 | -0.084* | -0.089 | 0.136** | 0.016 |
| Press event on the message day | 0.019 | 0.168 | -0.025 | 0.334** | 0.111 |
| Early release on the message day | -0.238 | 0.183 | 0.249** | -0.368 | 0.059 |
| Premiere on the message day | 0.056 | 0.044 | 0.088 | 0.028 | -0.362 |
| Movie fixed effects | Included | Included | Included | Included | Included |
| Channel fixed effects | Included | Included | Included | Included | Included |

** p-value<0.01; * p-value<0.05.

Table 8 displays the analysis for the Tone dimension with "Mention" as the reference category. The table indicates that spike messages are more opinionated than non-spike messages. This finding could indicate that spikes represent large scale exchange of opinions in

[12] Note that n slightly changes between Tables 7-10. This is due to that fact that the MTurk classification was done separately for each dimension, and sometimes not all messages are classified due to workers skipping posts, and MTurk technical issues.

the social system. However, spike messages are not more likely to express watching intentions or calls for others to go watch the movie.

Table 9 presents the results of the multinomial logit estimation for the Sentiment dimension. The reference category is “Neutral”. WOM messages that are part of a spike day are more positive, negative, or mixed than non-spike WOM. This supports the analysis on the Tone dimension showing that spike messages are more opinionated compared with non-spike messages. Note that the sentiment is more positive on the days of a press event and the movie premiere, and more negative on the days of a trailer release.

**Table 9:** A multinomial logit model for the probability that a WOM message has a certain sentiment, reference option is "Neutral" (*n*=86,866).

| Dependent variable: | Positive | Negative | Mixed |
|---|---|---|---|
| Days before release | -0.003** | -0.004** | -0.003* |
| Message is part of a spike day | 0.092** | 0.123** | 0.133** |
| Trailer launched on the message day | -0.014 | 0.115* | 0.197** |
| Press event on the message day | 0.217** | 0.095 | 0.271* |
| Early release on the message day | 0.027 | 0.317 | 0.369* |
| Premiere on the message day | 0.454** | 0.149 | -0.087 |
| Movie fixed effects | Included | Included | Included |
| Channel fixed effects | Included | Included | Included |

** p -value<0.01; * p-value<0.05.

The last content dimension we discuss is the External Driver. The results on the five categories in this dimension, with reference category “No External Driver”, are presented in Table 10. Table 10 shows spike messages are more likely to mention an external event (such as a trailer, press event, early release and especially the movie premiere, which is a major, notable event) as the motive for writing the post, as opposed to mentioning no such event.

An interesting validity check is to consider the relationship of an actual occurrence of an external event (the explanatory variables) on message posters’ probability of mentioning such as a driver for engaging in WOM on the movie. Table 10 shows that a press event on a certain day is likely to evoke messages noting the press event as the motive for writing. A significant coefficient is also found for the premiere. Interestingly, the coefficient of the trailer is insignificant. This means, that although the release of a trailer increases the probability that the topic of a given message will be the trailer (Table 7), the WOM message will not necessarily mention the trailer as the motive for posting.

**Table 10:** A multinomial logit model for the probability that a WOM message mentions an external event as the driver for posting the message, reference option is "No external driver" (*n*=82,599).

| Dependent variable: | Trailer | Press Event | Early Release | Movie Premiere | Another Movie Event |
|---|---|---|---|---|---|
| Days before release | 0.035** | -0.006** | -0.007** | -0.019** | 0.011** |
| Message is part of a spike day | 0.135** | 0.156** | 0.100* | 0.289** | 0.010 |
| Trailer launched on the message day | 0.077 | 0.181** | -0.141 | -0.180** | -0.055 |
| Press event on the message day | -0.699** | 0.242** | -0.349 | -0.126 | 0.177* |
| Early release on the message day | 0.254* | -0.192 | -0.005 | -0.645* | -0.554** |
| Premiere on the message day | 0.009 | -0.824** | -0.897 | 0.583* | 0.070 |
| Movie fixed effects | Included | Included | Included | Included | Included |
| Channel fixed effects | Included | Included | Included | Included | Included |

** p -value<0.01; * p-value<0.05.

To summarize this section, we conclude that spike WOM differs in content from non-spike WOM. Spike messages are more dedicated to specific aspects of the movie, are more opinionated, are more likely to express sentiment (either positive or negative), and are more likely to refer to external events as their drivers. These results align well with our theoretical framework which describes spikes as outcomes of social interactions that might relate to external events, but are much more than a mere collection of individuals' independent responses to these events.

## 5.4 Spikes and Box Office Revenues

### 5.4.1 A Model of Box Office Revenues with Spiky WOM

If indeed pre-release WOM spikes are bursts of focused attention in the social system, one could expect WOM spikes to be a predictor of box office revenues over and above the total WOM volume and other movie characteristics. To test this possibility, we regressed the opening-weekend box office revenues on the number of spikes during the 60-day period prior to the release date, controlling for the cumulative pre-release WOM volume (measured as the number of messages on the movie) in that same period and other movie characteristics. Table 11 presents the estimation results of six models for the box office revenue. Model 1 accounts only for movie characteristics. Model 2 adds the cumulative volume of pre-release WOM (total number of mentions) over the 60-day pre-release period. As expected, advertising spending and WOM volume significantly explain variations in box office revenues. All else equal, the pre-release WOM volume elasticity is .58. Descriptive statistics and correlation table are in the Web Appendix (part 3).

**Table 11:** A model for box office revenues, DV=log(opening weekend box office revenues)[13]

| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 |
|---|---|---|---|---|---|---|
| Intercept | 6.905 ** | 3.890 ** | 2.790 * | 2.675 * | 2.833 * | 3.992 ** |
| log(Production Budget) | 0.205 | 0.022 | 0.078 | 0.082 | 0.051 | -0.054 |
| log(Ad Spending) | 0.746 ** | 0.659 ** | 0.635 ** | 0.642 ** | 0.638 ** | 0.636 ** |
| log(Star Power) | 0.079 | 0.034 | 0.052 | 0.060 | 0.060 | 0.035 |
| MPAA Ordinal | -0.410 ** | -0.369 ** | -0.325 * | -0.315 * | -0.287 * | -0.306 * |
| Genre Dummies$^{+}$ | ** | ** | ** | ** | ** | ** |
| log(WOM volume) | | 0.580 ** | 0.525 ** | 0.548 ** | 0.584 ** | 0.512 ** |
| log(# Spikes) | | | 0.641 ** | | | |
| log(# Internal Spikes) | | | | 0.612 ** | 0.648 ** | 0.634 ** |
| log(# Event Spikes) | | | | -0.123 | -0.104 | -0.232 |
| log(Mean Spike Duration) | | | | | -0.129 | -0.047 |
| log(Mean Spike Magnitude) | | | | | -0.150 | -0.147 |
| # Trailer Releases | | | | | | 0.112 |
| # Press Events | | | | | | -0.039 |
| # Early Releases | | | | | | 0.032 |
| # Premieres | | | | | | 0.076 |
| F-statistic | 31.08 ** | 41.36 ** | 40.56 ** | 37.96 ** | 33.47 ** | 27.63 ** |
| $R^2$ | 0.704 | 0.778 | 0.789 | 0.792 | 0.795 | 0.805 |
| Adjusted $R^2$ | 0.681 | 0.759 | 0.770 | 0.771 | 0.772 | 0.776 |

* p<.05; ** p<.01; *n*=157. $^{+}$ The genre dummies are jointly significant (F-test, p<.01 for all models).

Our focus is on WOM spikes, which are explored in Models 3-6. Model 3 adds the total number of spikes over 60 days prior to the movie release date. A movie that has spiky WOM dynamics realizes – on average and all else equal – more box office revenues than a movie with smooth WOM dynamics. All else equal, the WOM spike elasticity is .66.

[13] Note, that the improvement in the $R^2$ achieved between Model 2 to models 3-6 (which contain spikes), is small. However, the internal spike coefficient is similar in magnitude to that of the WOM volume.

This result reminds of the findings from the advertising literature, that in certain market scenarios (for example, S-shaped response function) "pulsing" the advertising budget (i.e. dividing it into several intense doses) dominates alternative strategies of spreading advertising dollars over time (Dube, Hitsch, and Manchanda 2005, Feinberg 1992, Fremier and Horsky 2012, Villas-Boas 1993). Similar to advertising, the reason can be that strong pulses increase recall and persuasiveness and as a result lead to higher message effectiveness, (Janiszewski, Noel, and Sawyer 2003).

We showed in section 5.2 that spikes can emerge spontaneously or in response to external events. To differentiate the two types of spikes, we identify *internal spikes* and *event spikes*. The event spikes happen on days on which there is a firm-created movie event – a trailer release, a press event, an early movie release or the movie premiere. All the other spikes are labeled *internal spikes*. Model 4 includes the numbers of internal and event spikes as separate explanatory variables. The coefficient of the internal spikes is significantly positive, while the coefficient of the event spikes is not. This does not mean, however, that firm-initiated events are unimportant to spikes. As we showed earlier, these events are mentioned and discussed in spikes. However, only the internal spikes explain variation in sales. Note, that the results of the model are robust to the choice of threshold (Equation 7). See Web Appendix (Part 1) for details.

In Model 5 we test whether additional spike characteristics beyond the number of spikes explain variations in box office revenues. For each movie we include the average spike duration, measured in days, and the average spike magnitude. These spike characteristics do not add explanatory power to the model. Model 6 also accounts for the occurrence of external events. It includes the number of trailer releases, number of press events, number of early releases and number of premieres – none of which has a significant coefficient. To ensure that the results indeed stem from spikes, we checked alternative specifications of the model with variables dealing with the distribution of WOM over time, see Web Appendix (part 4).

### 5.4.2 The Predictive Power of Spikes

To assess the predictive power of pre-release WOM volume and spikes, we perform an out-of-sample analysis. To learn how well pre-release WOM predicts early on, we make predictions starting one month (30 days) before the movie release. We use a variation of the box office models presented above that allows for early predictions. In particular, to ensure our prediction is realistic, we consider the calendar date and, for each prediction, we use historical data only. The implementation steps of the prediction are as follows:

**Step 1.** Order all movies according to release date.

**Step 2.** Split the set of movies in half: the training set contains movies 1 to 79 and the test set contains movies 80 to 157.

**Step 3.** For each movie *i* in the test set and for each time *t*, where *t* runs from 30 days to 1 day before the movie release, do the following:

- **Step 3.1.** Fit box office models 1, 2, 3 and 4 from Table 11. Instead of using the entire data up to the release date, use only the data that were historically available *t* days before the release date of movie *i*. For example, if *i*=92, with a release date of July 5, 2010, and the prediction is conducted five days before release (i.e., July 1, 2010), use the data from all the movies 1 to 91, whose release date is prior to July 1, 2010.
- **Step 3.2.** Using the estimates in Step 3.1, predict the box office revenues of movie *i* and compute the mean absolute prediction error (MAPE).

**Figure 7:** Mean absolute prediction error, for box office revenues models, averaged on all movies in the dataset

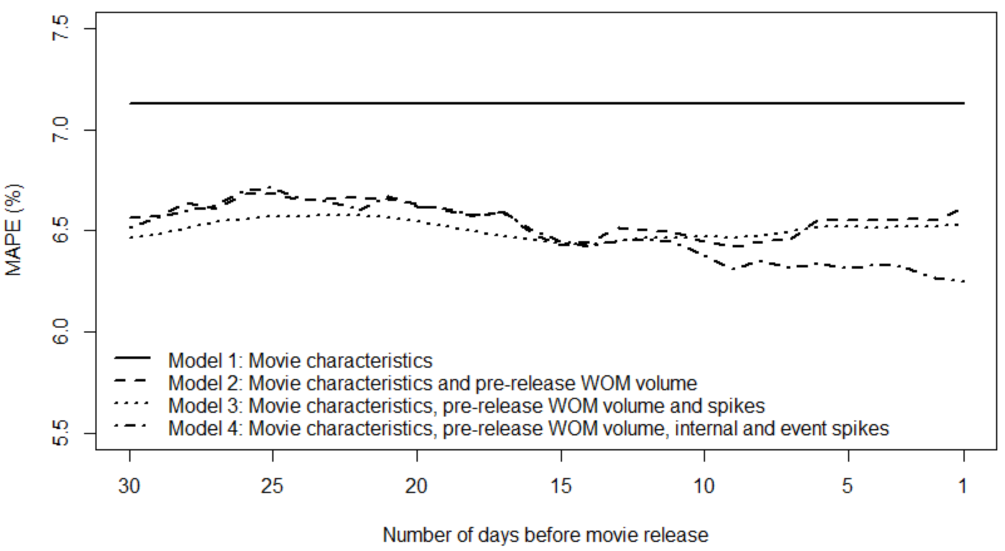


Figure 7 shows the MAPE for the Models 1 to 4 in Table 11. We see that the models that incorporate WOM volume consistently outperform the model with movie characteristics only. Starting two weeks before release, the model including internal and event spikes outperforms the others.

# 6 Discussion

This paper deals with an under-explored phenomenon of WOM: spikes. Our theoretical framework and empirical observations suggest that WOM spikes do not represent mere noise or measurement errors: They reflect focused awareness of, attention to, and interest in the product. Moreover, spikes are much more than large-scale aggregations of independent responses to external events – they are a manifestation of social interactions and consumer focus. They are an inherent and important part of WOM and are an early indicator of sales.

We addressed three questions, and here are our findings:

*What is the mechanism that creates WOM spikes*? Our conceptual framework suggests that the spiky pattern of WOM occurs as the result of an ignition, which can either take place in response to an external event, or occur spontaneously (internally). After ignition, mechanisms of positive feedback cause a sharp increase in word-of-mouth activity, which eventually decays due to recency.  According to the agent-based model, WOM will be spiky in social systems characterized by a medium level of positive feedback but with a strong tendency to discuss recent topics.  Empirically, we find that:

1. In our dataset of WOM on 157 Hollywood movies, 31% of the spikes co-occurred with days of a firm-created communication events. The remaining 69% do not.
2. As release approaches, spikes are more likely to occur. This observation aligns with our theoretical framework since with time, the level of interest in the movie increases. As a result the initial ignition is more likely to "catch" and evoke a discussion.
3. Spikes are likely to co-occur with firm-initiated events such as trailers, press events, the movie premiere and especially the early release.

*What are the characteristics of spikes? How does spike WOM differ from regular WOM?* We ran a large-scale content-analysis via Amazon Mechanical Turk to compare spike WOM with non-spike WOM. Our findings clearly show that spikes are not simply higher levels of the same WOM. In particular, we find that compared to regular WOM:

1. WOM in spikes is more specific: Spike messages are more likely to deal with movie specific aspects such as the actor, director, the trailer, and critics reviews.
2. WOM in spikes is more opinionated: Spike messages express stronger opinions and have a stronger sentiment (either positive or negative).
3. WOM in spikes relates more to external drivers: External events not only co-occur with some spikes, but the users explicitly mention them as the motive for posting the message. About two out of three spikes do *not* coincide with an external event, however. Thus, although spikes are not a direct collective response to an external event, these events have strong presence in the content of the spike messages.

*Do products with spikier pre-release WOM realize higher initial sales?* We find that movies with more spikes in their pre-release WOM tend to have higher box office opening revenues, controlling for total WOM volume and other movie characteristics. Moreover, we find that this is especiall the case for *internal* spikes – that is, spikes that do not co-occur with an external event. A box-office revenue model incorporating pre-release WOM volume, plus external and internal spikes more accurately predicts box office revenues.

Our work contributes to the social network literature as well as to marketing practice. From a research perspective, this paper contributes to the discussion of the dynamic aspects of WOM. Given the fact that WOM is a highly dynamic phenomenon, the literature on its dynamic pattern is surprisingly scarce. Here we explicitly model the *temporal pattern* of WOM and focus on deviation from the trend – the spikes. Studying WOM spikes contributes to the emerging discussion on the role and implications of irregularities, outliers, and out-of-trend observations (Barabasi 2005; Taleb 2007; Goldenberg, Lowengart, and Shapira 2009; Stephen and Galak

2012). Spikes, as well as other irregularities, carry valuable information on the dynamic social interactions in complex systems, and understanding them contributes to understanding these dynamics.

We believe our work is also of high value to practitioners. In the movie industry, the distribution and screening plans are all set 60 days before the release. At this point in time, the main managerial tool for maximizing the movie box office sales is the advertising and PR plan. Our results provide tools for forecasting movie ticket sales on the basis of pre-release WOM, as well as managing the PR activity. For example, studios can plan their PR with the aim of generating spikes instead of continuously feeding the WOM channels. As for trailers, studios currently release one to two trailers for each new movie; however, since our results suggest that trailers are more often discussed in spikes, studios might want to seriously consider the strategy of releasing a larger number of trailers. The content of spike messages can also affect the mix of PR topics: Our results indicate that messages that are part of a spike are more likely than non-spike messages to mention the actor, director, trailer and critics' reviews. Thus, studios might want to make more information available on these topic categories.

Our work opens several paths for further research. First, one might wonder whether spikes also exist in post-release WOM and what the differences are, if any, between pre- and post-release spikes. Second, it would be interesting to explore whether spikes also serve as indicators for market metrics other than sales, such as retention or brand equity. Third, we focused on the movie industry; however, we expect WOM spikes to occur for other products with a known release date as well. An investigation of the similarities and differences of WOM spikes across products could potentially yield useful insights.

# 7 References


Asur, S., B. A. Huberman. 2010. Predicting the future with social media. *Proc. Internat. Conf. on Web Intelligence*, ACM , Toronto, 492-499.

Barabási, A. 2005. The origin of bursts and heavy tails in human dynamics. *Nature*, **435** (7039) 207-211.

Barabási, A. 2010. *Bursts: the hidden patterns behind everything we do, from your e-mail to bloody crusades*. Penguin, New York.

Batra, R., M. L. Ray. 1986. Affective responses mediating acceptance of advertising. *J. Consumer Res.* **13** 234-249.

Biggs, M. 2002. Strikes as sequences of interaction: The American strike wave of 1886. *Social Science History* **26**(3) 583-617.

Biggs, M. 2003. Positive feedback in collective mobilization: the American strike wave of 1886. *Theory and society* **32**(2) 217-254.

Biggs, M. 2005. Strikes as forest fires: Chicago and Paris in the late nineteenth century. *American journal of sociology* **110**(6) 1684-1714.

Bikhchandani, S., D. Hirshleifer, and I. Welch. 1992. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political Economy* **100**(5) 992-1026.

Byrd, R. H., P. Lu, J. Nocedal, and C. Zhu. 1995. A limited memory algorithm for bound constrained optimization. *SIAM J. Scientific Computing* **16**(5) 1190-1208.

Cacioppo, J. T., R. E. Petty. 1979. Effects of message repetition and position on cognitive response, recall, and persuasion. *J. Personality and Social Psychology* **37**(1) 97-109.

Calder, B., B. Sternthal. 1980. Television commercial wearout: An information processing view. *J. Marketing Res.* **17** 173–186.

Chandler, J., P. Mueller, G. Paolacci. 2014. Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior research methods* **46** (1) 112-130.

Conley, C., J. Tosti-Kharas. 2014. Crowdsourcing content analysis for managerial research. *Management Decision* **52**(4) 675 – 688.

Corrigan, T., P. White. 2012. *The Film Experience: An Introduction*, 3rd edition, Bedford/St. Martin's, Boston, MA.

Crane, R., D. Sornette. 2008. Robust dynamic classes revealed by measuring the response function of a social system. *PNAS* **105**(41) 15649-15653.

Darley, V. 1994. Emergent phenomena and complexity. In *Artificial Life IV* 411-416.

Dellarocas, C., X. Zhang, N. F. Awad. 2007. Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *J. Interactive Marketing* **21**(4) 23-45.

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

Dube, J. P., G. J. Hitsch, P. Manchanda. 2005. An empirical model of advertising dynamics. *Quantitative Marketing and Economics* **3**(2) 107-144.

Ehrenberg, A. S., Goodhardt, G. J., and Barwise, T. P. 1990. Double jeopardy revisited. *The Journal of Marketing*, **54**(3) 82-91.

Elberse, A., J. Eliashberg. 2003. Demand and supply dynamics for sequentially released products in international markets: The case of motion pictures. *Marketing Sci.* **22**(3) 329-354.

Eliashberg, J., A. Elberse, M. Leenders. 2006. The motion picture industry: Critical issues in practice, current research, and new research directions. *Marketing Sci.* **25**(6) 638-661.

Eliashberg, J., J. J. Jonker, M. S. Sawhney, B Wierenga. 2000. MOVIEMOD: An implementable decision-support system for prerelease market evaluation of motion pictures. *Marketing Sci.* **19**(3) 226-243.

Feinberg, F.M. 1992. Pulsing policies for aggregate advertising models. *Marketing Sci.* **11**(3) 221-234.

Foutz, N.Z., W. Jank. 2010. Prerelease demand forecasting for motion pictures using functional shape analysis of virtual stock markets. *Marketing Sci.* **29**(3) 568-579.

Fremier, M., D. Horsky. 2012. Periodic advertising pulsing in a competitive market. *Marketing Sci.* **31**(4) 637-648.

Gelper, S., R. Fried, C. Croux. 2010. Robust forecasting with exponential and Holt-Winters smoothing. *Journal of Forecasting* **29**(3) 285-300.

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

Goldenberg, J., O. Lowengart, D. Shapira. 2009. Zooming in: Self-emergence of movements in new product growth. *Marketing Sci.* **28**(2) 274-292.

Granovetter, M. 1978. Threshold models of collective behavior. *American journal of sociology,* **83**(6) 1420-1443.

Hamilton, J. D. 1994. *Time series analysis,* Vol. 2. Princeton University Press, Princeton, NJ.

Izumi, K., U. Kazuhiro. 1998. Emergent phenomena in a foreign exchange market: analysis based on an artificial market approach. *Artificial Life VI* 398-402..

Janiszewski, C., H. Noel, A. G. Sawyer. 2003. A meta-analysis of the spacing effect in verbal learning: Implications for research on advertising repetition and consumer memory. *J. Consumer Res.* **30**(1) 138-149.

Kim, H., M. Hanssens. 2013. Advertising, searching, blogging and new-product sales Working Paper.

Leskovec, J., L. Backstrom, J. Kleinberg. 2009. Meme-tracking and the dynamics of the news cycle. *Proc.15th Internat. Conf. on Knowledge discovery and data mining* (SIGKDD), ACM, New York, 497-506.

Liu, Y. 2006. Word of mouth for movies: Its dynamics and impact on box office revenue. *J. Marketing* **70**(3) 74-89.

Lovett, M., R. Peres, R. Shachar. 2013. On brands and word of mouth. *J Marketing Res.* **50**(4) 427-444.

Mahajan, V., E. Muller, R. A. Kerin. 1984. Introduction Strategy for New Products with Positive and Negative Word-of-Mouth. *Management Sci.* **30**(12) 1389-1404.

Maronna, R. A., D. R. Martin, V. J. Yohai. 2006. *Robust Statistics: Theory and Methods,* 1st edition. John Wiley & Sons, Hoboken, NJ.

Moe, W. W., P. S. Fader. 2009. The role of price tiers in advance purchasing of event tickets. *J. Service Res.* **12** 73-86.

Myers, S., J. Leskovec. 2014. The bursty dynamics of the twitter information network, *proc. 23rd Internat. Conf. World Wide Web*, ACM, New-York, 913-925.

Myers, S., C. Zhu, J. Leskovec. 2012. Information diffusion and external influence in networks, *Proc. 18th Internat. Conf. on Knowledge Discovery and Data Mining* (SIGKDD ), ACM, 33-41.

Nordhielm, C.L. 2002. The influence of level of processing on advertising repetition effects. *J Consumer Res.* **29**(3) 371-382.

O'Connor, B., R. Balasubramanyan, B. R. Routledge, N. A. Smith. 2010. From tweets to polls: Linking text sentiment to public opinion time series. *Proc. of the Internat. Conf. on Weblogs and Social Media (ICWSM)*, 11, 122-129.

Oliveira, J. G., A. Vazquez. 2009. Impact of interactions on human dynamics. *Physica A* **388**(2) 187-192.

Onishi, H., P. Manchanda. 2012. Marketing activity, blogging, and sales. *Internat. J. Res. Marketing* **29**(2) 221–234.

Peacock, R. B. 2000. *The Art of Movie Making: Script to Screen*. Prentice Hall, Upper Saddle River, NJ.

Pechmann, C., D. W. Stewart. 1988. Advertising repetition: A critical review of wearin and wearout. *Current Issues and Research in Advertising* **11**(1-2) 285-329.

Peer, E., J. Vosgerau, A. Acquisti. 2014.Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior research methods* **46**(4) 1023-1031.

Pham, M. T., Lee, L., and Stephen, A. T. 2012. Feeling the future: The emotional oracle effect. *Journal of Consumer Research*, **39**(3), 461-477.

Shi, N., C. Gao, Z.Zhang, L. Zhong, and J. Huang. 2013. The Spontaneous Behavior in Extreme Events: A Clustering-Based Quantitative Analysis. In *Advanced Data Mining and Applications*, 336-347. Springer Berlin Heidelberg.

Stephen, A.T, J. Galak. 2012. The effects of traditional and social earned media on sales: A study of a microlending marketplace. *J. Marketing Res.* **49**(5) 624-639.

Strogatz S. H. 2004. *Sync: How Order Emerges from Chaos in the Universe, Nature, and Daily Life*. Hyperion, New York, NY.

Taleb, N. N. 2007. Black swans and the domains of statistics. *The American Statistician* **61**(3) 198-200.

Vakrastas, D., T. Ambler. 1999. How advertising works: What do we really know?*J. Marketing* **63** 26-43.

Villas-Boas, J.M. 1993. Predicting advertising pulsing policies in an oligopoly: A model and empirical test. *Marketing Sci.* **12**(1) 88-102.

Winter, M., S. Suri. 2012. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods* **44**(1) 1-23.

Xiong, G., S. Bharadwaj. 2014. Pre-release buzz evolution patterns and new product performance. *Marketing Sci.* 33(3) 401-421.

Zhang Y, E.T. Bradlow, D.S. Small. 2015. Predicting customer value using clumpiness: From RFM to RFMC. *Marketing Sci.* **34**(2) 195–208.