Australasian Marketing Journal 23 (2015) 179–187

Contents lists available at ScienceDirect

Australasian Marketing Journal

journal homepage: www.elsevier.com/locate/amj

Impact of user-generated and professional critics reviews on Bollywood movie success

Rakesh Niraj *, Jagdip Singh

Design and Innovation Department, Weatherhead School of Management, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, Ohio 44106-7235, USA

ARTICLE INFO

Article history:
Received 5 May 2014
Revised 6 February 2015
Accepted 26 February 2015
Available online 2 April 2015

Keywords:
User Generated Reviews
Professional reviews
Sentiment analysis
Movie marketing
Positivity ratio
Bollywood movies

ABSTRACT

User generated (e.g., peer-to-peer or word-of-mouth) reviews and professional experts' reviews are important sources of external information that consumers use in making consumption decisions, but studies of their simultaneous effects is scarce. We develop a conceptual model for hypothesizing the simultaneous effects of the volume and valence of user and expert reviews on sales (i.e., box-office) revenue in the consumption of entertainment products (i.e., movies). We gather and analyze the User Generated Reviews (UGRs) using a text processing algorithm and evaluate resulting semantic networks to validate our approach. Combining these data with other relevant data from multiple sources for all Bollywood movies released over a six month period, we test the hypotheses proposed in the model. We find that, after controlling for effect of expert reviews, both volume and valence of user generated reviews have significant positive association with box-office revenue; however, valence evidences a curvilinear (saturation) effect such that positivity effect shows decreasing returns at high positivity levels. We also find that expert reviews positively affect box office revenue with little evidence of saturation or amplifying effects.

CHINESE ABSTRACT

用户创作的（如同行彼此之间或口口相传）的评论和专家的评语是外部信息的重要来源，消费者可以在作出消费决定时对这些信息来源加以利用，但是鲜有对它们之间的协同效应的研究。我们开发了一个概念模型，以假设用户评价和专家评论对娱乐产品（如电影）的销售（即票房）收入的同步效应。我们使用文本处理算法收集和分析了用户创作的评论（UGRs），并对由此产生的语义网络进行评估，从而验证我们的方法。将这些数据与从多种来源取得的、在6个月内发布的所有与宝莱坞电影有关的其他数据结合起来，我们对模型中提出的假设进行了测试。我们发现，在控制了专家评语的效果之后，用户创作的评论无论是在数量上还是在价值上都与票房收入有显著的正相关性；然而，价值方面的证据显示了一个曲线（饱和度）的效果，即这样的积极效果在正相关性水平较高时会呈现一个下降的回报率我们还发现，专家评语会积极地影响票房收入，几乎不存在饱和度和放大效应的证据。评出了前100个营销系。

1. Introduction

Consumption of artistic experiences – such as theater, movies, concerts, stand-up comedies, and other similar performances – are fruitful, rich and unique contexts for the study of marketing phenomena. Many aspects of these contexts make them especially distinctive. First, artistic experiences are exemplars of experiential and hedonic consumption. For studies that seek to understand and enhance customers' in-consumption experiences, research on artistic performances may offer new insights. Second, to be successful, artistic performances have to strike a balance between innovation and bottom line returns. Without innovation, artistic performances fail to rise above the competition; yet, innovation needs to be tempered with mass appeal to generate sufficient payoffs that support innovative activity. Novelty versus return is a common dilemma of the service industry, and artistic experiences provide a rich context to study them. Third, marketing is an increasingly important aspect of artistic experiences. Consider movies for instance. By some accounts, studios spend 50 cents to 70 cents of their budget dollar on marketing their releases in the United States and Canada

* Corresponding author. Tel.: +1 216 368 0799; fax: +2163684785.
E-mail address: rkn10@case.edu (R. Niraj).

http://dx.doi.org/10.1016/j.ausmj.2015.02.001

alone.[1] Little by way of valid scientific evidence currently exists to quantify the ROI of this significant marketing spend, thereby providing opportunities for research to inform practice. Fourth, a majority of artistic performances are prone to fail, highlighting the importance of studying customer response to innovation. Continuing with the movie example, often the "theater life" of most movies is no more than 2 weeks, breakeven can take "years – sometimes full decade," and, were it not for post-theater payoffs, most movies would be failed investments (Davidson, 2012). In a sense, the study of artistic experiences is well suited to study the science of hit innovations (Sharda and Delen, 2006). Finally, artistic performances are subjective and intangible experiences that are not easily evaluated for quality using relatively objective criteria (e.g., as in Consumer Reports). As such, consumers' consumption decisions are significantly swayed by social influence including peer input (word-of-mouth) and judgments of respected experts (professional evaluations). Artistic experiences are rich contexts for examining the role of social influence in consumer decision making.

The aim of this study is to contribute to the study of social influence in consumption of artistic performances, specifically the so called "Bollywood" movies.[2] Four aspects of our study, including its context, are notable in their contribution. First, we seek to isolate the quality and quantity effects of peer influence or user-generated reviews (UGRs) on revenue and to examine its valence effects. Past research has provided some insights into the process of word-of-mouth (WoM) generation by consumers' peers (Dellarocas and Narayan, 2006; Godes and Mayzlin, 2004; Godes and Silva, 2012; Liu, 2006) and how it influences commercial successes of experiential products (Chevalier and Mayzlin, 2006; Liu, 2006). Liu (2006) is one of the first to incorporate volume and valence of WoM directly while investigating movie success. He finds that volume of WoM matters, while its valence does not. Dellarocas et al. (2007) find that including online review information significantly improves forecasting of revenue based on opening weekend data. Chintagunta et al. (2010) also use UGRs but relying on sequence of release for different regional US markets, they find that valence of UGRs matters while volume does not. More recent attempts to link UGRs to music (Dhar and Chang, 2009) and book sales (Chevalier and Mayzlin, 2006) have often found inconsistent support for the influence of quality and/or quantity of UGRs on commercial success.

Second, our study aims to provide a more complete understanding of social influence by going beyond peer influence to examine the role of professional-expert evaluations in consumer decision making. Although past research has recognized the importance of professional-experts (Chakravarty et al., 2010; Deasi and Basuroy, 2005; Eliashberg and Shugan, 1997), often studies fall short in accounting for the multiple sources of social influence. For instance, in an early work, Basuroy et al. (2003) examine the relative role of critics, star power and budgets on movie success. They do not have direct WoM or UGR data available and they do not consider it. Likewise, Moul (2007) also lacks these data but infers information transmission among consumers using a statistical (error component) approach and finds the variance attributable to word of mouth information transmission to be statistically significant. Chakravarty et al. (2010) finds experimental support to the idea that frequent moviegoers tend to be more influenced by critics, while infrequent moviegoers show a tendency to be swayed more by WoM.

Third, our study is also more inclusive in considering multiple sources of social influence data. Much past empirical work examines social influence data available on Yahoo! Movies. One specific feature of this is that users provide a star-ratings (on a one to five star scale) in addition to, and more often in lieu of, written comments. Absence of written comments does not impede the computation of valence scores for social influence (ratings are sufficient); however it arguably misses the more nuanced sentiments expressed in user reviews. Machine-learning assisted sentiment-extraction strategy is now being used in other contexts in marketing literature to analyze unstructured text such as written comments (Shesahdri and Tellis, 2012). No published study on movie marketing context has utilized this method to our knowledge. Also, no previous study uses Twitter of YouTube comments as data source for social influence in movie marketing. Twitter is an immediate, non-computer-dependent, source and increasingly used by consumers to share information and transmit opinions. Likewise, YouTube is a frequently used source for watching official and unofficial movie clips and promos where users often post comments.

Finally, our study context is unique in that it has rarely been explored in previous research on movie marketing despite its significant economic impact. The Hindi cinema or Bollywood is an important and growing industry in a big emerging economy (Punathambekar, 2013). The 2010 KPMG report on Indian Media and Entertainment Industry notes that successful Indian movies get significant portions of their box office revenue from Australia,[3] the UK, the US and increasingly the Middle-East (KPMG International, 2010).

Our results show that both volume of user generated reviews and their valence have significant positive association with box-office performance of the movies. Further, we find evidence for a curvilinear (saturation) effect of the valence of UGRs on box-office performance. We also find evidence for a positive influence by professional critics' opinions. We begin next with development of relevant theory to guide our hypotheses for the influence of user-generated and professional-expert reviews on movie revenues. That is followed by a section describing our data sources and measures and introducing semantic networks based on UGRs. After discussing the validity of UGR concepts, we present our main empirical model and results. The final section provides a discussion and conclusions from the study.

2. Theory development

Search and use of external information in decision making assumes importance in consumption of experiential products like movies, where past performance and observable attributes are less reliable in judging product quality and success. Movies with star-actors are not immune to box-office failures, sequels are not necessarily as popular as the original, and seasoned directors often deliver flops. Evident by their widespread use, consumers increasingly rely on external sources of information in their decision-making such as expert reviews and also, increasingly, on content generated by their peers who have experienced the product. In this section, we will first develop hypotheses about the influence of the two external information sources, followed by a brief description of other factors that have been often posited to have a role in explaining a movie's box office success and which we use as control variables in our model.

We start with comments by peer reviewers or UGRs. Frequency of positive thoughts relative to negative ones, i.e., the positivity ratio, has been found to be important by researchers in the field of emotions (Fredrickson, 2004) and subjective well-being or happiness (Diener et al., 1991). Similar findings are reported in interpersonal, e.g., couples conflict (Gottman et al., 1977) and team

[1] Accessed from http://www.reuters.com/article/2010/06/11/us-industry-idUSTRE65A13Q20100611 and http://www.forbes.com/sites/quora/2014/02/11/how-has-movie-marketing-and-distribution-evolved-over-time/.

[2] As we explain later, "Bollywood" is also referred to as "Hindi Cinema" to indicate that it includes movies produced usually in India with Hindi or Hinglish (combination of Hindi and English) as its main language. In 2013, over 700 Bollywood movies were released for worldwide distribution.

[3] See http://www.theaustralian.com.au/arts/review/bollywood-celebrates-its-100th-birthday/story-fn9n8gph-1226611908187.

performance contexts (Losada, 1999). Losada (1999) finds that teams with high ratio of positive to negative speech pattern during their meetings perform well on outcome dimensions such as profitability, customer satisfaction and assessment by peers. This stream of research points out that assessing the entire conversation for positive and negative emotions, and in particular the ratio of positive to negative thoughts, is a good predictor of outcomes. We seek to test if this also applies in our context. Since we do not rely on a numerical- or star-rating based valence used by prior literature on UGRs about movies, and instead access the reviewers' writing in its entirety, we are likely to catch more nuanced views and can assess positivity ratio of the comments on any given movie. We, therefore, propose the positivity ratio as a measure of valence of UGRs in our context and expect that the intuitive relationship between positivity ratio of the WoM about a movie on the web will be positively related with its box office performance.

Fredrickson and Losada (2005), in further investigating the role of positivity in human flourishing, find the important role for *appropriate negativity* (their italics), and find that there is an upper limit to how much positivity is better. They posit that "without appropriate negativity, behavior patterns calcify". Might there be similar upper limits to positivity in our context as well? We argue that the credibility of positivity as informational value declines with lower instances of negative UGRs. This might be due to suspicion on the part of readers that a forum with only positive UGRs is perhaps infiltrated and controlled, or worse, bought over or otherwise strategically influenced by the promoters of the movies. Thus we propose a curvilinear relationship reflecting a saturation effect on how valence, as captured by positivity ratio, is associated with movie performance.

H1a. The revenue from product consumption experiences is positively associated with the positivity ratio in user-generated-reviews as well as the volume of UGRs.

H1b. The association between revenue and positivity ratio of UGRs becomes weaker with increasing positivity value of UGRs (i.e., negative quadratic effect).

The other external source of information about quality of movies is provided by professional movie critics, who are qualitatively and quantitatively different from average moviegoers. As experts, they are usually employed or retained by large media organizations, and are expected to provide detailed, thoughtful and original analysis of pros and cons of a new release that generally eschews emotive or top-of-mind reactions. Evidence of the role of critics is somewhat mixed, with popular media reporting anecdotes of critics getting it wrong all the time (Bailey, 2012). While Simmons's (1994) reports based on a survey in the US that more than a third of Americans seek the advice of critics when selecting a movie, Eliashberg and Shugan (1997) found that critics have a role in predicting box-office success, but not in influencing it. Generally, given their independent opinions and expertise, consumers perceive critics as credible communicators (Levin et al., 1997). We, therefore, expect that in the Indian context also, there will be positive relationship between critics' rating and box-office revenue, proposing:

H2. The revenue from product consumption experiences is positively associated with the valence of professional reviewers' opinion.

In addition to the two information sources, we also include three other factors in our model as covariates. Genre denotes types or classes of sub-products within a given literary product (Abrams, 1999). It is a defining element in how the movies are produced and distributed (Deasi and Basuroy, 2005). The differences across genres in production costs, width of appeal and type of audience interest is so great that marketing literature has routinely treated genre at least as a control variable (Neelamegham and Chintagunta, 1999). We also use this factor as a control variable and use six different genres including comedy, drama, action and others. Star actors and directors are akin to high-equity brands that enjoy name recognition and positive image (Levin et al., 1997). While it is axiomatic to say that stars are very important in a movie's success in any movie industry, Bollywood not being an exception, attempts to find systematic evidence have often resulted in mixed results (Litman and Kohl, 1989; Ravid, 1999). Ravid (1999) also examines quality signaling or capturing expected rents as two possible mechanisms for star's role in a movie success. Once again, we treat star power of both actors and directors as control variables.

Advertising and publicity done by the movie producers and studio is the final factor we use as a control in our model. These expenses have been climbing and sometimes reported to be as high as 30–40% of the production budget. This helps create pre-release excitement very necessary for a typically short-life product and past literature has also routinely used it in some form as a control (Gopinath et al., 2013; Moon et al., 2010).

3. Data sources and measures

As mentioned earlier, our data come from a different context and our data sources are similar yet completely different from what previous literature has used. India has a longstanding moviemaking tradition, and in fact if one combines the movies made in all different languages every year, the country is by-far the biggest moviemaking nation in the world. With 1255 Indian feature films going through certification in India in 2011 (CBFC, 2011), it left the second largest movie producing country (USA) as a distinct second with only about 520. We focus on the movies made in Hindi language, a segment of the industry now known worldwide as Bollywood, and collect their performance as well as UGRs about them systematically from various sources. We followed in almost real time all 48 Hindi movies released between October 15th 2010 and April 15th 2011.

3.1. Data from conventional sources

One reason why previous research has not looked at this industry in Indian context is the very fragmented nature of the data for various aspects of the analysis. While reliability of cost and profitability data even for Hollywood movies is well known to be dubious,[4] the Indian movie making and distribution industry has long been in quasi-formal sector and hence even the box-office revenue data have typically not been available in any reliable manner. In fact the only published article we could find on this issue uses Bollywood movie's box office sales in two international markets, the US and the UK (Fetscherin, 2010). As a result, we needed to look for specialized sources for various aspects of the movie marketing and performance data, besides spending an enormous amount of time in tracking and analyzing the sentiments from UGRs.

Box office performance data in terms of weekly ticket sales was secured from an industry trade magazine, *Film Information*,[5] which we supplemented and cross-checked with a trade website, Bollywoodtrade.com. The unit of sales in both sources is Crores (i.e., 10 million) of Indian Rupees (1 AU$ ≈ 53 INRs in October 2014) obtained from reports of tickets sold at cinemas all over India. In terms of coverage, these sources do not include revenues from Bollywood movies released simultaneously in some foreign markets,

[4] North American box office data from sources such as *Variety* magazine or online from Boxofficemojo.com is generally considered very reliable, but the cost and profitability data are often referred to pejoratively as Hollywood Accounting (http://en.wikipedia.org/wiki/Hollywood_accounting).

[5] http://www.naachgaana.com/2012/09/03/film-information-komal-nahta-indias-top-film-trade-magazine-on-ett-two-week-total-184-83-crore/.

especially in the US, the UK, and the Middle-East. In our sample, less than one-third of the movies had simultaneous foreign release and for those we could not verify the sales revenues because of inconsistent reports in industry trade magazines. Rather than include data that introduce significant error variance, we restricted our analysis to ticket sales in the Indian market.

Star power of actors and directors of a movie are considered important factors in determining movie revenues in early weeks, and we considered two types of measures for them in line with previous research (Ravid, 1999). One was based on nominations for top 5 critical and popular awards garnered by the actors and directors, including National Film Awards, and those that are hosted by major movie-focused magazines (Filmfare) or TV channels (Zee). Another approach was based on the total box-office revenue (in Rupees Crores) generated by all the movies of the actor/director concerned released within the last three years. Award nomination information was collected from IMDB.com, while the recent history of box-office revenue was collected primarily from the website Bollywoodtrade.com, which was also the back-up source for the current domestic box-office. Movies were divided into six genres based on either a categorization in the magazine Film Information, or in case of multiple-categorizations listed in the primary source, based on judgment of two independent raters about the predominant genre of the movie. A publicity variable was created on a 5-point scale based on descriptive assessment of pre-release publicity and marketing support as reported in Film Information.

3.2. Professional reviewer comments

For professional reviewer comments, we first identified prominent movie critics with large following through exploration of relevant websites and trade publications. Reviews from about half a dozen of these critics, usually associated with leading English language dailies (like Mayank Shekhar of *Hindustan Times*, Nikhat Kazmi of *Times of India*) or major web-based film information sources (like Yahoo! Movies India, Glamsham.com etc.) were tracked. Fortunately, at that time, an Indian website that facilitates buying movie tickets online, bookmyshow.com, conveniently aggregated these prominent reviews and so it was relatively straight-forward to get the full text of these critics' movie reviews. The reviews also included an overall rating on a 1–5 "star" scale, with greater stars indicating more positive evaluation. Not all movies were reviewed by all six critics, but we had at-least two reviews for each movie. Following Litman (1983) and others, we use the overall rating provided by all reviewers on a 5 star scale with half-a-star increments, which we averaged across the available ones for a numerical score between 0.5 and 5. While we chose to use numerical scores for professional reviews in our regression, two independent judges also counted the number of positive and negative sentences in the reviews about various aspects of the movie to create positivity ratios as a way of summarizing their sentiments and present them as descriptive statistics. Critics typically write their reviews originally in English for popular English-language sources (e.g., newspapers, magazines). A legitimate concern is that since the movies are in Hindi language, critics' reviews may not reach the intended movie audience (i.e., those not fluent in English). While this remains a limitation to some extent, two factors ameliorate this limitation. First, urban youth, one of the primary markets for these movies, is invariably bilingual and prone to give greater weight to English-medium conventional and digital media mostly as they relate these sources to Western influence. Second, opinions of prominent movie critics are routinely translated in Hindi for inclusion in Hindi-language newspapers and magazines, allowing their influence to reach a wider audience.

3.3. Textual data: User Generated Reviews (UGRs)

As mentioned before, as a distinct feature of our UGR data, we collected the UGRs on movies from a variety of different sites. The sources included the movie segments of popular web-portals like yahoo.co.in, rediff.com; websites of popular newspapers like *The Times of India* and *Hindustan Times*; and a popular movie centered discussion forum called BollySpice.com, in addition to Twitter and YouTube. For the last two sources, comments were tracked starting three weeks before the release dates. For these two, and for every other source where comments mainly started just after release, comments were tracked till four weeks after the movie release date. Text was scraped from all the web sources mentioned above using customized Java-based programs. As user interface, and behind the scene web-publishing technology, was diverse and in some cases kept changing through the roughly eight months of total data collection, the programming to scrape content needed to periodically adjust.

All scraped text was then organized by movie, source and week first in Excel files and content analyzed and coded for sentiment using the text-processing program Automap developed by the Center for Computational Analysis of Social and Organizational System (CASOS) at Carnegie Mellon University. The text itself included comments that were usually written about certain characteristics of the movie, like actors, director, songs (this being Bollywood), storyline and sometimes simply about the movie itself. These were identified as subject of the sentence or fragments thereof. The primary challenge in this was to accommodate alternate words for the same subject – for example a director "Ram Gopal Verma" may be referred by his full name or as director, "Ram", "Ramu", "Verma", or simply as "RGV." We needed a list of adjectives (words or phrases) that could indicate positive or negative sentiments about a subject. Unlike some other studies that analyze a multitude of products (Seshadri and Tellis 2012), the adjectives were similar and potentially limited in number since the products being commented on were similar. However, a peculiar problem we experienced was the routine mixing of Hindi or Hinglish words (a unique mix of Hindi and English) in otherwise primary use of English-language words. This was most common in the use of adjectives. For example, the use of the words/phrases like "dhansu" or mind blowing for superlative, "bakwaas" for its opposite, and "paisa wasool" for get your money's worth are all examples of phrases used for which no readymade dictionary could be used to capture such idiosyncratic formulations.

Thus we created our own dictionary and coded about 700 words and phrases (that included variations and misspellings) of adjectives that could then be matched with a movie-specific subject list of typically under 100 words per movie to infer positive or negative sentiment being expressed by the user in their comments. Since users sometimes simply wrote a few descriptive sentences with no judgments, and sometimes commented on other commentators or reviewers which were not a judgment about the movie per se, Automap functionality for removing irrelevant words but keeping a neutral placeholder was used to pre-process the text. Das and Chen (2007) report and compare several methods of extracting sentiments from stock message boards. We largely follow what they describe as Adjective-Adverb Phrase Classifier, and which performs reasonably well in their tests. A weighted version of this classifier was coded in Excel Visual Basic and used to further process all pre-processed comments using special capabilities (1) to assign a sentiment to most proximate subject, (2) to count only subject-adjective pairs within the same sentence and (3) to take cognizance of negative qualifiers to reverse the meaning (like "not" and "good" in close proximity were inferred to imply negative sentiment). After offsetting negative and positive sentiment within the same comment, each individual comment was then classified as (net) positive or negative and both the total number of comments and the ratio of positive to negative comments were calculated for comments aggregated across all sources for every movie every week.

3.4. UGR-based semantic networks

Before testing the hypotheses, we examined key concepts and semantic networks of the UGR content as extracted by our Automap-based approach for meaningful and valid representation of user reviews. Since our focus was on the content (e.g., story, director) and evaluative (e.g., good/bad) features of movies, we extracted concepts from UGRs that relate either to nouns/pronouns referring to movie, director, actor and such, or adjectives/adverbs and evaluative phrases that qualify the nouns and pronouns. Since the writing tended to be informal and grammatically convenient rather than correct, we first applied a purpose-built thesaurus to pick the dominant concepts and corresponding evaluative terms. In all, we first mapped the evaluative terms and phrases referred to earlier (about 700 of them) to the valence (e.g., positive/negative) and strength of the evaluation (e.g., superb/pathetic). For example, hats off and amazing were recoded as superb; got it right and really recommend were recoded as better, while interesting and well-made were recoded as good.

Applying a bidirectional semantic network approach, and using sentences as the unit of analysis with concept proximity of up-to five words, we first visualized the network using Automap and its companion software, ORA. Semantic networks for a very small sample of movies, falling on the two ends of the box office success spectrum, is generated and evaluated first. These networks are presented here for illustrative purposes only. Our evaluation was guided by two expectations: (a) semantic nodes in user reviews would represent concepts that characterize central features of a movie including its content (e.g., actor, director, story, music) and evaluative dimensions (e.g., good, bad), and (b) semantic networks would vary predictably for the more successful versus less successful movies (in terms of revenues) such that the former would be associated with greater density and positivity of network linkages.

Figs. 1 and 2 display illustrative semantic networks respectively for successful and unsuccessful movies. Several features of semantic network are specified below to allow meaningful evaluation and comparison of these networks. First, it is important to ensure that the networks are not biased by limited number of comments. Typically, and as expected, a successful movie attracted many times more UGR comments than an unsuccessful movie. To balance this, we combined extracted date from multiple unsuccessful movies (i.e., 7 to be precise) to match approximately the number of comments for one successful movie. Second, to facilitate interpretation, semantic nodes were drawn such that each node's diameter is proportional to the frequency with which it appears in the network. Third, the thickness of the line linking any two nodes is proportional to the relative frequency of that link in the semantic network relative to total frequency of all links.

4. Overall validity of UGR-based concepts

Concepts measured using textual data require careful validity checks to ensure that the subsequent results of hypotheses testing are robust and credible. Textual data based measures rely on validated word libraries (as discussed above), and their measurement validity is ascertained by meaningfulness and discrimination of overall patterns so obtained. From this perspective, the sematic networks obtained from analysis of UGR data provide an opportunity to examine them closely for validation evidence. We discuss this next.

Individually and in comparison, Figs. 1 and 2 indicate that our procedures are effective in extracting meaningful and valid semantic networks of UGRs. Note, as expected and regardless of movie success, an overall gestalt concept of movie itself emerges as a central concept in the semantic network. Moreover, consistent with the unique emphasis of music in Bollywood movies, music is the second most prominent node regardless of movie success (both figures). Other concept nodes that emerge are predictable – actor, story, and director – consistent with their prominent role in characterizing the nature of a movie. In addition, positive and negative evaluative words emerged with significant frequency and are displayed as part of semantic networks. Interesting insights for generalizability and validity emerge from examining the linkages between concepts and evaluative terms, which we address next.

Comparison of semantic networks of a successful movie (Fig. 1) and unsuccessful movies (Fig. 2) provides three insights. First, although the concepts in the two figures appear with comparable frequency (evidenced by their diameter), the evaluative terms differ significantly. The positive evaluative terms (e.g., good, better and superb) appear with significantly higher frequency in a successful movie network (Fig. 1) than in semantic network for unsuccessful movies (Fig. 2). By contrast, the negative evaluative terms (e.g., bad, worse and pathetic) appear more frequently in semantic network for unsuccessful compared to successful movie. Second, the network linkages for successful movie networks show significantly stronger connections with positive evaluative terms relative to the corresponding pattern for unsuccessful movie. For instance, three linkages dominate the semantic network for highly successful movie: "movie-superb," "actor-superb," and "movie-good"). By contrast, the corresponding linkages in semantic network for unsuccessful movies are weak and infrequent. We can, however, observe that there are not as many evaluative words linked with music as there are with movies. The movie concept is linked with high-frequency with predominantly positive evaluative words in Fig. 1, though there are connections with negative evaluative words as well. Similarly, both positive and negative concepts are linked with movie in Fig. 2 as well; however, relative strength of positive to negative concepts is much weaker in these unsuccessful movies. Third, the semantic network density for successful movie shows positive asymmetry while for the unsuccessful movies it is diffused and negatively asymmetric. Specifically, Fig. 1 shows that for a successful movie the strength of network linkages is concentrated on the right hand side where linkages involve positive evaluations. By contrast, for unsuccessful movies, the strength of network linkages is more diffused but still shows a relatively higher concentration on the left side where linkages involve negative evaluative terms. Taken together, these three insights confirm that the two semantic networks reasonably capture the key conceptual and evaluative themes in UGRs, and their linkages correspond to expectations about the likely prevalence of positive and negative sentiments of successful and unsuccessful movies. We next present the empirical model to test our formal hypotheses presented in previous sections.

5. Empirical model and results

5.1. Empirical model

A pooled regression was run for all the 48 movies for all the weeks together whenever positive revenue was reported. Since many movies survived for less than four weeks, the total number of observations (movie-weeks) is 139. Log of the weekly domestic box office revenue was the dependent variable and genre dummies, other covariates typically identified by previous literature and our variables of interest were the independent variables. The regression equation identifying time-varying and time-constant elements in the estimation is given below:

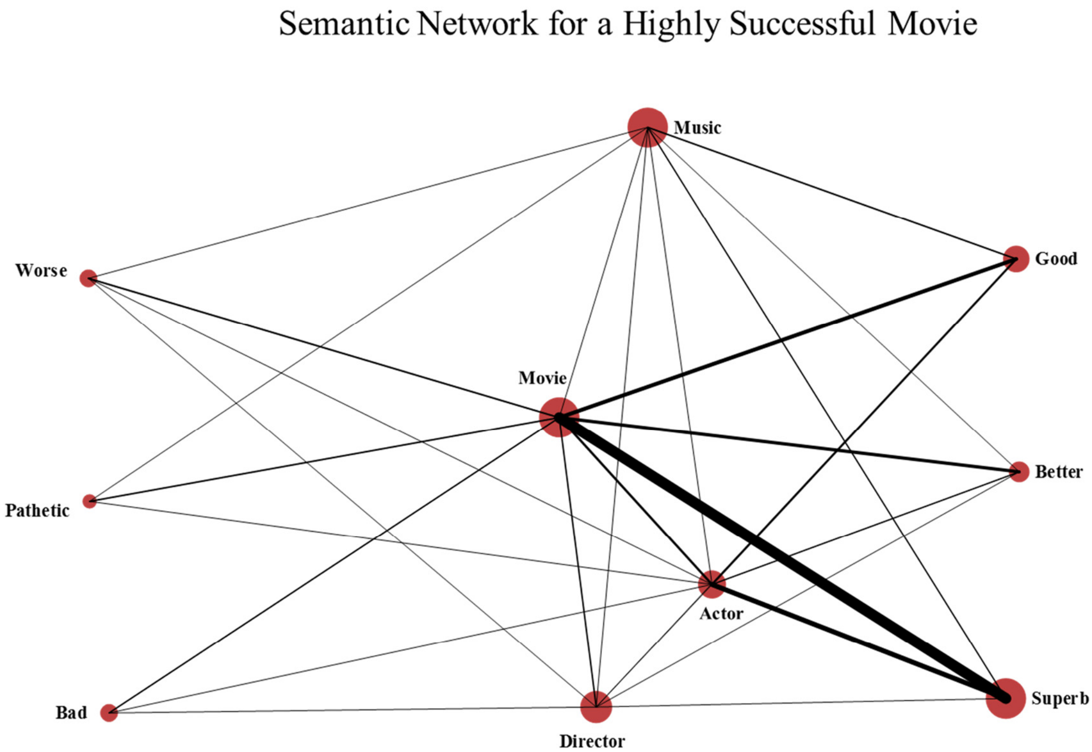

Fig. 1. Semantic network for a highly successful movie.

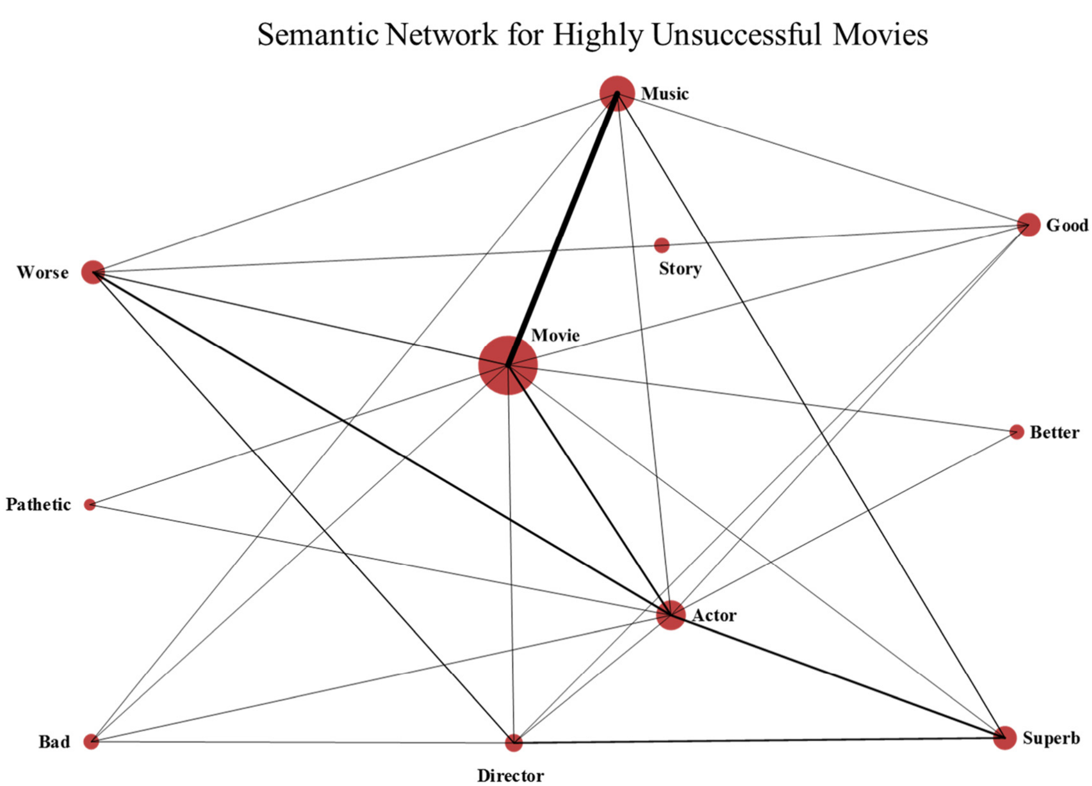

Fig. 2. Semantic network for highly unsuccessful movies.

$$\text{Log}(\text{Rev})_{it} = \alpha + \sum_{k=1\text{ to }5} \gamma_k \text{Genre Dummy}_k + \beta_1 \text{Publicity}_i + \beta_2 \text{Running Week}_{it} + \beta_3 \text{Director Star Power}_i + \beta_4 \text{Actor Star Power}_i + \beta_5 \text{Critic Rating}_i + \beta_6 \text{Log}(\text{Volume of UGRs})_{it} + \beta_7 \text{Positivity Ratio of UGRs}_{it} + \beta_8 (\text{Positivity Ratio of UGRs})^2_{it} + \varepsilon_{it}$$

In the equation above, i is the index of the movie, t is the index for the movie's running week and k is the index for the genre of the movie. Notice that there are a total of six genres of the movies so k can take values of 1 to 5 only with the sixth one being the baseline. The variables publicity, director star power, actor star power and critic ratings are time-invariant as they are fixed for the entire duration of the movie. The values of both publicity and critical reviews are out just after the movie release and remain constant throughout the movie run. Director and actor star powers are known for the movie even pre-release and do not change. Running Week takes the value between one and four indicating how recently the movie was released, and was included to capture and filter-out a very strong downward trend in box office revenue for all movies. Since UGRs are tracked and analyzed week by week, the volume and positivity ratio related measures are week-specific.

5.2. Descriptive characteristics

Some description of revenue, by genre of the movie, is in Panel A of Table 1. Summary statistics about the measures used in the study is provided in Panel B of the Table 1. Comedy turns out to be the most popular genre both by number of movies made, and for total revenue. Drama and romance are other more popular genres. Looking at other data, about half the movies do not have significant reportable movies beyond three weeks. While the distribution of number of reviews is very spread out and has high standard deviation, it is interesting to note that overall, there are three to four times more positive online UGRs about a movie than negative ones. This is consistent with findings in other contexts where similar or even higher margins of positives to negatives have been reported (Keller, 2007). There is also a lot of variability in weekly and total box office revenue for the movies that ranged from a low of INR 0.02 Crores to a high of INR 106 Crores. Also, most movies make much lower amounts in box office revenue compared to the average (of INR 13.88 Crores), the median being less than 27% of the average. On a week by week basis, box office revenues come down by about 70% on an average from week 1 to week 2 and the trend continues in future weeks. The volume of UGRs, however, shows a slight uptick in week 2, before coming down in weeks 3 and 4, while its positivity remains roughly constant on an average week to week. Professional reviewers generally give more negative verdicts on the overall quality of the movies, and the range goes from 0.5 (half a star) to 3.67, indicating that there was considerable disagreement among experts about the quality of the movies at the higher end of the quality spectrum.

5.3. Regression results and test of hypotheses

The regression results are summarized in Table 2. The model fits the data quite well with adjusted-R^2 in excess of 0.75. A large negative parameter estimate for running week indicates the usual downward weekly trend in box-office receipts, as noted earlier ($B = -1.22$, $p < .001$). Pre-release publicity and other marketing efforts also have a robust positive influence on box office revenue ($B = 0.40$, $p < .01$). We had included the star power of both the director and the actors in the movie in the regression. Both were operationalized by calculating recent box-office performance of the actor/director. Both effects are positive and significant ($B = .001$ and $.0005$ respectively, $p < .05$). This effect remains positive and significant with the

Table 1
Data description and summary statistics.

Panel A: Movie summary by genre

Genre	No. of movies	Average total-run revenue (Rs. crore[a])
Action	11	4.71
Animation	3	1.67
Comedy	13	32.28
Drama	13	7.78
Horror	2	0.28
Romance	6	14.65

Panel B: Summary of revenue and user review data

48 Movies, up to 4 weeks of data N = 143 Item description (unit)	Minimum	Maximum	Median	Mean	Standard deviation
No. of weeks of positive revenue	1	4	3	–	–
Positive user reviews in a week	1	3094	89	374	586
Negative user reviews in a week	0	2263	41	155	340
UGR positivity ratio (N = 139)	0.23	13.0	3.1	3.5	2.3
Professional reviewer ratings	0.5	3.67	2.50	2.35	0.72
Profession reviewer positivity ratio	0.07	10.83	0.82	1.73	2.29
Director star power (Rs. crore)	0	117.7	0	17.6	30.1
Actor star power (Rs. crore)	0	238.1	55.4	66.5	63.7
Weekly box office revenue (Rs. Crore)	0.01	62.50	0.71	4.66	9.28
Total box office revenue over the run (Rs. Crore)	0.02 (Bhoot & Friends)	106.0 (Golmaal 3)	3.73	13.88	20.80

[a] A crore is a commonly used number in India, and equals 10 million. Thus using approximate conversion rate at the time, one crore Indian Rupee is approximately US Dollars 200,000.

Table 2
Regression results.

DV: Log (weekly box office collections in Rs. crore)
Model F-value = 33.50***,[a]
N = 139
R^2 = 0.777
Model degrees of freedom: 13[b]
Adjusted R^2 = 0.754

Independent variable	Parameter	Standard error	t-Value
Publicity	0.40***	0.10	3.94
Running week	–1.22***	0.11	–10.95
Director star power	0.001**	0.00	2.41
Actor star power	0.0005**	0.00	2.57
Professional review rating	0.77***	0.20	3.91
(Log) volume of UGR comments	0.30***	0.09	3.26
UGR positivity ratio	0.49***	0.14	3.49
(UGR positivity ratio)^2	–0.03***	0.01	–3.16

[a] **Statistically significant at <.05; ***Statistically significant at <.01.
[b] Model estimated with genre dummies as well. Those parameters are suppressed for conserving space.

other, award nomination based measure of star power as well. Next we address the effects of the hypothesized variables of interest pertaining to information-sources that consumers have about the movie.

First, consider Hypothesis 1 about the impact of UGRs. Like most extant studies on this effect, volume of comments is found to be positive and highly significant ($B = .30$, $p < .01$). In addition, we find statistically significant parameters for both contemporaneous positivity-ratio and its squared term ($B = 0.49$ and $-.03$ respectively, $p < .01$). The squared term is negative, which indicates a concave, or tapering off effect, which is consistent with the idea that some negativity is good in the effect of positivity ratio of UGRs on box office revenue. Thus, not only do we find positive effect of both volume and valence on weekly box-office (support of H1a), but our specific hypothesis about curvilinear effects (H1b) is also supported.

Finally, we note that the effect of Professional Review Ratings is also positive and significant ($B = .77$, $p < .01$). This supports our hypothesis 2. We tested for its quadratic effect but did not find any support for a curvature in this effect. As such, the quadratic term was dropped from our final model in the interest of parsimony. In summary, we find support for positive informational effects for both UGRs and professional review ratings in moviegoers' decisions. In addition, we find both the volume and valence of UGRs are important; albeit, the positive effect starts tapering off at high positive valence.

6. Discussion

This study aimed to examine the simultaneous influence of user generated content and professional critics on the success of Bollywood movies in India as measured by weekly revenues. Our study advances past research by showing that both volume and valence of user generated content contribute significantly to movie success after controlling for the effect of professional critics (which is also positive and significant) and for other alternative explanations found to be important in past research. Moreover, this study goes further to show how volume and valence differ in their influence on movie success. The insight here is that valence is more effective when it is not unabashedly positive. Rather, an increasingly balanced valence of user generated content that blends positive comments with negative notations while maintaining a net positive tone is apparently more credible with peers to motivate them to engage in consumption decision. In other words, consumers are prone to discount unbalanced comments from peers that focus solely on positive feelings especially later in a movie's life cycle. Expert opinions of professional critics are valued by movie audience. Their independent positive effect suggest that movie audiences evaluate peer-to-peer feedback and opinions of professional critics for different reasons, and are able to integrate these different sources of information in their decision to invest in consumption of a movie experience. While the scales for both sources are different so direct comparison in effect size is not very reliable, we do see that the relative effect of the two are comparable in magnitude.

The preceding results clarify past findings that have yielded mixed and muddled effects of valence and volume effects of peer-to-peer influence. These clarifications stem from several unique features of our study. First, this study utilizes multiple sources spread over different type of websites, including newspapers, movie section of major Internet portal, prominent movie focused blogs, YouTube and Twitter. The latter two have especially become major sources of movie related user discussions in recent years. Also, since we extract sentiments directly from the comments, we do not rely on numerical or star ratings as has generally been the case in the past. Instead, we divide individual comments on positive or negative based on the sentiment expressed in a comment and then take their ratio as a measure of the valence of the peer comments. Since our interest is also in identifying the impact of professional reviews, we also capture those from multiple sources and investigate their influence on the movie box office revenues.

Second, in order to examine the focal relationships after controlling for other known factors, we pick appropriate covariates identified by prior literature that include genre of the movies, marketing effort in promoting the movies as well as star power of actors and directors. We find all the covariates have impacts in expected directions. Contrary to prior literature, we find support for both volume and valence of UGRs being positive and significantly related with revenue. We are the first to look at more nuanced higher order effects of volume and valence on box office revenue by including higher order terms to examine saturation or threshold effects. In addition to the linear positive effects, we also find a saturation effect in the impact of valence, signifying that some degree of negativity in user comments is actually helpful to movie performance, perhaps for credibility of commentators. This parallels the finding of "appropriate negativity" in communication patterns researchers have found helpful in contexts such as conduct of business meetings or couples conflict (Fredrickson and Losada, 2005).

Third, this study contributes to the literature by bringing a new measure of valence in the UGR/movie literature from personal and group psychology literatures. Since peer communications on the Internet is an important way on how people get their information and form opinions, we believe that this richer mode of sense-making going beyond simple numerical ratings is an important avenue of understanding communications in more varied contexts. Our findings of both valence and volume being positively associated with box office performance is reassuring and signifies that people perhaps want to watch not only movies that are popular and hence draw a lot of user generated content, but also movies that early audiences like. This also reconciles the contradictory finding in prior literature that found either volume or valence of UGRs to be significant. Beyond user generated reviews, we also find the role of professional critics and find evidence that their opinions and ratings are also positively associated with movie success. While this may not be the last word on their influence, at least in this context it seems that the wisdom of the crowd has not supplanted the wisdom of the wise. Perhaps all these significant effects at the aggregate level point toward existence of multiple segments of movie-goers; some simply want to watch highly anticipated and popular movies, while some others wait for their peers to weigh-in on the quality of movies and some others get influenced by what professional knowledgeable critics have to say.

Finally, this study also contributes to research by using multiple sources of social influence beyond just relying on predominantly Yahoo! Movies and the numerical/star ratings therein. Since a machine-learning assisted sentiment extraction strategy is utilized, we are able to utilize multiple sources which may have their own different styles of eliciting user responses and often do not have a star rating system. We believe that borrowing the positivity ratio idea to get at valence provides us with a robust tool to utilize multiple sources, with or without length limitations in the user generated comments. Thus, Twitter comments, restricted to 140 characters, can be used side by side with entries on movie-themed blogs which could be quite expansive and could go to several hundred words. Finally, our use of a different emerging marketing context helps generalize findings in the movie marketing area. India has been a major movie-producing country for many decades. Many elements of entertainment marketing are equally relevant in that market. At the same time, the Indian movie industry is becoming global. Rigorous research studies focused on this industry are rare, primarily due to data and other institutional constraints discussed in our research method section. Thus, a systematic study focused on this important part of the entertainment market is also one of our contributions.

While an important step toward mitigating the paucity of an important part of the market, our study also has certain limitations that we hope future researchers will address. First, our data are for a limited number of movies and larger, more comprehensive, studies would be helpful in confirming our findings. We did focus on all movies released for a domestic box-office during a six month period and hence our analyses are more comprehensive, leading to some confidence about our findings. Secondly, we utilized one method of sentiment extraction from the unstructured comments and future researchers could use new emerging methods including naïve-Bayes methods in their studies. In our very specialized context, our customized dictionary creation based approach worked reasonably well. However, as technology progresses in this area, more refined approaches may be utilized.

As user generated content becomes one of the dominant ways consumers use the Internet, we expect other aspects of WoM like personal networks and recommendations using friend networks on specific sites, like Facebook, for example, to be a significant direction of future work. We also expect that examination of relative influences of critics, user generated reviews and marketer initiated communications over time will be examined more carefully by future researchers.

Acknowledgements

We thank Nitin N S Jain and S. Abid of bollywoodtrade.com for some discussion on box-office sources and figures, Weatherhead Research fund for financial support and the following research assistants: Priyanka Pawar for developing and implementing UGR collection effort, Suraj Shetty, and Suchit Verma for finding all other information and especially Sandeep Kumar for painstakingly processing all the UGR data in Automap and specially developed Excel Macros to calculate weekly level variables used in the model.

References

Abrams, M.H., 1999. A Glossary of Literary Terms. Harcourt Brace, Orlando, FL.
Bailey, J., 2012. 12 great movies the critics got dead wrong. Available from http://flavorwire.com/272219/12-great-movies-the-critics-got-dead-wrong?all=1 (accessed 12.12.13.). Favorwire.com.
Basuroy, S., Chatterjee, S., Ravid, S.A., 2003. How critical are critical reviews? The box office effects of film critics, star power, and budgets. J. Mark. 67 (October), 103–117.
Central Board of Film Certification (CBFC), 2011. Government of India – annual report.
Chakravarty, A., Liu, Y., Mazumdar, T., 2010. The differential effects of online word-of-mouth and critics' reviews on pre-release movie evaluation. J. Interactive Mark. 24 (3), 185–197.
Chevalier, J.A., Mayzlin, D., 2006. The effect of word of mouth on sales: online book reviews. J. Mark. Res. 43 (3), 345–354.
Chintagunta, P.K., Gopinath, S., Venkataraman, S., 2010. The effects of online user reviews on movie box office performance: accounting for sequential rollout and aggregation across local markets. Mark. Sci. 29 (5), 944–957.
Das, S.R., Chen, M.Y., 2007. Yahoo! For Amazon: sentiment extraction from small talk on the web. Manage. Sci. 53 (9), 1375–1388.
Davidson, A., 2012. When you wish upon 'Ishtar'. New York Times Sunday Magazine, July 1, 2012. Pg. MM 16.
Deasi, K.K., Basuroy, S., 2005. Interactive influence of genre familiarity, star power, and critics' reviews in the cultural goods industry: the case of motion pictures. Psychol. Mark. 22 (3), 203–223.
Dellarocas, C., Narayan, R., 2006. A statistical measure of a population's propensity to engage in post-purchase online word-of-mouth. Stat. Sci. 21 (2), 277–285.
Dellarocas, C., Zhang, X.Q., Awad, N.F., 2007. Exploring the value of online product reviews in forecasting sales: the case of motion pictures. J. Interactive Mark. 21 (4), 23–45.
Dhar, V., Chang, E., 2009. Does chatter matter? The impact of user-generated content on music sales. J. Interactive Mark. 23 (4), 300–307.
Diener, E., Sandvik, E., Pavot, W., 1991. Happiness is the frequency, not the intensity, of positive versus negative effect. In: Strack, F. (Ed.), Subjective Well-Being: An Interdisciplinary Perspective. Pergamon Press, Oxford, pp. 119–139.
Eliashberg, J., Shugan, S.M., 1997. Film critics: influencers or predictors? J. Mark. 61 (2), 68–78.
Fetscherin, M., 2010. The main determinants of bollywood movie box office sales. J. Glob. Mark. 23, 461–476.
Fredrickson, B., 2004. The broaden and build theory of positive emotions. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 359, 1367–1377.
Fredrickson, B., Losada, M., 2005. Positive affect and the complex dynamics of human flourishing. Am. Psychol. 60 (7), 678–686.
Godes, D., Mayzlin, D., 2004. Using online conversations to study word-of-mouth communication. Mark. Sci. 23 (4), 545–560.
Godes, D., Silva, J.C., 2012. Sequential and temporal dynamics of online opinion. Mark. Sci. 31 (3), 448–473.
Gopinath, S., Chintagunta, P.K., Venkataraman, S., 2013. Blogs, advertising, and local-market movie box office performance. Manage. Sci. 59 (12), 2635–2654.
Gottman, J.M., Markman, H., Notarius, C., 1977. The topography of marital conflict: a sequential analysis of verbal and non-verbal behavior. Marriage Fam. 39, 461–477.
Keller, E., 2007. Unleashing the power of word of mouth: creating brand advocacy to drive growth. J. Advert. Res. 47 (4), December, 448–452.
KPMG International, 2010. Film: on the path to resurrection. Chapter 2 from FICCI-KPMG Indian Media and Entertainment Industry Report.
Levin, A.M., Levin, I.P., Heath, C.E., 1997. Movie stars and authors as brand names: measuring brand equity in experiential products. Adv. Consum. Res. 24 (1), 175–181.
Litman, B., 1983. Predicting the success of theatrical movies: an empirical study. J. Pop. Cult. 17, 159–175.
Litman, B., Kohl, L.S., 1989. Predicting financial success of motion pictures: the '80s experience. J. Media Econ. 2 (2), 35–50.
Liu, Y., 2006. Word of mouth for movies: its dynamics and impact on box office revenue. J. Mark. 70 (3), 74–89.
Losada, M., 1999. The complex dynamics of high performance teams. Math. Comput. Model. 30, 179–192.
Moon, S., Bergey, P.K., Iacobucci, D., 2010. Dynamic effects among movie ratings, movie revenues, and viewer satisfaction. J. Mark. 74 (January), 108–121.
Moul, C., 2007. Measuring word of mouth's impact on theatrical movie admissions. J. Econ. Manag. Strategy 16 (4), 859–892.
Neelamegham, R., Chintagunta, P.K., 1999. A bayesian model to forecast new product performance in domestic and international markets. Mark. Sci. 18 (2), 115–136.
Punathambekar, A., 2013. From Bombay to Bollywood. New York University Press, New York. Published in Postmillenial Pop series.
Ravid, S.A., 1999. Information, blockbusters, and stars: a study of the film industry. J. Bus. 72 (4), 463–492.
Seshadri, T., Tellis, G., 2012. Does chatter really matter? Dynamics of user-generated content and stock performance. Mark. Sci. 31 (2), 198–215.
Sharda, R., Delen, D., 2006. Predicting box-office success of motion pictures with neural networks. Expert Syst. Appl. 30 (2006), 243–254.
Simmons, J., 1994. A 'thumbs up' pulls in the audience. Wall St. J. B1.