Are you A 'Viral Star'? Conceptualizing and Modeling Inter Media Virality

Amit M. Joshi

Michael Trusov*

July 12, 2011

* Amit M. Joshi (ajoshi@bus.ucf.edu, phone: (407) 823-5355) is an Assistant Professor, College of Business Administration, University of Central Florida, P.O. Box 161400, Orlando, FL 32816;

Michael Trusov (mtrusov@rhsmith.umd.edu, phone: (301) 405-5878, fax: (301) 405-0146) is an Assistant Professor, Robert H. Smith School of Business, University of Maryland, College Park, MD 20742.

The authors gratefully acknowledge the financial support of the Marketing Science Institute.

Are you A ‘Viral Star’? Conceptualizing and Modeling Inter Media Virality

The spread of social media has meant that User-generated Content (UGC) has become an important form of communication. Most of the past research in social media has concentrated on analyzing *message virality* or the causes of why certain messages go viral. We posit that accounting for *media virality*, a phenomenon where messages are transferred beyond the original media they were carried in, has become imperative. In this paper we argue that media virality is a function of product characteristics and propose a framework for capturing this type of virality using just two dimensions, the inter media elasticity and inter media duration, together referred to as an entity’s Inter Media Reactivity (IMR).

We illustrate the application of our concept using data on movie stars across several media. We calculate the IMR for each star, and demonstrate how media virality differs across each, and also analyze the star-specific characteristics that drive media virality. Subsequently, we relate the media virality of stars to the performance of their movies. Our research thus provides a theoretical contribution to the literature by exploring media virality while also providing several managerially relevant substantive insights about the motion picture industry.

Key Words: Social media, advertising, motion picture industry

Introduction

In August 2007, Cadbury Schweppes, the leading British confectionary manufacturer, launched a £ 6.2 million advertising campaign for their Cadbury's Dairy Milk brand that featured a Gorilla playing drums. Though several reviewers considered it abstract ("Spot the link between a gorilla and chocolate" *The Independent*, 05/14/2007), the campaign was a huge success, generating 500,000 Youtube views in its first week, spawning several spoofs ("Net fans go ape over gorilla ad" *Birmingham Mail* 11/02/2007), and reportedly increasing the sales of the Dairy Milk brand by 9% ("Cadbury's ape drummer hits the spot" *Media Week*, 09/25/2007). The campaign clearly 'went viral' and Cadbury spokesman Tony Billsbrough claimed that although the actual advertising campaign was mainly run on television, "*It was always the plan that it would become a bit of a cult hit…"*

In sharp contrast, in March 2009, Skittles re-launched its website in a way that enmeshed the brand with social media ("Skittles cozies up to Social Media" *The Wall Street Journal,* 03/02/2009). The idea was to leverage organically generated social media content to build the brand. Specifically the site featured content created by users on Youtube, Facebook Twitter and Flickr. All Twitter users who used the word "Skittles" had their tweets featured on the home page. This social media experiment started failing just a few hours into the campaign, with tweets like "*That Skittlesthing is so six hours ago*" and "*Congratulations Skittles for lowering the bar for terrible ideas*", until the campaign had to be shut down just a day later, when consumers started tweeting the word "Skittles" along with profanities ("What are you doing (#Skittles)?" *The Wall Street Journal*, (03/03/2009).

The examples above present interesting contrasts, but also have several similarities. On the one hand, Cadbury's campaign originated with a TV ad that then caught on with social media and was immensely successful. The Skittles campaign started with social media, never caught on in the commercial media and was a failure. However, both campaigns were for products in the fast-moving consumer packaged goods category, both were conceived with the intention of 'going viral' and both received substantial coverage in the print news media. What can then explain the difference in consumer reactions to the two campaigns above? Clearly, campaign-specific factors and execution would have played a part in the success / failure of the advertisements in ‘going viral’ and their eventual impact on product performance. However, is it also possible that certain brands are inherently more likely to ‘go viral’, especially in certain media? Would Cadbury’s have been as successful had they begun their campaign online instead of on TV, and would Skittles have still failed if they started on TV instead of on social media? Or was the Cadbury’s campaign destined for success and Skittles for failure irrespective?

The two examples above illustrate both the necessity of managing social media as well as the inherent problems therein. At a very basic level, managers may need to understand whether their brands are better suited to being advertised in social media or commercial media. Media decisions are often taken at an aggregate level and well before results of such expenditures are known. As a consequence, knowing that a particular medium may more efficiently transmit messages about a brand would be a powerful tool for marketers in deciding future promotion portfolios. In this research we demonstrate that some brands are inherently ‘social media oriented’ such that messages for these brands originating in social media are often picked up by commercial media. Other brands are ‘commercial media oriented’, wherein the reverse frequently occurs. To gain an understanding of why this may occur, we begin with a brief discussion on social media and virality

Social Media and Virality

The features that differentiate social media and user-generated content (UGC) from commercial media are (1) accessibility of the medium; (2) consumer-driven creativity of message; and (3) non-professional orientation (OECD 2007). A characteristic that is closely associated with UGC and one that is important to managers and academics alike is the concept of virality. Virality can be parsimoniously defined as the spread of a message beyond its original recipients (Dobele et al 2005) and is related to UGC since consumers have a much greater level of control over the message content and who it is disseminated to. Previously, an interesting advertisement or controversial news would be consumed relatively passively, with the only modes of consumer reaction possibly being letters to newspapers and news stations or verbal communication (word-of-mouth) among a comparatively small group of people (the two-step inter-media theory proposed by Katz and Lazarsfeld, 1955). Today, however, the web has empowered the consumer such that they have a variety of options to respond, discuss or retaliate to the mass media, using blogs, discussion forums or twitter feeds (Kaplan and Haenlein 2010).

What drives successful 'viral' campaigns or messages? Recent and past research notes several factors that may influence the 'virality' of a message[1]. Important drivers include strength of ties to the referrer for both traditional (Granovetter 1973) as well as electronic referrals (DeBruyn and Lilien 2004); geographical proximity rather than demographic proximity (Yang and Allenby 2003); customer loyalty (Bowman and Narayandas 2001); message and product characteristics (Dobele et al 2005; Berger and Milkman 2010, Stephen and Berger 2010, Berger and Schwartz 2010); and product and merchant ratings (Chevalier and Mayzlin 2006).

All of the research cited above focuses on studying *message* virality. However, another kind of virality that has not received much attention in academia is inter-media virality. If virality is the spread of information beyond its original recipients, media virality would imply the transfer of the same core message beyond the original media it was carried in. For instance, an interesting or controversial advertisement may be uploaded on Youtube, discussed on several blogs and heavily searched for using Google. With the popularity of UGC this type of virality has become important, with consumers commonly using various social media (such as Youtube, Twitter or blogs) to discuss and spread information obtained from other (generally commercial) media. While the transfer of information from commercial media to social media through UGC is easy to visualize and understand, we are now also observing media virality from social to commercial media. Indeed, social media and UGC are now often the original source of information and are widely cited in mass media such as newspapers ("Staying

[1] Please see Leskovec et al 2007 (pp 4-7) for an excellent review of drivers of virality.

Informed Without Drowning in Data", *The New York Times*, 12/18/2008). The proposition that social media has become a strong driver of commercial media is also supported by private conversations with leading industry practitioners – for example, it is very common for magazine and newspaper reporters to search for new stories on blogs and discussion forums.

Why is the concept of media virality important for managers and academics alike? As demonstrated in the examples above, the empowered consumer and the consequent media virality has meant that a firms' control over its communication mix has been diminished. A message in one medium can quickly be picked up by another medium, and the sum of the messages in various media can have synergistic effects (Naik and Peters 2009), leading to several new and important implications for academics and managers. First, this implies that media are now even more interdependent, making the various sources of information endogenous. Second, the direction as well as duration of information flow (social to commercial or vice versa) may have as much of an impact on eventual message dissemination as the content of the message. Thus, information passing from a commercial medium such as print to a social medium such as a blog may have a different impact than that passing from a blog to print media, causing the media reactions to be asymmetric depending on the *direction* of information transfer (social to commercial or vice versa). This implies that the traditionally held knowledge regarding media effectiveness and communication mix optimization may now be outdated under the presence of such media virality.

How then, should firms design and control the overall communication mix? For which entity or product will a message in commercial media (say, print news) lead to a flurry of UGC and therefore 'buzz' in social media (such as blogs) or vice versa, *on average*? Will that cross-media dissemination be symmetric from commercial to social and social to commercial or not? If there is indeed asymmetry in cross-media dissemination, which media will lead in information transfer and what will that asymmetry depend upon? How long will that buzz last in the other media? To effectively manage media expenditures and therefore control product performance, it is important that managers know the answers to these questions.

While the media landscape has dramatically changed in the past decade, our "…measurement systems are struggling to catch up…" (Sharp and Wind 2009, pp.120). In this research, we propose that media virality can be captured parsimoniously by using just two dimensions associated with an entity – its inter-media elasticity and inter-media duration, which are together referred to as Inter Media Reactivity. To empirically estimate these dimensions and therefore understand media virality, we employ traditional methods used for market response models (please see Hanssens, Parsons and Schultz 2001 for a detailed discussion). When studying evolving markets, it is common to link the change in the volume of advertising over time to product performance (for instance Lodish et al 1995) while implicitly assuming that average advertising quality is common across competitors and therefore excludable from the model. Similarly, we use econometric models to relate the fluctuations in volumes of social and commercial media over time to understand media virality. As with market response models, we ignore the valence of information and execution details of a particular media campaign[2], but instead consider the average reaction of one medium to others over time, enabling us to make predictions about media virality at the level of each individual product. We empirically answer the questions raised above by applying our method to the entertainment industry by calculating the media virality of movie stars. Our choice of industry is driven equally by the managerial and academic importance of media virality in this particular industry and the availability of data over an extended period of time, across several media.

The rest of this paper is organized as follows. We begin by reviewing literature on social media, and its synergies with commercial media and use that to motivate our conceptual framework where we extend the concept of message virality to the idea of media virality. We then discuss our data and estimation approach, followed by our results. After a discussion of our main results, we demonstrate how

[2] In subsequent analysis, we include valence to test the robustness of our findings.

our measure of media virality impacts product performance and the key managerial implications of our findings. We finish with limitations and suggestions for future research.

Social Media and Commercial Media

The rise in interest in social media and its relationship with commercial (or traditional) media has been well reflected in academic research with several studies analyzing the impact of social media on performance. One of the earlier papers in the field, Godes and Mayzlin (2004), uses conversations from Usenet to predict ratings for TV shows. The authors acknowledge the endogenous nature of WOM, but find no relationship between WOM and future ratings. In contrast, Liu (2006) finds that volume of WOM (from Yahoo! movies message board) offers significant explanatory power when predicting both opening weekend as well as total revenue for motion pictures. In a similar vein, Duan et al (2008) find that while the valence of online ratings has no significant impact on movie's box office revenue, the volume of these ratings influence the revenue, indicating that an awareness effect rather than an information effect drives consumers.

While the above studies mainly looked at social media in isolation, other research has investigated social media in conjunction with commercial media. Even though the study of the impact of various media on each other predates social media (specifically, the research stream on Integrated Marketing Communications, IMC), this research has typically aimed at identifying and analyzing the synergies between various media, without looking at media virality (see for instance, Naik and Raman 2003). Recent research in IMC has incorporated new media, but UGC has not been analyzed (for instance, Naik and Peters (2009) analyze the synergy from advertising in traditional and online media). In contrast, research on social media that also incorporates commercial media has mainly focused on product performance rather than media virality[3]. For instance, Stephen and Galak (2010) analyze the interaction of

[3] Hereon, we use the terms social media to imply user-generated content on social media.

commercial media (specifically, print media mentions and TV events) with blog posts and discussion forum posts for a financial lending website and conclude that commercial media has a large impact on product performance per event, while the impact of social media is considerably smaller in magnitude. In contrast to this study, Trusov et al (2010) find that the elasticity of WOM is about 20 times larger than that for commercial media. In addition, this research finds that the WOM effects also have larger carryover effects than traditional media. Table 1 details the various studies dealing with social media.

Insert Table 1 here

Two recent studies have investigated the impact of valence and the possible synergy effects between social and traditional media. Using market-level (DMA) data for movie launches, Chintagunta, Gopinath, and Venkataraman (2010) find that the valence of movie reviews are bigger drivers of box office performance at the local level, than volume. However, when analyzed at the national level, their findings mirror past research that volume trumps valence. The importance of volume over valence is also reported by Onishi and Manchanda (2010), who find that the impact of valence (.01) is overwhelmed by the impact of volume (0.40) when predicting movie sales volume in the Japanese market. Furthermore, this research finds that blogs and TV advertising work synergistically when driving sales, across three product categories.

As noted previously, the focus of all the studies cited above is to understand the impact of different media and their synergies on product performance, but does little to understand how messages may travel from one medium to another. Most of the past research in social media or UGC has concentrated on analyzing *message virality* or the causes and drivers of why certain messages go viral, while ignoring the *media virality*. We posit that media virality, or the spillover of information from one medium to another is especially important in an era where UGC is a key method of communication. From a marketing perspective, new product information, updates or recall information is typically carried in commercial media (such as news), and can spillover into social media via UGC. On the other hand, product reviews, opinions and the latest trends may originate in social media and may then spillover into commercial media.

If products are represented on a media map on the basis of the volume of information about that product in each media at a given point in time (Figure 1a)[4], then the change in that product's location in response to a change in information volume in one medium can take two different paths. In a non-interdependent world, where there exists no media virality, a change in media volume is simply captured by a shift of the entity along that media dimension alone (Figure 1b). In a more realistic, UGC driven media-viral world however, such a change may also lead to a change in one or more other media, as shown in Figure 1c.

Furthermore, recent research has indicated that entity characteristics may play an important role in driving media virality. While extant research on WOM and virality was focused on the performance impact of social transmission, latest research has found that characteristics such as emotional appeal (Berger and Milkman 2010), general interest in a product ('talkability'; Stephen and Berger 2010) and omnipresence (Berger and Schwartz 2010) all play a role in driving virality. Applying these findings to the framework above implies that the same change in focal media volume for two entities in the same industry, in the same time period may lead to different media interactions, as exemplified in Figure 1d.

Insert Figures 1a-1d here

Based on the research above, we propose that entity specific characteristics determine media virality for that entity. Thus, for two products in the same industry the same message, possibly having identical content (say for example, a bad review or product recall) may show completely different cross-media dissemination. We propose using two dimensions to succinctly capture an entity's media virality – its inter-media elasticity and inter-media duration. Together, we term the elasticity and duration as an entity's Inter Media Reactivity (IMR), since they capture how any medium reacts to information about

[4] While the figure shows the 3 media as orthogonal for expositional purposes, we acknowledge that all media are clearly not orthogonal. In the analysis that follows we do not assume any properties of a geometric system.

that entity in another medium and for how long it remains stimulated in response. We discuss both the elasticity and durations in detail next.

The first dimension of IMR that captures media virality – the inter media elasticity – determines how a media responds to information about an entity in another media. The inter media elasticity captures both the magnitude of the impact as well as direction of impact. For instance, a significant and positive social-commercial elasticity for an entity would imply that information about that entity in social media (for e.g. blogs) is typically picked up by commercial media (like news). An insignificant elasticity would imply that information for that entity does not travel through across that media route. The elasticity metric therefore explicitly captures the information flow across media over time, which is a necessary condition for understanding the causality of information transfer. While we do not aim to capture actual causality of information transfer from one media to another in the strictest sense, we do identify the conditions under which presence of information in one medium is associated with the presence of that information in other media at a later date. We expect inter media elasticity for pairs of media to be asymmetric, since entity characteristics that stimulate social media are different from those that drive commercial media as discussed above, and thus the same information in any medium may have different effects on other media.

IMR elasticity may be conceptually considered to be similar to cross-elasticity, which is common in economics and marketing. While cross elasticities are typically applied to prices or sales volumes, IMR elasticity captures how responsive various media are to changes in one a focal medium. This variable therefore parsimoniously captures both the direction as well as magnitude of media response. For instance, a positive and significant social media to commercial media elasticity for a product implies that an increase in social media mentions for that product shall lead to a corresponding increase in commercial media. However, unlike cross elasticities, IMR elasticities need to be analyzed in conjunction with the inter-media durations, which are discussed next.

The importance of media reactivity duration can be best demonstrated by two examples. The first is the experience of American Express, as described by Christopher Frank, VP of Global Marketplace Insights, in a panel discussion at the Marketing Science Institute (MSI) and Wharton Interactive Media Initiative "The Emergence and Impact of User-Generated Content" conference, held at Wharton, University of Pennsylvania, December 10-12, 2009. After that company started tightening credit lines midst the Great Recession, they faced a tremendous negative backlash in both the social and commercial media. Even as they prepared to respond to the negative publicity, to their surprise, they noticed that the backlash was relatively short-lived in both media (though it lingered for a bit longer in the commercial media as compared to social media). While American Express was able to emerge relatively unscathed from this experience, they were unable to predict how the two media would interact and react to their strategy of lowering credit, and were caught off guard by the duration of media response as well. The second example describes the launch experience of the movie "District 9", which was heavily marketed using social media networks. As an LA Times article notes, the studios miscalculated their media support timing by starting the viral buzz campaign too soon (10 months before launch), which led to at least some fans becoming disenchanted with the sporadic nature of updates on the movie. Clearly, the marketers of this movie miscalculated the duration for which social media would remain simulated in response to their promotion efforts. Thus, the length of time that media transmit information is as important as which media transmit that information and the duration is therefore an important dimension of the IMR. The duration captures how long the reacting media remains stimulated following an information spurt in the leading media. The duration is especially important in the context of UGC and social media as noted above, since some seemingly innocuous campaigns can take a life of their own, while other campaigns fizzle out quickly.

We now discuss the modeling techniques that may be used to capture the IMR characteristics. The method discussed below uses established econometric techniques to capture the IMR elasticity as well as duration.

Measuring IMR

The choice of our modeling approach to measuring IMR characteristics is driven by the following considerations. We need a model that (1) treats the various media as endogenous, so that they are influenced by both past values of same media as well as other media, as implied in media virality, (2) captures the complex feedback loops that may exist among media, (3) is capable of inferring the temporal causality, and (4) can estimate the duration of reactivity effects. Furthermore, the model chosen should be flexible enough to enable inclusion of any exogenous variables (in addition to the endogenous media variables) and would need to be estimated separately for each product / brand. All the above conditions can be parsimoniously captured using a distributed lag Vector Autoregressive (VAR) Model (Dekimpe and Hanssens 1995, Nijs et al 2007) as shown below.

$$\begin{bmatrix} M1_{it} \\ M2_{it} \\ M3_{it} \end{bmatrix} = \begin{bmatrix} \gamma_{M1,i} \\ \gamma_{M2,i} \\ \gamma_{M3,i} \end{bmatrix} + \sum_{j=1}^{J} \begin{bmatrix} \pi_{11}^{ij} \pi_{12}^{ij} \pi_{13}^{ij} \\ \pi_{21}^{ij} \pi_{22}^{ij} \pi_{23}^{ij} \\ \pi_{31}^{ij} \pi_{32}^{ij} \pi_{33}^{ij} \end{bmatrix} \begin{bmatrix} M1_{it-j} \\ M2_{it-j} \\ M3_{it-j} \end{bmatrix} + \delta_i [E_i] + \begin{bmatrix} u_{M1,it} \\ u_{M2,it} \\ u_{M3,it} \end{bmatrix} \quad (1)$$

This representation captures cross media-response effects for entity *i*, where the constant term is represented by γ, M represent the various media, with a 3 media model being shown for exposition, *t* presents time. In the system of equations (1), $[u_{M1}, u_{M2}, u_{M3}]' \sim N(0, \Sigma_u)$, and the order of the system, J (denotes the number of lags of endogenous variables) is determined by minimizing Schwartz' Bayes Information Criterion (SBIC). In the coefficient matrix, the diagonals (π^{ij}_{kk}) capture the own media momentum, or how a medium reacts to past information in the same medium. The other elements of the coefficient matrix capture the lagged relationships between various media. Contemporaneous effects are captured in the error terms. The matrix [E] represents exogenous variables, such as seasonality, trend, environmental factors, etc. While the VAR methodology has been well-accepted in the field of marketing and has been extensively used especially for modeling market response in evolving markets, with numerous recent applications (for instance, Nijs et al 2007, Srinivasan and

Hanssens 2009 and Joshi and Hanssens 2010), Web Appendix A discusses the estimation issues for a VAR system of models in greater detail.

Estimating this system of equations enables us to generate impulse response functions (IRFs), which trace the over-time impact of a unit shock to any endogenous variable on the other endogenous variables. IRFs are econometric "what-if" simulations, predicting the reaction of an endogenous variable (say, blogging activity) to a 'shock' to another endogenous variable (say, news coverage). If the VAR model is estimated in logs, then the IRFs are elasticities (Nijs et al 2001). The IRF values in this case would indicate how a % change in volume for one medium (e.g. news) would lead to a reaction (% increase or decrease) in another medium (e.g. blogs). Finally, the duration of the shock indicates the number of periods that the shock continues to have an impact on the response variable, and therefore captures the reactivity duration. The IRFs therefore parsimoniously capture the combined effects of the level (or volume) of media support, the interaction and interdependence effects as well as the duration of those effects across media. This method also captures causality by identifying the direction of information flow (through analysis of how each media responds to a shock to every other media, transmitted through the entire system) and the duration of impact (through length of significance of IRF lags) [5].

Following Dekimpe and Hanssens (1999), we use generalized IRFs (or simultaneous shocking) to ensure that the ordering of variables in the system does not affect the results and also to account for contemporaneous or same-period effects. The duration of the shock (maximum lag k) can be determined as the last period in which the IRF value has a $|t|$ statistic greater than 1 (Dekimpe et al. 1999, Nijs et al. 2001). While the VAR provides consistent and efficient estimates and accounts for the endogeneity among variables, we need to ensure that the residuals are white noise by testing them for serial correlation using the Ljung-Box Q Test (Ljung and Box 1978), which is the commonly

[5] In our empirical application we use the term "causality" in the Granger (temporal) sense (Granger 1969, Hanssens et al. 2001). In essence, Granger causality implies that knowing the history of a variable X helps explain a variable Y, over and above Y's own history. This 'temporal causality' is the closest proxy for causality that can be gained from studying the time series of the variables (i.e., in the absence of manipulating causality in controlled experiments).

accepted method for testing the randomness in residuals. We now demonstrate how our methodology may be applied to capture the IMR for a large set of entities.

Setting for Empirical Application

The data for our empirical application must be such that the various aspects of IMR discussed above are clearly demonstrated. Thus, we need an industry or product category with significant presence in both social and commercial media, data availability over time and across a (relatively) large number of entities (products), to demonstrate heterogeneity of IMR. In keeping with these requirements, we demonstrate the concept of IMR as well as the suitability of our chosen model using data from the motion picture industry, using movie stars as the 'entities'[6]. Stars are generally regarded as the backbone of the industry (Elberse 2007) and receive tremendous social and commercial media coverage. While movie stars themselves receive no direct advertising support, they are widely mentioned in other commercial media such as TV, news and print. Further, stars are heavily discussed in social media. As a recent study (Liu 2006) relating pre-launch movie reviews with performance notes, "Movie stars are the frequent topic of interpersonal communication…" For instance, stars like Tom Cruise and Brad Pitt are blogged and searched-about even when they have no movies forthcoming in the immediate future.

Media

Our primary aim is to understand media virality for products, which is driven by the interaction of social and commercial media. We therefore include blogs and discussion forums (social) and news (commercial) as the main media outlets for product information, based on extant research discussed above[7]. Blogs (or web logs) are the most popular form of user-generated commentary on the internet, with the blog search engine Technorati tracking over a 112 million blogs by 2008. While blogs are typically

[6] In the remainder of the text we therefore use the terms 'entity' and star or actor /actress interchangeably.
[7] There exist several other popular types of social media such as instant messaging (IM), and social networking sites such as Facebook. However, these media are either one-to-one (IM) or carry significant restrictions on public viewing (Facebook). Other social networks such as Twitter were still too small at the time when this study was conducted to be included in the empirical analysis. Clearly, our aim in this research is not to exhaustively analyze media reactions for all media, but to document its existence and drivers.

the opinion of an individual or a group of individuals, discussion forums allow for a dialogue among users, and can thus be important shifters of opinion[8]. Finally, we include news as the main outlet for commercial media. The 'news' medium includes information appearing in news sites and print media. Despite the growing popularity of social media and user-generated content for information dissemination, commercial media such as news still has a reach that far exceeds that of social media, and is therefore an important part of our analysis. Since our focal entities are movie stars, we do not include TV advertising, since there is no star-specific TV advertising[9]. In addition, we include the search trends for the stars in our model. While search trends themselves will primarily be an *indication* of the level of public interest (buzz), we empirically allow for a situation where they may also be a source of information. Search engines often prominently display the most popular search terms, as do several main pages of top portals. Furthermore, top search terms on various search engines are frequently mentioned in the news. For instance, a recent article in the Washington Post (April 30, 2009) discussed the top searched terms on Microsoft's Live and Yahoo! Search engines. Most websites and portals have a list of 'most viewed' links and 'most emailed' articles, which ensure that actions taken by consumers (or "word of mouse"; Gelb and Sundaram 2002) in social media themselves become informative for future consumers. We therefore include Google search trends as an endogenous media outlet in our model, and are, to the best of our knowledge, the first study in the field to do so.

Unit of Analysis

An important issue that needs to be addressed pertains to the unit of analysis. We define each individual message as a data point, irrespective of its content. Thus, similar messages occurring 5 times across different media are considered as being 5 data points. This follows from the established practice in aggregate advertising response literature that advocates the use of advertising expenditures (Little 1979). Findings related to the effects of advertising repetition, which finds that message repetition has a positive

[8] While blogs may also include firm-run 'fertilized' blogs, we are unable to distinguish such content from UGC. However, based on the volume of UGC in our data, we expect that fertilized blogs are a minuscule portion of the total social media we consider.
[9] We do include all types of advertising (including TV advertising) for the star movies in subsequent analysis below.

impact on evaluation of known brands (Campbell and Keller 2003), lend further credence to our approach.

Data

We started by identifying top male and female stars in the movie industry. Past research has used several methods for classifying actors as ‘stars’, such as possession of ‘marquee value’ according to the trade magazine *Variety* (Sawhney and Eliashberg 1996), ranking on the *Hollywood Reporter Star Power Survey (2002)* (Ainslie et al 2005) or the existence of a “StarBond” for an actor on the *Hollywood Stock Exchange* (HSX.com) (Elberse 2007). In order to be inclusive, we use each of these methods to arrive at a list of 75 actors (Table 2). For each of these stars we collect media data from 2004-2008 involving any of the stars in our list. The time period is chosen since data on search and user-generated content are only available from 2004 onwards. For each star, we also obtained information on their professional achievements (awards won, number of movie releases during our data interval, the box office gross of their highest grossing movie), personal milestones (including community service, scandals and other issues), and demographics (age, gender) from IMDb.com and Wikipedia.org.

Insert Table 2 here

Next, media data were collected from various sources. We obtained data on star-related search trends from Google.com. Google is the most popular and widely used search engine on the web, with a market share of about 80% in the worldwide search market (Hitslink.com 2009). Furthermore, past search data has just recently became available to the public through the “Google Trends” tool on Google.com. Google Trends provides a normalized and scaled time series trend of the relative search volume for the keyword entered. Google only provides scaled data to protect privacy and as a consequence, we do not have the actual search volume for any particular star. To ensure that all our actor trends are scaled on the same basis and therefore comparable, we chose to obtain data from Google Trends using the comparative search feature. For this reason, we randomly chose an actor (Mila Kunis) who was not a part of our list of

75 and compared all other actor search trends to her search trend. Mila Kunis was chosen as a 'baseline' for two reasons. First, while she is not on our top actor list, there are still enough searches for her name so that there is a continuous non-zero time series for her name search over the duration of our data period. Secondly, using a very popular name as a baseline may result in an all-zero series if the focal name is much less popular than the baseline. Thus, while any name or term can be used as a baseline, it makes more sense to use a relatively lesser known name so as to ensure that the trends for the 75 actors in our list are significant in comparison to the baseline. Figure 2 provides some examples of our search trends.

Insert Figure 2 here

Using Google News (automated news aggregator service using data from more than 4,500 English-language news sites and historical archives of print media) we obtained time-series data on actor related news (commercial media). Data on user generated content was obtained from omgili.com (discussion forum data) and Bloglines.com[10]. Omgili.com is a specialized search engine that focuses on searching for conversations (as opposed to searching for keywords like other search engines) using a crawler to go through over "...100,000 boards, forums and other discussion based resources." Omgili.com has only recently been used in academic research (Stephen and Galak 2010) and is a valuable source of data since discussion forums are the major form of online communication where users can have a dialogue, and interested parties have a chance to respond to information. Bloglines.com allows for the search of "*All-in-one Blog and online subscriptions, news reader, blog publishing and social sharing tools*" (http://www.bloglines.com/about). While Bloglines provides us with actor related data from several sources; to avoid possible overlap with Google News data we set our search criteria to exclude news. We aggregate blogs and forums data into one variable that captures social media presence, and is hereafter referred to as 'social' media data.

[10] We have also considered alternative popular source of blogs data - Google Blog Search. Data comparison for a few randomly selected actors has shown that while both sources are fairly consistent, data collection procedures were easier to implement from Bloglines.

Tables 3a-3c provide examples of stars with highest and lowest average weekly volumes in each of the three media that we study. While these numbers do not allow us to infer the absolute reach of the corresponding media (e.g., the total number of message recipients) or the media virality, they offer interesting insights from a comparative perspective across stars. First, it is clear that social media lead the way in volume of media mentions for stars, with Jessica Alba being referenced on average 4246 times a week. Indeed, even the least mentioned star in our list, Ed Harris was referenced 85 times per week on social media. In contrast, commercial media (news) generates comparatively lower volumes, with the top stars generating a weekly average of 216 (Brad Pitt) mentions in this medium. The second insight is that while some stars (such as Brad Pitt) generate a high volume of chatter across multiple media, other stars are clearly better at generating interest in one particular media. For instance, George Clooney consistently generates high volumes of commercial media mentions (as exemplified by his top 5 ranking in that media), but is conspicuous with his absence in other media. Similarly, Cameroon Diaz is often referenced in social media (average of 3310 mentions a week), but is nowhere near the top in commercial media. The across-actor variation and within actor relative time-homogeneity of media volumes form an ideal preamble to the discussion of results involving media virality for actors, which are proposed to have similar properties.

Insert Tables 3a-3c here

Results

Following equation (1) we construct a VAR model containing 3 endogenous variables – Actor Commercial, Actor Social and Actor Search, using data collected from these media as described above.[11] To enable analysis at the most disaggregated level possible, we estimate a separate VAR model for each actor, resulting in 75 models estimated. In addition to seasonality and trend, as exogenous variables (E) we have also included some significant events such as professional

[11] Results of unit root tests (Augmented Dickey Fuller, ADF) indicate the absence of a unit root after adjusting for trend and seasonality which suggest that the endogenous variables can enter the model in levels. Please refer to Web Appendix A for details.

achievements (award won/nominated, movie released) and personal milestones (community service, scandals and other issues) which traditionally stimulate media activity. An unrestricted model yields a total of 9 sets of IRFs for every actor. From a mathematical perspective, each IRF represents the elasticity of an endogenous variable (say commercial) to a shock to another such variable (say social). The practical interpretation in this case would be that the IRFs are estimates of inter-media elasticities for the specific pairs of media and capture the inherent media virality. For instance, the IRF of impact of Actor Commercial on Actor Social represents the elasticity of social media to information in commercial media. We should note that the IRF analysis takes into consideration all endogenous variables reacting to the initial shock in one medium and, in turn, possibly driving other media. Hence, the obtained pair elasticities reflect the ultimate result of these complex interactions. While our method provides us with 9 IRFs for each actor, we focus our attention on 4 key elasticities: (1) the impact of commercial media on social media, (2) the impact of social media on commercial media, (3) the impact of commercial media on search (which is an indicator of interest), and (4) the impact of social media on search.

With the help of the IRFs obtained from above we demonstrate the key aspects of IMR that follow from our conceptualization. Specifically, we show that IMR elasticities – (1) are asymmetric across media pairs for same entity, depending on leading and reacting media (2) are heterogeneous across entities (actors) for the same media combination, and (3) have varying duration of reactivity for media pairs across entities.

Asymmetry across Media for Same Entity

Figures 3a and 3b depict IRFs for 2 randomly selected stars – Hugh Jackman and Jack Nicholson. Both sets of IRFs show the asymmetry in reactivity among media pairs. The response of commercial media to information originating in social media is significantly different from the response of social to commercial media for both stars. The asymmetry in media responses is further highlighted by the response of search to both social and commercial media. We argue that this asymmetry may result from the specific characteristics of the entity (stars). Some stars are clearly better at transmitting information through social media, while others are much better at transmitting through commercial media. This knowledge is of value to practitioners in the context where information is organically transmitted across media, as is in the user-generated content world.

Insert Figures 3a and 3b here

Heterogeneity across Entities for Same Media

While the above results show the asymmetry in media reactivity for the same entity, the media map in Figure 4 shows how all the entities (stars) in our data differ across the *same* two media dimensions – response of commercial media to social media (Actor Commercial->Actor Social) and social media to commercial media (Actor Social->Actor Commercial). The IRF data on this plot have been mean centered and standardized across all the actors, so that the actor values are relative to the group. The four quadrants can be interpreted as follows. Quadrant 1 represents stars that have the strongest media reactions from both social and commercial media to information coming from the other medium. Thus information placed in either news or blogs is transmitted by the other medium, albeit with varying strengths for different entities. Hillary Swank and Sandra Bullock have the strongest media reactivity (large relative magnitude of elasticity) relative to other actors while stars like Denzel Washington and Jodie Foster who reside in the 3rd quadrant lag their peers in spreading information across both media. We want to reiterate that being in quadrant 3 does not imply that media interest for these stars is low, but rather that inter media reactions are relatively weak. Stars in the 2nd quadrant (such as Jessica Alba and Brendan Fraser) have strong Commercial-> Social elasticity, but weak Social-> Commercial elasticity. Finally, actors in quadrant 4 (for e.g., Laurence Fishburne) have strong relative Social-> Commercial elasticity, but not vice versa. As noted above, this product snapshot information demonstrates how IMR may vary for the same media combination across entities, and can be useful when managing the communication mix and deciding which media to stimulate for a particular entity at a given point in time.

Insert Figure 4 here

Duration of Effects

In Tables 4a and 4b we show the media interaction durations (in weeks) for select stars in our dataset for the same two pairs of media as above (Commercial and Social). These durations therefore capture the time that a medium (say social) remains stimulated following a spurt of information in another medium (commercial). For each media elasticity pairs, we show the stars with the top 5 and bottom 5 durations. Thus, Morgan Freeman, with a 12-week duration for both media combinations, has the highest longevity of media reactivity, while Robin Williams and Jack Nicholson (1 each) have the shortest durations for commercial to social and social to commercial respectively. Similar to the inter-media elasticities, the durations too are asymmetric across media pairs. For instance for Angelina Jolie the response of commercial media to social media lasts for 12 weeks, but social media only remains stimulated for 3 weeks to a spurt of information in commercial media. In contrast, for Bruce Willis, the social-commercial interaction lasts only for about 3 weeks, while the commercial-social interaction lasts 10 weeks. These durations coupled with the media elasticity for each entity provides managers with easy to understand and use metrics to manage the communication mix.

Insert Tables 4a and 4b here

The discussion above demonstrates the key aspects of IMR, and also highlights the uniqueness of this concept in describing how social and commercial media interact. Through this analysis we establish that media reactivity characteristics are asymmetric for media pairs within entities, heterogeneous across entities, and are of varying durations. These characteristics of IMR also reiterate the importance of this concept in studying communication mix when information is expected to be transferred across media. In general, IMR characteristics can aid managers by predicting media that are strong carriers for specific entities and identifying the duration of media transfer.

The above analysis demonstrates that for some actors (entities) media virality is predominantly driven by information originating in social media, while for others, it is for information in commercial media. What drives this difference? We now explore this aspect of media virality by relating various personal and professional characteristics of the actors to their IMR elasticity and durations.

Drivers of IMR Elasticity and Duration

To understand how stars' personal and professional characteristics drive IMR elasticities and durations derived above, we model the IMR elasticities and durations as functions of star specific characteristics obtained above, such as AGE, GENDER, PERSONAL AWARD, RELEASE, and MAXUSGROSS, besides the static media volume variables (AVGNEWS, AVGBLOGS, AVGTRENDS). These variables capture three distinct aspects of an actor's personality – demographics, professional and personal highlights, and presence in media (static volume), and are described in detail in Table 5.

Insert Table 5 here

Since the IMR elasticities are not data, but the outcomes of model (1) that are estimated with error, we use Simulated Maximum Likelihood (SMLE) for our estimation (details are provided in Appendix B). The estimating equation can be parsimoniously represented as:

$$\textbf{IMR Characteristic}_{\mathbf{s,p}} = \boldsymbol{\delta_0} + \boldsymbol{\alpha} * [\textbf{Demographic Variables}_{\mathbf{s}}] + \boldsymbol{\beta} * [\textbf{Highlights}_{\mathbf{s}}] + \boldsymbol{\gamma} * [\textbf{Static Media}_{\mathbf{s}}] + \varepsilon_{s,p} \qquad (2)$$

where s = 1 to 75 represents the 75 stars in our database; p = 1 to 8 represents the 4 IMR elasticities and 4 durations.

Results are displayed in Table 6, and reveal that distinct star specific characteristics can help explain the heterogeneity in social and commercial media response for the actors. Specifically, we find that stars personal and professional activities drive their IMR elasticities and durations. We find that actors with high incidences of high-profile activities not related to movies (PERSONAL) have a higher social to commercial media elasticity. In contrast, actors who are critically well received (AWARD) have a higher commercial to social media elasticity. This implies that commercial media is more sensitive to information in social media for actors involved in a high profile personal activities such as scandals, while social media is sensitive to information from commercial media for those with high profile professional activities, such as awards. Regressions for the duration of IMR elasticities also reveal the importance of awards in prolonging media virality (increasing reactivity duration). We find that either professional success (AWARD) or other high profile activities (PERSONAL) serve to extend the duration of social to commercial media virality.

We have also tested the robustness of our findings to the inclusion of valence information. As noted above, research has consistently reported that valence either has no effect or a very effect on the impact of commercial and social media on performance. While obtaining the valence of each message in our dataset was impractical, we collect data on key events in each actor's life using Wikipedia.org. Examples of key events include scandals, community service, awards, accidents / mishaps, marriage / divorce, etc. Next, the each event was rated as positive or negative to arrive at valence information for each of the events. This valence information was included as an explanatory variable in equation (2). Results indicate that valence is an insignificant driver of IMR elasticity and duration in all cases.

While this analysis is mainly exploratory, it demonstrates that actor characteristics play a part in driving both inter media elasticities as well as durations. It is important to note that by conducting our analysis over 4 years of data for each actor, we are aggregating over several movies, personal and professional events. These results thus represent the reactions of media to entity specific information, over and above the media reactions on account of a single event or social contagion. The preliminary findings that variables such as AWARD and PERSONAL affect media virality lends credence to our conceptualization that entity specific characteristics drive the transfer of information from media to media.

Insert Table 6 here

Managerial Implications: Impact of Actor IMR on Movie Performance

We next turn to demonstrating the managerial importance of media virality as captured by IMR. As we discussed in the previous section, the media virality of an actor describes how on average one media react to information about that actor emerged in another media. There is a general consensus in the entertainment industry, with support in the academic literature (for instance, Ainslie et al 2005, Elberse 2007), that star participation in a movie has a strong connection with the movie's box office performance. Insofar as a star's media characteristics (beyond other factors) impact movie performance, we expect the star's media virality as captured by the IMR characteristics to impact movie related performance metrics, specifically, the opening weekend box office revenues.

To test if IMR elasticities impact movie performance for each star, we re-estimated each star's IMR using data for the 54 weeks *leading* up to the launch of a new movie. This method enables us to provide managers with two important pieces of information – (1) which media elasticities are significant for a particular star just before movie launch, and (2) what is the impact of these star-specific media elasticities on movie performance. In addition to the IMR variables, we include variables that past research (such as Elberse and Eliashberg 2003; Basuroy et al 2003; Liu 2006; Joshi and Hanssens 2009) has shown to be predictive of movie box office (BO) performance. Thus, for all the movies containing the stars on our list (star movies henceforth), released between 2004-2008, we obtained data on variables such as number of screens at launch, production budget, MPAA rating, genre(s), source (based on book, sequel, remake, etc), opening weekend box office revenue (in the US), etc from the-numbers.com. Movie critic and user ratings were obtained from Yahoo.com. To demonstrate the differential impact of star movies from movies without stars and to avoid possible sampling bias, we collected the same data for another set of randomly selected non-star movies from the same time period. This resulted in a total set of 497 movies for analysis. Next, we purchased advertising data for each movie, from TNS Media Intelligence. These data include the total dollar value of media expenditure across 11 different media over the entire lifecycle of the movie in question. Total advertising has often been used to model movie performance in place of pre-launch advertising and research has indicated that it is a good proxy for pre-launch advertising (Joshi and Hanssens 2009). Tables 7a and 7b provide descriptive statistics for our data. As is clear from Table 7b, movies of a wide range of genres and MPAA ratings are well represented in our sample.

Insert Tables 7a and 7b here

Our forecasting variables fall into three categories. First, we include movie-related variables that extant research has shown to impact movie performance, as discussed above. Second, we include media volume variables which include the peak level of search for the *movie* just prior to launch and the 'baseline' level of social and commercial media activity for the *star* present in that movie (for star movies). The 'baseline' media volume for stars is estimated as the average volume for each media (news, blogs, search and forums) for the first 4 weeks (of the 54 week data interval) for each star. Advertising expenditures for a movie typically begin only a few weeks prior to launch, and by using the first 4 weeks (from our 54 week time series) we ensure that the entity-specific volume is accurately captured, even for movies that begin advertising spends well in advance of launch[12]. The third and final category of variables used (for star movies) is the four focal IMR elasticities. Our choice of opening weekend box office revenue as the dependent variable is driven by findings that media activities generally affect opening weekend revenues alone, with other factors such as word-of-mouth more important as subsequent drivers of revenue (Joshi and Hanssens 2009).

[12] We also considered other intervals from 4 weeks to 20 weeks for the entity media volume, and our results are robust to differing windows. Furthermore, as we argued above mentions of the movie and actor-movie combinations are extremely rare in both social and commercial media until shortly before a theatrical release, thereby lending support to our use of this timeline.

Forecasting Model

Our model, where the subscript m denotes movies, can be parsimoniously represented as below.

$$\mathbf{Performance_m} = \boldsymbol{\delta_0} + \boldsymbol{\alpha}^* \, [\mathbf{Traditional\ Variables_m}] + \boldsymbol{\beta} * [\mathbf{Volume_m}] + \boldsymbol{\gamma}^*[\mathbf{IMR_m}] + \varepsilon_m \qquad (3)$$

As with the analysis on drivers of IMR characteristics above, we use SMLE for estimating this model (Appendix B). The explanatory variables in equation (3) correspond to the three categories discussed above. We estimate separate models for star movies and non-star movies, with the later being used to establish the robustness of our findings vis-à-vis past research. All variables (except for the dummy variables and IMR variables) are taken in logs, so that coefficients have easy interpretation as elasticities. A detailed description of all the traditional variables in (3) is provided in Table 8.

Insert Table 8 here

Impact of IMR on Performance

In estimating equation (3), we had to check and control for possible endogeneity among variables. Past research in the motion picture industry (Elberse and Eliashberg 2003, Krider et al 2005) has noted that *screens* are often endogenous with box office revenue (albeit, when revenue and screens are modeled week-by-week)[13]. Therefore, we follow Elberse and Eliashberg (2003) and use time-invariant variables from our equation (3) as our instrumental variables to conduct the Hausman test to check for endogeneity of our variables. Results indicate the absence of endogeneity among the variables in our model. To rule out multicolinearity we estimated Variance Inflation Factors (VIF) for our model. The VIF obtained was 2, much less than the critical value of 10 typically used to signify the presence of multicolinearity. Hence, by ruling out both endogeneity and multicolinearity, we ensure that our model performs adequately.

The results from our model for non-star movies indicate that model fit as well as coefficient magnitude and direction are in line with expectations from past research[14], and as a consequence, we only discuss results from the model for star-movies. As shown in Table 9, variables such as Screens,

[13] Other variables such as advertising and production budget, while intuitively appealing as possibly endogenous, have consistently been shown to be exogenous (see Elberse and Eliashberg 2003, Basuroy et al 2003).

[14] Detailed results available upon request.

Production Budget, Season and Genre are movie related variables that impact opening weekend box office revenue, with signs and magnitudes that are comparable to past studies. In addition, the diagnostic statistics (R^2) indicate suitable model fit.

Among the media volume variables, the peak level search for a movie prior to launch, the variable that captures pre-release interest, has a positive impact on opening weekend revenues, but search for the actor has no impact on performance. Further, the volume of baseline social media activity associated with the star in the movie has a positive impact on opening weekend revenues. Monitoring search and blogging activities for the movie and movie star respectively, prior to the theatrical release would allow studios to make more accurate predictions regarding box office performance, which is in accordance with recent findings (Karniouchina 2011).

Two of the four IMR elasticities are significant in predicting performance for our set of movies. The presence of stars for whom news in commercial media leads to a spurt of activity in social media, or those for whom information in social media leads to greater search, are associated with higher opening weekend revenues.

Insert Table 9 here

Our findings from the performance forecasting model can be summarized as follows, (1) overall media volume (social and commercial) as well as IMR elasticities of the actor help to predict movie performance, (2) search trends for movies have an impact on movie performance, (3) IMR elasticities have an impact performance *over and above* the impact of traditional and other media variables. More importantly, we demonstrate the managerial importance of media virality by relating IMR elasticities to financial outcome variables.

Conclusions and Future Research

The omnipresence of social media, user-generated content and subsequent consumer media empowerment has meant that managers and academics alike need to alter the well known theories about how to successfully and efficiently communicate product information to consumers. Messages are now regularly moving from one medium to another, often without the control of those who originate the message. Under the presence of media virality, it is necessary that we know how a message will be transmitted through various media, and for how long various media will carry that message. We demonstrate how media virality can be parsimoniously captured using inter-media reactivity, which has two dimensions – elasticity and duration. We further demonstrate that IMR characteristics provide forecasting power for performance predictions.

This research makes several contributions, both conceptual and substantive. This is the first study to present a uniting generalized framework that integrates social (specifically UGC) and commercial media, and proposes a methodology to estimate entity level media reactions. We propose using IMR to capture media virality, and demonstrate that IMR is asymmetric across media, heterogeneous across products and entities, and lasts for varying durations for media and entities. Through our empirical application in the motion picture industry, we calculate IMR for a large number of entities (75 actors).
We demonstrate that actor characteristics play a part in driving both inter media elasticities as well as durations. Further, we show that the IMR characteristics aid in performance forecasting. We demonstrate that IMR variables show predictive power after controlling for other variables traditionally used in the past research on the motion picture industry forecasting.

The linkage of IMR parameters to performance also generates several substantive results that are managerially important. Managers should note the importance of information disseminated by social media, while being cognizant of the differing impact of search for movie and search for actor. We find that movie search is positively related to performance, while actor search is unrelated. Most importantly, practitioners should note that each media pair has a different impact on performance, highlighting the need to revisit the empirical generalizations for advertising elasticities in light of media virality. Overall, our methodology provides managers with the knowledge of which entities are intrinsically better at transmitting through what types of media over a given period of time.

Limitations in our work point to avenues for future research. Identifying the drivers of an entity's media virality in greater detail than done here becomes an important next step in understanding media dynamics. While our research methodology allows us to analyze a rich dataset, we are unable to quantify the *valence* of this information, focusing instead on the volume of information. Future research on IMR should include the valence of information to address its impact on the dynamic and perhaps integrate some elements of micro level analysis, using, for example Event Structure Analysis techniques (Heise, 1989). Another shortcoming specific to our dataset is that there is no direct measure of performance for our chosen entity. We would clearly expect a stronger linkage between IMR and performance if measured for the same entity, something that can be addressed in the future. Finally, we only test our concept on a single industry (although we have a large number of entities) and the application of our framework and modeling technique to other industries and product categories can confirm the robustness of both. Future research could also study whether pulsing (e.g., Feinberg 1992, Bronnenberg 1998) is still the optimal strategy for advertising spending under media reactivity or calculate the long run effects of multi-media information dissemination on sales. The still nascent field of social and commercial media interactions thus promises to be an important avenue for research in the coming years.

References

Ainslie, Andrew, Xavier Dreze and Fred Zufryden (2005), "Modeling Movie Life Cycles and Market Shares," *Marketing Science*, 24, 3, 508-517.

Basuroy, Suman, Subimal Chatterjee and S. Abraham Ravid (2003), "How Critical are Critical Reviews? The Box Office Effects of Film Critics, Star Power and Budgets," *Journal of Marketing*, 67, 103-117.

Berger, Jonah and Katy Milkman (2010), "Social Transmission, Emotion, and the Virality of Online Content," Working Paper, Wharton School of Business.

Berger, Jonah and Eric Schwartz (2010), "What Products Get Talked About? How Cues, Public Visibility, and Interest Shape Immediate and Ongoing Word-of-Mouth," Working Paper, Wharton School of Business.

Bowman, Douglas. And Das Narayandas (2001) "Managing customer initiated contacts with manufacturers: The impact on share of category requirements and word-of-mouth behavior," *Journal of Marketing Research 38*, 3, 281–297.

Bronnenberg, Bart J. (1998), "Advertising Frequency Decisions in a Discrete Markov Process Under a Budget Constraint," *Journal of Marketing Research*, 35, 399-406.

Campbell, Margaret C., and Kevin Lane Keller (2003), "Brand Familiarity and Advertising Repetition Effects," *Journal of Consumer Research*, 30, 292-304.

Chevalier, Judith A. and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43, 3, 9.

DeBruyn, A and Gary Lilien (2004), "A Multi-Stage Model of Word of Mouth through Viral Marketing," *International Journal of Research in Marketing*, 25, 3, 143-225

Dekimpe, Marnik G., and Dominique M. Hanssens (1995), "Empirical Generalizations about Market Evolution and Stationarity," *Marketing Science*, 14(3), 1-21.

------------------ (1999), "Sustained Spending and Persistent Response: A New Look at Long Term Marketing Profitability," *Journal of Marketing Research*, 23, 397-412.

Dekimpe, Marnik. G. and Dominique M. Hanssens (2004), "Persistence Modeling for Assessing Marketing Strategy Performance," *in Cool Tools in Marketing Strategy Research, Lehmann, D. and Moorman, C. (Eds.), Marketing Science Institute*

Dickey, D.A. and W.A. Fuller (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, p. 427–431.

Dobele, Angela, David Toleman and Michael Beverland (2005), "Controlled Infection! Spreading the brand message through viral marketing," *Business Horizons*, 48, 143-149.

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

Elberse, Anita (2007), "The Power of Stars: Do Stars Drive Success in Creative Industries?" *Journal of Marketing*, 71, 4.

Elberse, Anita and Jehoshua Eliashberg (2003), “Demand and Supply Dynamics for Sequentially Released Products in International Markets: The Case of Motion Pictures,” *Marketing Science*, 22, 3, 329-354.

Elaishberg, Jehoshua, Anita Elberse and Mark A.A.M. Leenders (2006), “The Motion Picture Industry: Critical Issues in Practice, Current Research, and New Research Directions,” *Marketing Science*, 25, 6, 638-661.

Feinberg, Fred M. (1992), “Pulsing Policies for Aggregate Advertising Models,” *Marketing Science*, 11, 3, 221-234.

Gelb, Betsy and Suresh Sundaram (2002), “Adapting to “word of mouse”,” *Business Horizons*, 45, 4, 21-25.

Godes, David and Dina Mayzlin (2004), “Using Online Conversations to Study Word-of-Mouth Communication,” *Marketing Science*, 23, 4, 545.

Godes, David and Dina Mayzlin (2009), “Firm-created word-of-mouth communication: evidence from a field test,” *Marketing Science*, 28 (4) 721-739.

Granger, Clive W.J. (1969), “Investigating Causal Relations by Econometric Models and Cross-Spectral Methods,” *Econometrica*, 37, 424-438.

Granovetter, M. S. (1973), “The strength of weak ties,” *American. Journal of Sociology, 78*, 1360–1380

Hanssens, Dominique M., Leonard J. Parsons, and Randall L. Schultz (2001), *Market Response Models: Econometric and Time-Series Research*, 2nd ed., Kluwer Academic Publishers, Boston, MA.

Heise, David R. (1989), “Modeling event structures,” *Journal of Mathematical Sociology*, 14: 139-169.

HitsLink.com (2009), “Search Engine Market Share,” http://marketshare.hitslink.com/report.aspx?qprid=4, accessed August 19, 2009.

Joshi, Amit M. and Dominique M. Hanssens (2009), “Movie Advertising and the Stock Market Valuation of Studios: A Case of “Great Expectations”?,” *Marketing Science,* 28, 2, 239-250.

----------------------(2010), “The Direct and Indirect Effects of Advertising Spending on Firm Value,” *Journal of Marketing*, 74, 20-33.

Kaplan, Andreas M. and Michael Haenlien (2010), “Users of the world, unite! The challenges and opportunities of Social Media,” *Business Horizons*, 53, 59-68.

Karniouchina, Ekaterina V. (2011), “Impact of star and movie buzz on motion picture distribution and box office revenue,” *International Journal of Research in Marketing*, 28, 1, 62-74.

Katz, E., & Lazarsfeld, P. (1955), *Personal Influence,* The Free Press, New York.

Krider, Robert E., Tieshan Li, Yong Liu and Charles B. Weinberg (2005), “The Lead-Lag Puzzle of Demand and Distribution: A Graphical Method Applied to Movies,” *Marketing Science*, 24, 4, 635-645.

Lee, Lung-Fei (1999), “Statistical Inference with Simulated Likelihood Functions,” Econometric Theory, Vol. 15, No. 3 (June), 337-360.

Leskovec, Jure, Lada A. Adamic and Bernardo A. Huberman (2007), "The Dynamics of Viral Marketing," *ACM Transactions on the Web*, 1, 1, 1-39.

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

Little, John D. C. (1979), "Aggregate Advertising Models: The State of the Art," *Operations Management*, 27, 4, 667.

Ljung, G. M. and G. E. P. Box (1978). "On a Measure of a Lack of Fit in Time Series Models," *Biometrika* 65, pp. 297–303.

Lodish, Leonard M, Magid M Abraham, Jeanne Livelsberger, Beth Lubetkin, Bruce Richardson, and Mary Ellen Stevens (1995), "A Summary of Fifty-five In-market Experimental Estimates of the Long-Term Effects of TV Advertising," *Marketing Science*, 14, 3, G133-140.

Lutkepohl, Helmut (1993), *Introduction to Multiple Time Series Analysis*, 2nd Ed., Springer-Verlag, New York.

Mayzlin, Dina (2006), "Promotional chat on the internet," *Marketing Science*, 25 (2), 155-163.

Naik, Prasad A. and Kalyan Raman (2003), "Understanding the Impact of Synergy in Multimedia Communications," *Journal of Marketing Research*, 40, 4, 375-388.

Naik, Prasad A. and Kay Peters (2009), "A Hierarchical Marketing Communications Model of Online and Offline Media Synergies," *Journal of Interactive Marketing*, 23, 288-299.

Nijs, Vincent R., Marnik G. Dekimpe, Jan-Benedict E. M. Steenkamp and Dominique M. Hanssens (2001), "The Category-Demand Effects of Price Promotions," *Marketing Science*, 20, 1, 1-22.

Nijs, Vincent R., Shuba Srinivasan and Koen Pauwels (2007), "Retail-Price Drivers and Retailer Profits," *Marketing Science*, 26, 4, 473-487.

OECD (2007), "Participative web and user-created content: Web 2.0, wikis, and social networking," Paris: Organization for Economic Co-operation and Development.

Sawhney, M. and J. Eliashberg (1996), "A Parsimonious Model for Forecasting Gross Box-Office Revenues of Motion Pictures," *Marketing Science*, 15 (2), 113-131.

Sharp, Byron and Yoram (Jerry) Wind (2009), "Today's Advertising Laws: Will They Survive the Digital Revolution?," *Journal of Advertising Research*, 49, 2, 120-126.

Srinivasan, Shuba and Dominique M. Hanssens (2009), "Marketing and Firm Value: Metrics, Methods, Findings and Future Directions," *Journal of Marketing Research*, 46, 3, 293-312.

Stephen Andrew and Jeff Galak (2010), "The Complementary Roles of Traditional and Social Media in Driving Market Performance," *Working Paper*, INSEAD, 2009/52/MKT.

Stephen, Andrew and Jonah Berger (2010), "Creating Contagious: How Social Networks and Item Characteristics Combine to Drive Persistent Social Epidemics," Working Paper, Wharton School of Business

Trusov, Michael, Randolph E. Bucklin and Koen Pauwels (2009), "Effects of Word-of-Mouth versus Traditional Marketing: Findings from an Internet Social Networking Site," *Journal of Marketing*, 73, 5, 90-102.

Yang, Sha and Greg Allenby (2003), "Modeling interdependent consumer preferences," *Journal of Marketing Research,* 40, 3, 282–294.

Table 1. Research on Impact of Social Media

Article	Product Category	Advertising	Blogs / WOM*	Search	Discussion Forums	Media Interactions	Differential Impact Across Products Studied?	Asymmetry in Media Interactions Studied?
Godes and Mayzlin (2004)	TV shows	N	Y	N	N	N	N	N
Chevalier and Mayzlin (2006)	books	N	Y	N	N	N	N	N
Liu (2006)	movies	N	Y	N	N	N	N	N
Dellarocas, Zhang and Awad (2008)	movies	Y	Y	N	N	N	N	N
Duan, Gu, Whinston (2008)	movies	N	Y	N	N	N	N	N
Stephen and Galak (2009)	microfinance loans	Y	Y	N	Y	Y	N	N
Trusov, Bucklin and Pauwels (2009)	online community	Y	Y	N	N	Y	N	N
Naik and Peters (2009)	automobile	Y	N	N	N	Y	N	Y
Gopinath, Chintagunta and Venkataraman (2010)	movies	Y	Y	N	N	N	N	N
Onishi and Manchanda (2010)	Green tea, movies, cell phones	Y	Y	N	N	Y	Y	N
This research	movies	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**

** WOM has been operationalized as user reviews or message board discussions in the studies mentioned.

Table 2: List of Movie Stars Selected for Analysis.

#	Star Name	#	Star Name
1	Adam Sandler	41	Keira Knightley
2	Al Pacino	42	Kevin Costner
3	Angelina Jolie	43	Kevin Spacey
5	Arnold Schwarzenegger	44	Kirsten Dunst
6	Ben Affleck	45	Laurence Fishburne
7	Ben Stiller	46	Leonardo DiCaprio
8	Brad Pitt	47	Mark Wahlberg
9	Brendan Fraser	48	Matt Damon
10	Bruce Willis	49	Matthew McConaughey
11	Cameron Diaz	51	Meryl Streep
12	Cate Blanchett	52	Morgan Freeman
13	Charlize Theron	53	Nicolas Cage
14	Chris Rock	54	Nicole Kidman
15	Denzel Washington	56	Pierce Brosnan
16	Drew Barrymore	57	Reese Witherspoon
17	Ed Harris	58	Renee Zellweger
18	Eddie Murphy	59	Richard Gere
19	Edward Norton	60	Robert De Niro
20	Ewan McGregor	61	Robin Williams
21	George Clooney	62	Samuel L. Jackson
22	Gwyneth Paltrow	63	Sandra Bullock
23	Halle Berry	64	Scarlett Johansson
25	Heath Ledger	65	Sean Penn
26	Hilary Swank	66	Seth Rogen
27	Hugh Grant	67	Susan Sarandon
28	Hugh Jackman	68	Sylvester Stallone
30	Jennifer Aniston	69	Tim Robbins
31	Jessica Alba	70	Tom Cruise
32	Jim Carrey	71	Tom Hanks
33	Joaquin Phoenix	72	Uma Thurman
34	Jodie Foster	73	Whoopi Goldberg
35	John Travolta	74	Will Ferrell
36	Johnny Depp	75	Will Smith
37	Julia Roberts		
38	Kate Hudson		
39	Kate Winslet		
40	Keanu Reeves		

Table3a: Top and Bottom Stars Referenced in Commercial Media (News)

(average weekly volume preceding the indicated date)

Movie Star	News
top…	
Brad Pitt (09/22/2007)	216
Matt Damon (12/15/2007)	180
Angelina Jolie (06/23/2007)	168
Nicole Kidman (12/08/2007)	155
George Clooney (10/06/2007)	149
...and bottom	
Laurence Fishburne (03/24/2007)	23
Jodie Foster (04/05/2008)	23
Ewan McGregor (01/19/2008)	22
Keanu Reeves (04/12/2008)	22
Tim Robbins (05/10/2008)	20

Table 3b: Top and Bottom Stars Referenced in Social Media (Blogs and Forums)

(average weekly volume preceding the indicated date)

Movie Star	Blogs
top…	
Jessica Alba (06/21/2008)	4246
Johnny Depp (07/04/2008)	4147
Brad Pitt (09/22/2007)	3846
Cameron Diaz (05/10/2008)	3310
Will Smith (12/15/2007)	2811
...and bottom	
Hugh Grant (04/22/2006)	166
Matthew McConaughey (12/23/2006)	160
Mel Gibson (03/18/2006)	130
Brendan Fraser (08/02/2008)	118
Ed Harris (11/11/2006)	85

Table3c: Top and Bottom Stars in Search Trends

(average relative weekly volume preceding the indicated date)

Movie Star	Search
top…	
Jessica Alba (06/21/2008)	17.4
Angelina Jolie (06/23/2007)	16.4
Jennifer Aniston (06/03/2006)	9.2
Brad Pitt (09/22/2007)	8.4
Will Smith (12/15/2007)	5.5
...and bottom	
Pierce Brosnan (03/08/2008)	0.6
Cate Blanchett (10/13/2007)	0.6
Anthony Hopkins (10/27/2007)	0.6
Susan Sarandon (05/10/2008)	0.4
Whoopi Goldberg (09/16/2006)	0.3

Table 4a: Top and Bottom Stars in Duration of Commercial to Social Parameter

Movie Star	Duration
top…	
Morgan Freeman	12
Chris Rock	12
Ewan McGregor	12
Hilary Swank	12
Matthew McConaughey	11
...and bottom	
Drew Barrymore	1
Richard Gere	1
Keanu Reeves	1
Robin Williams	1
Sylvester Stallone	1

Table 4b: Top and Bottom Stars in Duration of Social to Commercial Parameter

Movie Star	Duration
top…	
Matthew McConaughey	12
Morgan Freeman	12
Reese Witherspoon	12
Susan Sarandon	12
Tom Cruise	12
...and bottom	
Al Pacino	1
Brendan Fraser	1
Drew Barrymore	1
Hugh Jackman	1
Jack Nicholson	1

Table 5: Description of Star Characteristics Used to Model Drivers of IMR Elasticities

Variable	Description
AGE	Age, in years as of December 2008
GENDER	Male = 1; Female = 0
PERSONAL	Taking the value of 1 for the week in which there are high-profile personal activities* by the star
AWARD	Takes the value of 1 for the week in which the actor wins a major** award
RELEASE	Number of movies released between 2004-2008
MAXUSGROSS	The maximum revenue ($) ever earned by any of the actor's movies
AVGNEWS	Average count of number of mentions for the star in commercial media
AVGBLOGS	Average count of number of mentions for the star in social media
AVGTRENDS	Average count of number of searches for the star's name

* 'High-profile' activities are defined as those that warrant a mention on the actor's Wikipedia page. Examples of such activities include scandals, community service, accidents / mishaps, marriage / divorce, etc. All professional activities (such as Awards) were excluded from this count.
**For the purpose of this analysis, Golden Globes and the Academy Awards were considered as 'major' awards

Table 6: Drivers of IMR Elasticities and Durations

	Elasticities				Durations			
	Social on Commercial	**Commercial on Social**	**Social on Search**	**Commercial on Search**	**Social on Commercial**	**Commercial on Social**	**Social on Search**	**Commercial on Search**
Const	0.0995**	0.1049*	0.0545*	0.1784**	0.77594	2.9150	0.6228	7.5978**
	(2.1935)	(1.9532)	(1.8477)	(3.5856)	(.3076)	(-1.3260)	(0.2610)	(4.4051)
Age	-0.0003	0.0008	-0.0004	0.0008	-0.018438	-0.0263	0.0190	-0.0088
	(-0.4219)	(0.0817)	(-0.7766)	(-1.6348)	(-.3926)	(-.6436)	(0.4280)	(-.2755)
Award	0.0139	0.0406*	0.005	0.0127	1.707419*	1.2712	1.7335*	2.6024**
	(0.7778)	(1.8737)	(0.433)	(0.6259)	(1.7140)	(1.4641)	(1.8391)	(3.8204)
Personal	0.007*	0.0007	-0.0002	0.0003	0.396716*	-0.1293	0.4205*	-0.0881
	(1.6581)	(0.134)	(-0.105)	(0.0711)	(1.6867)	(.6306)	(1.8892)	(-.5476)
Release	-0.0005	0.0032	-0.0001	-0.0003	0.199453*	0.0641	0.1977*	-0.1069
	(-0.2769)	(1.2998)	(-0.1317)	(-0.1506)	(1.7029)	(0.6278)	(1.7838)	(-1.3347)
Gender	0.0412**	0.0158	0.0067	0.0097	2.026329*	1.6217	-0.2088	0.4784
	(2.0463)	(0.6595)	(0.5001)	(0.3002)	(1.7683)	(1.6237)	(-.1926)	(0.6105)
MaxUSGross	-0.0001	-0.0001	-0.0002	0.0004	-2.78E-09	0.0000	0.0000	0.0000
	(-0.9757)	(-1.3876)	(-0.4779)	(0.4377)	(-.7080)	(0.6823)	(-1.0755)	(0.6778)
AvgSocial	-0.0003*	-0.0003	-0.0003	0.0001	-0.000871	-0.0018	0.0003	0.0001
	(-1.9253)	(-1.6011)	(-0.0253)	(0.5524)	(-.7878)	(-1.8675)	(0.2652)	(0.1090)
AvgNews	-0.0002	0.0002	0.0001	-0.0001	-0.000221	0.0025	-0.0027	-0.0154**
	(-0.1283)	(1.2098)	(1.4797)	(-0.0856)	(-.0205)	(.2652)	(-.2609)	(-2.0829)
AvgTrends	0.0128**	0.009	0.0002	-0.0041	0.837401**	0.7219**	0.4511	0.1017
	(1.9698)	(1.2112)	(0.0583)	(-0.6639)	(2.2699)	(2.2450)	(1.2922)	(.4033)

*p<.10; **p<.05

Table 7a: Description of Data: Actors and Movies

	Minimum	Maximum	Mean	Median
Actors				
News*	0	223.04	57.4	51.75
Social Media (blogs)*	0	5,178.86	462.07	171
Social Media (discussion forums)*	0	429	9.04	3.5
Search Trend Index*	0	15.98	2.31	1.35
Number of Movies	1	9	2.84	2.00
Movies (with stars)				
Advertising Budget ($K)	7.5	49,107.8	20,210.0	20,523.9
Production Budget ($K)	4,000	258,000	53,766	40,000
Opening Weekend Box Office ($K)	11.1	151,116.5	15,876.6	9,390.1
U.S. Gross ($K)	356	891,930.3	135,070.7	76,884.8
Search Index Reached prior to Opening	0.2	160.0	11.2	5.2
Yahoo! Critics Rating§	1.3	3.7	2.6	2.7
Yahoo! User Rating§	1.3	3.7	2.89	3
Movies (without stars)				
Advertising Budget ($K)	1,011	59,526	19,705	18,428
Production Budget ($K)	400	207,000	41,808	30,000
Opening Weekend Box Office ($K)	9.65	135,634.55	15,438.66	10,005.9
U.S. Gross ($K)	381.42	436,721.7	52,521.15	33,302.17
Search Index Reached prior to Opening	0.2	200	9.71	3.74
Yahoo! Critics Rating§	1	4	2.4	2.3
Yahoo! User Rating§	1	3.7	2.79	3

* *– weekly averages for the baseline period*

§ *– The following conversion from letter ratings was used:*
"A" - 4, "A-" - 3.7, "B+" - 3.3, "B" - 3, "B-" - 2.7, "C+" - 2.3, "C" - 2, "C-" - 1.7, "D+" - 1.3, "D" - 1.

Table 7b: Description of Data: Movie Genre, MPAA Rating and Source

	Share of Movies
Genre	
Action	10%
Adventure	10%
Comedy	33%
Documentary	3%
Drama	29%
Horror	6%
Musical	1%
Thriller	9%
Western	1%
MPAA Rating	
PG13	40%
R	37%
PG	18%
G	4%
NR	1%
Source	
Original Screenplay	51%
Based on Book/Short Story	20%
Sequel	8%
Remake	6%
Other	15%

Table 8: Description of Traditional Variables Used in Analysis

Variable	Description	Source
Advertising	Advertising support received by a movie, in $	TNS MI
Screens	Number of screens for movie on opening weekend	The-numbers.com
MPAA	Rating from the Motion Picture Association of America	The-numbers.com / IMDB
Genre	Classification by movie type	The-numbers.com / IMDB
Production Budget	Estimated budget to produce the movie	The-numbers.com
Source	Dummy variables for Sequel, Book-based movie, Remake, Original Script and Other	The-numbers.com / IMDB
Seasonality	Season in which the movie opened, with a dummy variable for the main movie release seasons	IMDB

Table 9: Impact of Traditional and IMR Variables on Movie Performance.

Predictors Category	Predictors	Opening Weekend(log)
Traditional Prediction Variables	Intercept	1.7423** (2.21)
	SEASON:Jan-Mar	-0.3887** (-1.99)
	SEASON:Apr-May	-0.6868** (-3.07)
	SEASON:Aug-Nov	-0.8738** (-4.34)
	SEASON:ThanksGiving-Dec	-0.4677 (-1.58)
	Screens(log)	0.8276** (14.36)
	RATING:PG	0.3421 (1.09)
	RATING:R	-0.077 (-0.29)
	RATING:PG-13	0.3770 (1.24)
	GENRE:Comedy	0.2357 (1.42)
	GENRE:Drama	-0.0681 (-0.35)
	GENRE:Action	0.8416** (3.18)
Volume Variables	ProdBudget(log)	0.0254** (3.40)
	Ads Budget(log)	0.1019** (2.28)
	Peak Movie Search(log)	0.0701** (2.92)
	Actor Commercial (log)	-0.0644 (-0.39)
	Actor Social (log)	0.0979** (2.05)
	Actor Search (log)	-0.0577 (-0.73)
IMR Variables	Social on Commercial	0.4188 (0.80)
	Commercial on Social	1.5927** (2.07)
	Commercial on Search	0.7800 (0.92)
	Social on Search	-2.1432* (-1.89)
	Adjusted R-squared	0.90

*p<.10; **p<.05

Figure 1a: Location of Product in Media Space on Basis of Volume

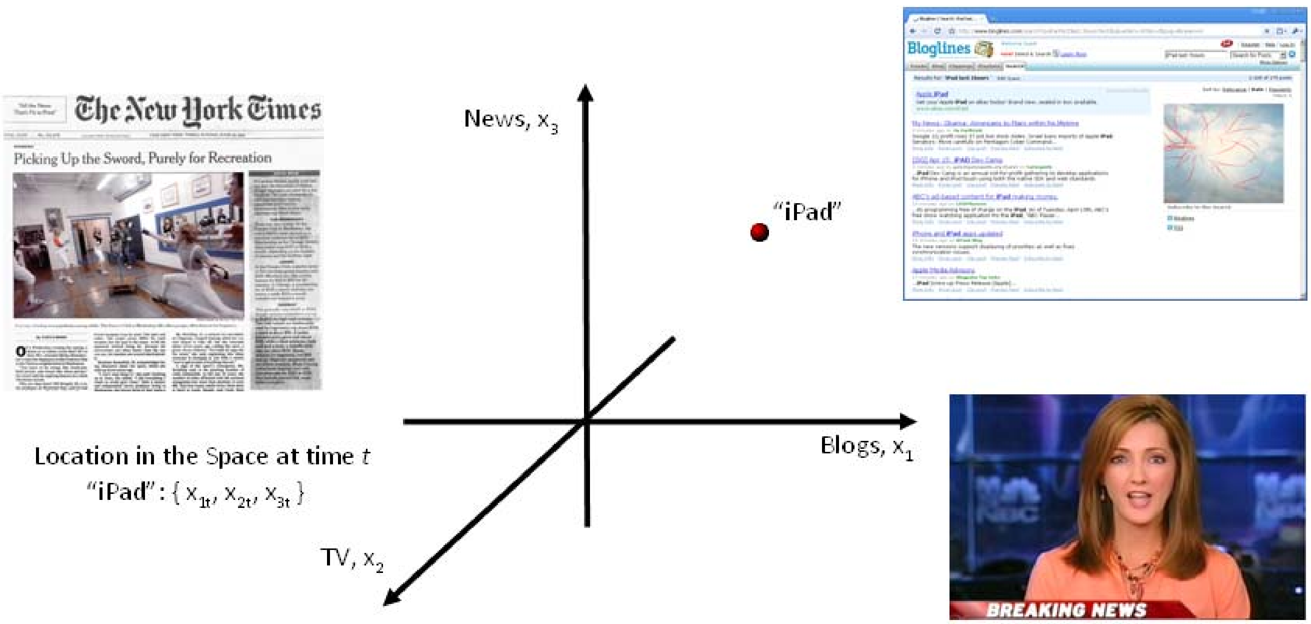


Figure 1b: Change in Product Media Location in Non-Interdependent World in Response to Change in Volume of One Media

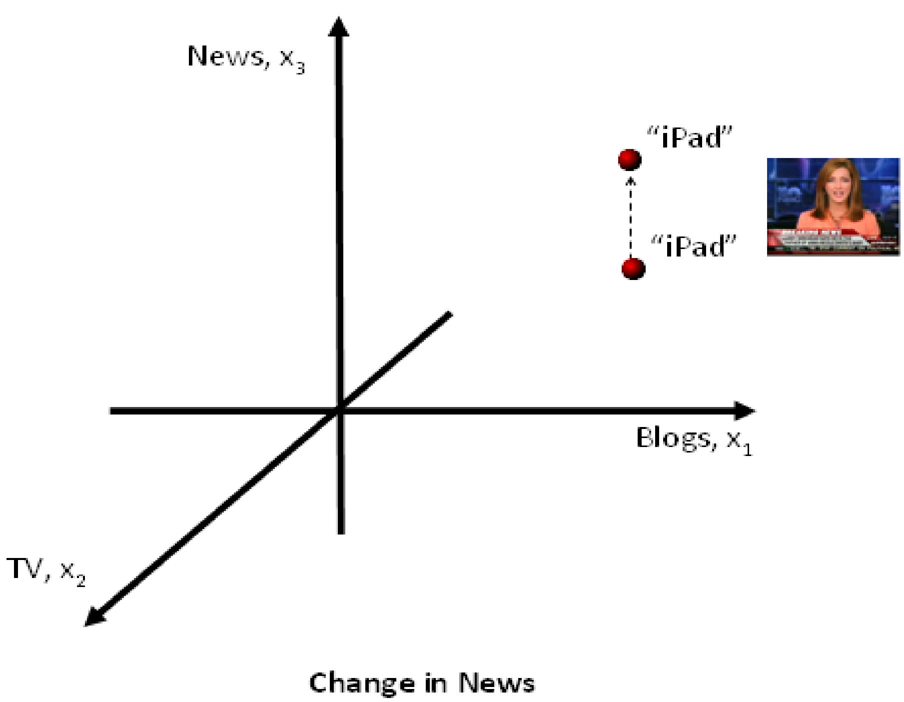
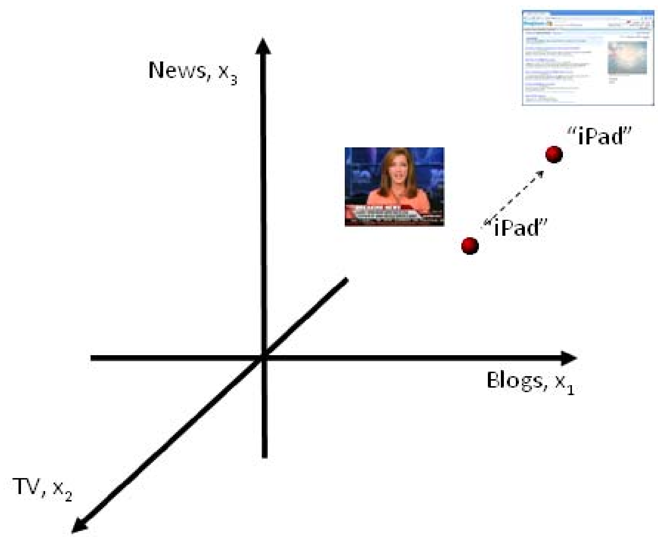


Figure 1c: Change in Media Location in Interdependent World with Media Virality



Change in News Leads to More Blogging

Figure 1d: Same Media Change, Different Impact for Different Products

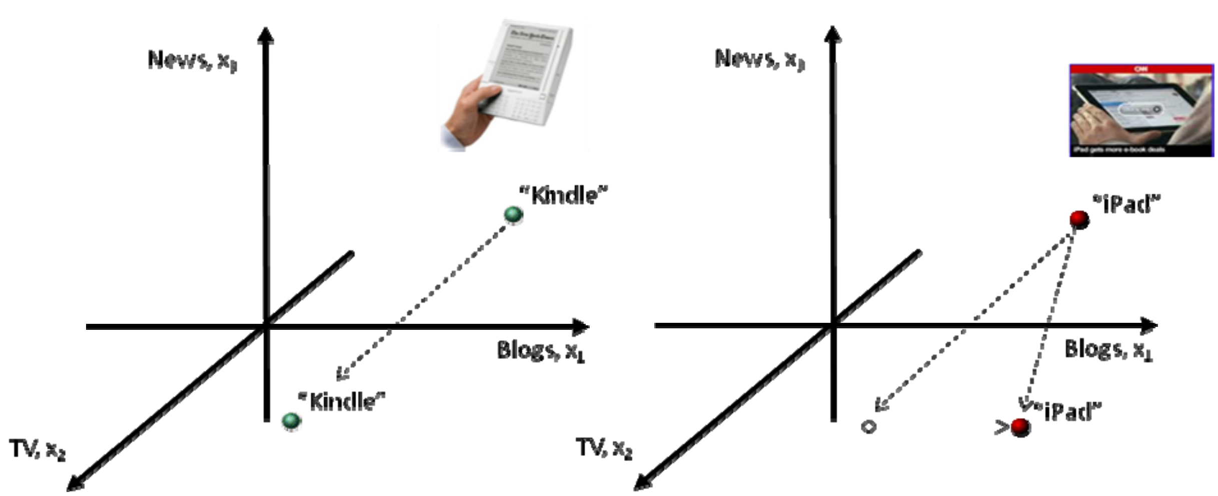
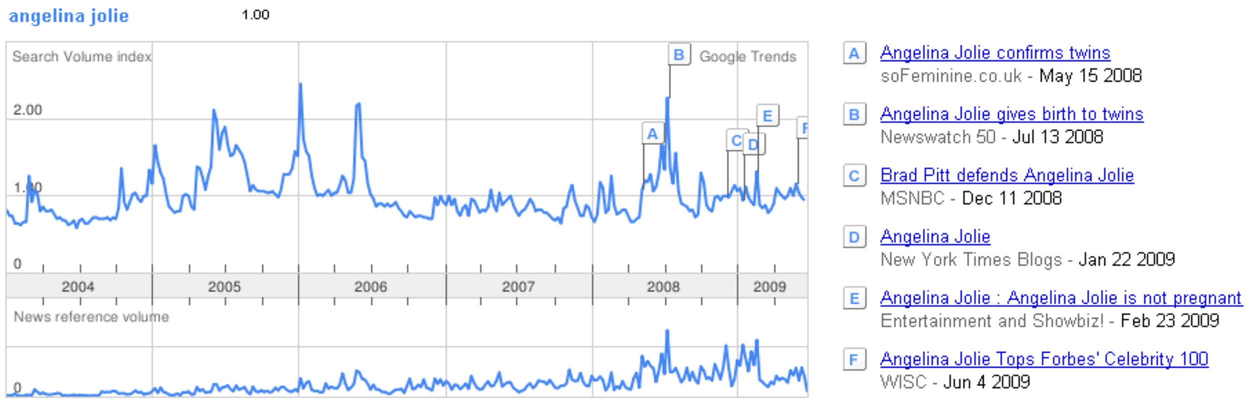


Figure 2: Google Trends – Stars Generate a lot of Non-Movie Related Buzz

(a) Search Volume Index for Angelina Jolie



(b) Search Volume Index for Meryl Streep

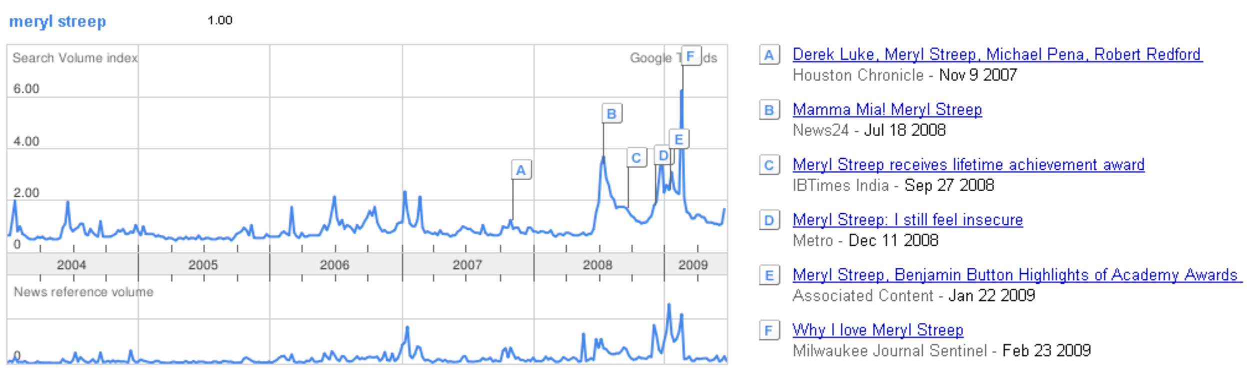
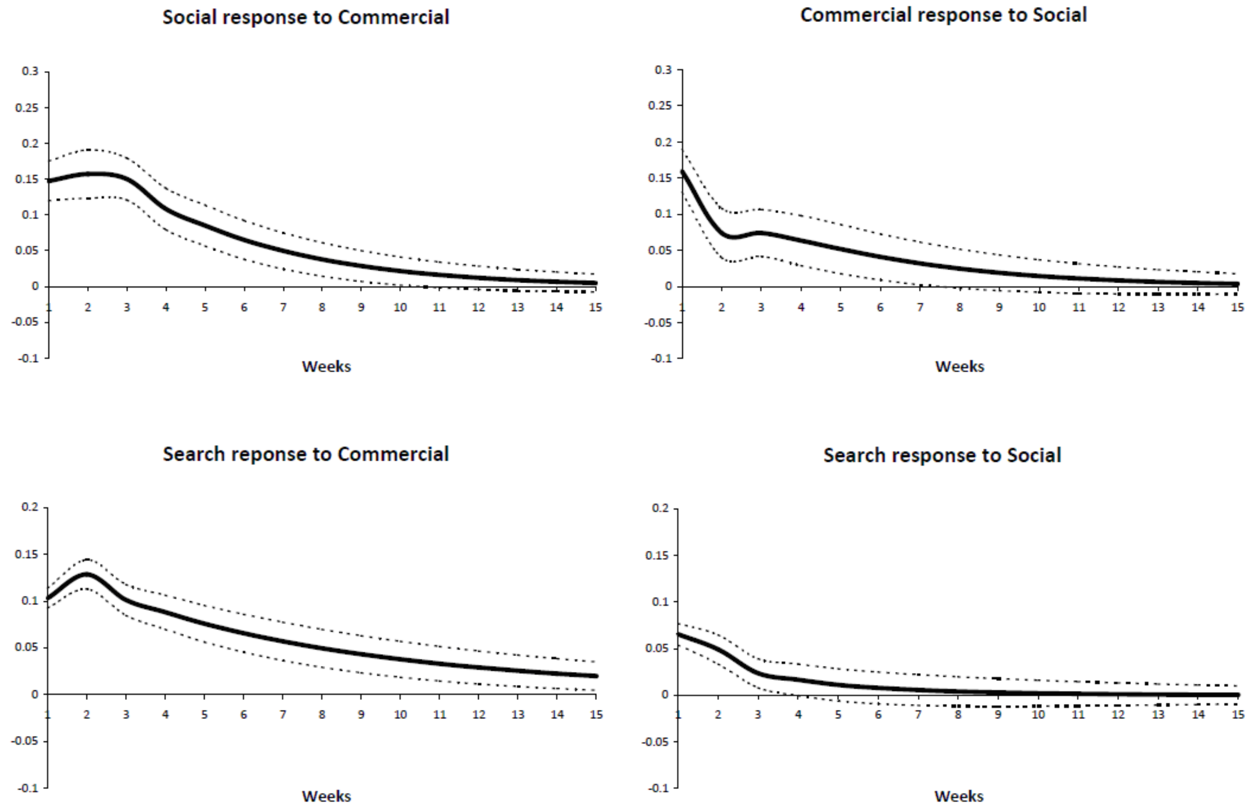
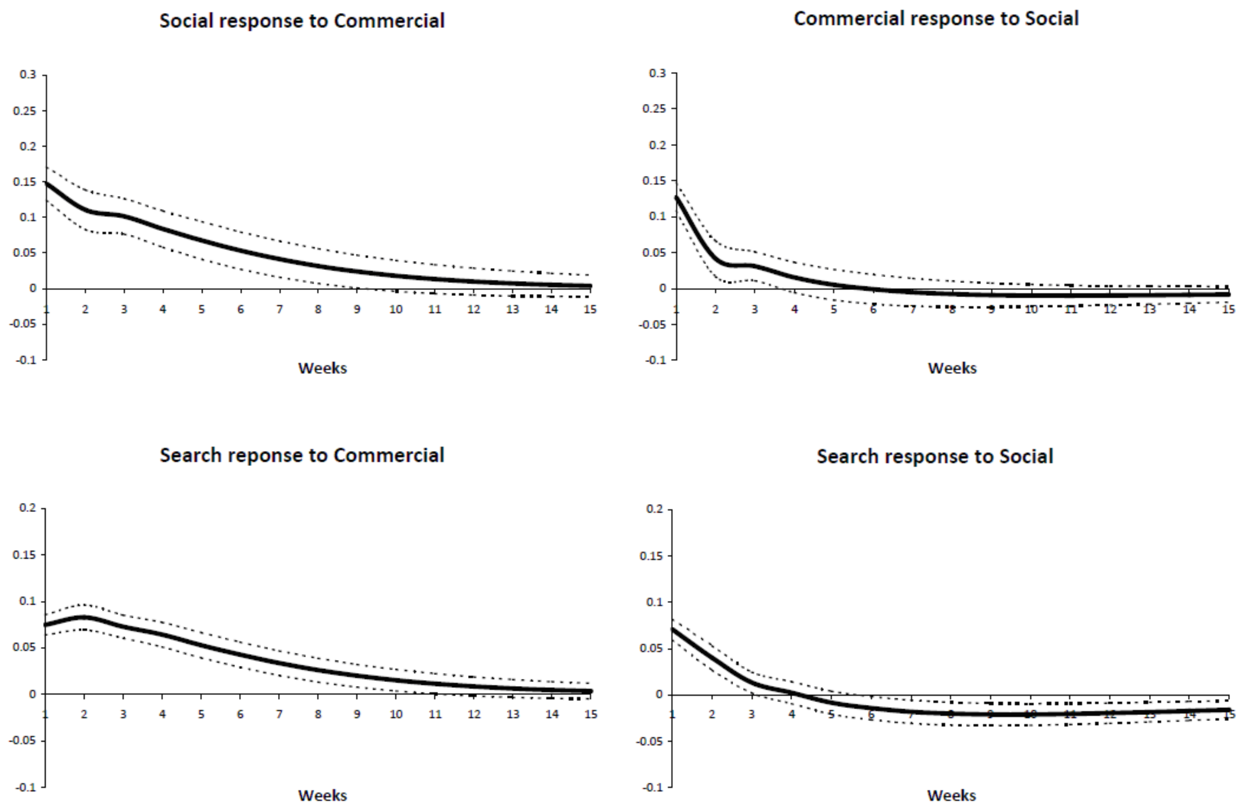


Figure 3a: Media Asymmetry for Hugh Jackman



Note: IRFs can be interpreted as follows. Looking at IRF (,1,2), a 1% increase in social media volume will lead to a .15% increase in commercial media volume, which will last for approximately 7 weeks.

Figure 3b: Media Asymmetry for Jack Nicholson



Note: IRFs can be interpreted as follows. Looking at IRF (1,2), a 1% increase in social media volume will lead to a .13% increase in commercial media volume, which will last for approximately 4 weeks.

Figure 4: IMR Map Demonstrating Heterogeneity among Stars