Using Online Conversations to Study Word of Mouth Communication

David Godes and Dina Mayzlin*

August 2003

*David Godes is an Assistant Professor in the Graduate School of Business Administration, Harvard University, Soldiers Field, Boston MA 02163 email: dgodes@hbs.edu, phone: 617-495-8040, fax: 617-496-5637. Dina Mayzlin is an Assistant Professor in the School of Management, Yale University, New Haven, CT 06520 dina.mayzlin@yale.edu. The authors would like to thank Guido Kuersteiner, Subrata Sen, Duncan Simester, Christophe Van den Bulte, Birger Wernerfelt, Steve Shugan, the Associate Editor, and four anonymous referees for comments on earlier drafts of this paper as well as participants in the MIT Marketing Lunch, the Wiesbaden 2001 Marketing Science Conference, HBS Marketing Seminar, Yale Marketing Seminar, 2002 NYU/Columbia/Wharton/Yale Consortium, UC Berkeley Marketing Seminar, Stanford Institute for Theoretical Economics Summer Conference, and the Young and Rubicam Brand Buzz Summer Lunch Workshop. This research also benefited from conversations with Sendhil Mullainathan in the early phases of development. Both authors contributed equally and their names are listed in alphabetical order.

Using Online Conversations to Study Word of Mouth Communication

Abstract

Managers are very interested in word of mouth communication because they believe that a new product's success is related to the word of mouth that it generates. However, there are at least three significant challenges associated with measuring word of mouth. It is our primary objective in this paper to address these challenges. First, how does one even gather the data? Since the information is exchanged in private conversations, direct observation is – or at least has traditionally been – quite difficult. Second, even if one could observe the conversations, what aspect of them should one measure? The third challenge comes from the fact that word of mouth is not exogenous. While the mapping from word of mouth to future sales is of great interest to the firm, we must also recognize that word of mouth is at the same time an outcome of past sales. We find that on-line conversations may offer an easy and cost-effective opportunity to measure word-of-mouth. However, simply counting on-line conversations may not be informative. On the other hand, measuring the "dispersion" of these conversations across communities is. Specifically, we show that a measure of dispersion has explanatory power in a dynamic model of sales. As a context for our study, we have chosen new TV shows during the 1999/2000 seasons. Our source of word-of-mouth conversations is Usenet, a collection of thousands of newsgroups with very diverse topics.

Keywords: word-of-mouth, diffusion of innovations, measurement, networks and marketing, new product research, Internet marketing

1 Introduction

Among the many and varied channels through which a person today may receive information, it is hard to imagine any that carry the credibility and, as a result, the importance of interpersonal communication, or word of mouth (WOM). There is little debate as to whether WOM *matters* to the firm. In fact, there is good reason to believe that it has more potential impact than any other communication channel. Katz and Lazarsfeld (1955) showed nearly half a century ago that WOM was the single most important source of information for certain household items. More recently, Kotler (2000) cites a study of 7,000 consumers in seven European countries in which 60 percent said they were influenced to buy a new brand by family and friends. Similarly, a 1999 study by Jupiter Communications found that 57% of people visiting a new web site did so based on a personal recommendation, far higher than any other source of influence. As these studies suggest, managers are interested in WOM for one simple reason: success of a product is related to the WOM that it generates. In fact, it is commonly believed that WOM directly impacts sales. It might affect awareness in some cases, or preferences in others. On the other hand, WOM may simply serve as a leading indicator of a product's success. Whatever the specific mechanism, there is significant empirical evidence, as well as a strong intuitive sense, for the hypothesized link between WOM and product sales.

The managerial implication of the above putative relationship is that the firm should measure WOM. As a leading indicator, WOM measurement would be important for purposes of market research. As a sales driver, WOM measurement would be an essential prerequisite to effective "buzz management." After all, to paraphrase Edward Deming, "you can't manage what you can't measure." However, there are at least three significant challenges associated with measuring WOM. First and foremost, how does one even gather the data? Since the information is exchanged in private conversations, direct observation is – or at least has traditionally been – quite difficult. As a result, most marketers and researchers have either relied on consumer recall or have inferred the process of information exchange from aggregate data. One fascinating and important implication of the rise of on-line communities is that this development makes feasible the observation by marketers of consumer-to-consumer conversations. In this paper, we investigate the potential use of these conversations in measuring WOM.

Second, even if we could observe the conversations, what aspect of them should we measure? Sales are easy to capture quantitatively. How does one *measure a conversation*, a set of statements between people? Should we count the words? The number of people involved in it? It is a significant challenge to determine which of the possible transformations of a conversation are meaningful and managerially useful. Currently, the most common approach is to use simple counts. This approach is much like the news-clipping services that monitor how many times a firm's products are "mentioned." For example, Yahoo! Buzz Index keeps track of how many times users query a particular topic on the Yahoo! Search engine. We investigate the informativeness of this naïve measure. As well, we investigate another dimension of WOM: *dispersion*. We define this construct as the extent to which conversations about the product are taking place across a broad range of communities. We expect that *less dispersed* WOM – discussions focused within a narrow and homogenous population – is likely to have less of an impact than those that are more broadly dispersed.

The third challenge comes from the fact that WOM is not exogenous. While the mapping from WOM to *future* sales is of great interest to the firm, we must also recognize that WOM is at the same time an outcome of *past* sales. This has implications for the measurement of WOM as well as for the interpretation of any measurement. High WOM today does not necessarily mean higher sales tomorrow. It may, in fact, just mean that the firm had high sales yesterday. Thus, to truly understand the nature of the link, we need to understand the full dynamic relationship between sales and WOM. Further, we must allow for the fact that the role – and/or impact – of WOM may change over a product's life.

As a context for our inquiry, we have chosen new TV shows during the 1999/2000 seasons. WOM appears to be especially important for entertainment goods: a recent Forrester report concludes that approximately 50% of young Net surfers rely on WOM recommendations to purchase CDs, movies, Videos/DVDs and games (Forrester Research, 2000). Thus, television shows are a natural candidate to use for testing the dynamic nature of word-of-mouth. In addition, the "purchase" of a TV show is a repeat purchase. This is interesting in this context because the consumer's purchase experience in period t will affect not only her decision to talk about it but also her consumption experience in period t+1. Our source of word-of-mouth information is Usenet, a collection of thousands of newsgroups with diverse topics.

It is our primary objective in this paper to address the above challenges associated with measuring WOM generally. In so doing, we will evaluate the informativeness of the measures – volume and dispersion – to the manager. Specifically, we envision a manager attempting to learn from aggregate data the underlying process governing the sales of her product. If she had the opportunity to measure WOM, this paper offers unique insight into which aspects of it she should measure. Given this focus, we are seeking measures that are practical to implement from both the perspective of hard costs and effort. We make no claim that the measures we investigate here are in any sense optimal. Instead, we hope to show when and if they have explanatory power in the dynamic sales model. In addition, we also obtain more costly measures related to the underlying content of the conversations.

Another objective of the paper is to investigate the usefulness of on-line conversations in the study of WOM communications. The context we study is characterized by a purchase decision made off-line, yet we are measuring WOM on-line. Thus, to the extent that we find that certain measures are informative, we argue that this supports the idea that at least some aspects of on-line

WOM are proxies for overall WOM (off-line as well as on-line). Given the obvious operational advantages of measuring WOM on-line, we hope to spur a significant increase in focus on the Web as a laboratory for WOM research.

The paper addresses these challenges as follows. After reviewing the relevant literature in Section 2, we discuss in detail our research objectives in Section 3. Based on existing theories, we're interested in understanding the extent to which dispersion and volume are important and informative dimensions of WOM communication. In Section 4, we describe in detail the two sources of data used in the study: Nielsen ratings and WOM data from Usenet newsgroups. In Section 5, we present the main empirical results. We find that higher WOM dispersion is related to higher future sales. We also find that the impact of dispersion declines over time. This argues for the explicit consideration of WOM early in a product's life. Surprisingly, we find that volume is *not* consistently associated with higher future sales. We discuss this surprising result in Section 6. One potential explanation for the null result could be the fact that positive and negative volume have different and offsetting associations with future sales. Since the valence of the post is unobserved in our main analysis, these effects may cancel each other out. To test this, we utilize a sampling scheme to collect valence data. Nonetheless, these regressions which account for positive and negative WOM separately do not yield the expected associations with future sales. Another explanation might be that there is less additional information from a volume measure – as compared with a dispersion measure – conditional on past sales. A regression of the contemporaneous explanatory variables supports this. Finally, it could be that the simple linear relationship that we specify between WOM and sales is not sufficiently rich to capture reality. We conclude in Section 7 with a discussion of the findings, their implications, their limitations and suggestions for future work.

2 Literature Review

Our work draws on three streams in the WOM literature: (1) WOM as a driver of future sales, (2) the importance of social structure in the flow of WOM, and (3) the multi-dimensional nature of WOM. In addition, we discuss the traditional approaches that have been taken to measure WOM.

2.1 WOM as a Driver of Sales

There exists ample theoretical support for the idea that WOM may impact a firm's sales. Banerjee (1992, 1993) presents two models that suggest that people may place significant weight on the opinions of others. In fact, rational agents may ignore their own private information in favor of information drawn from observation of others' actions. This may lead to "herding" in which all agents select the same action, even if each individually has information that favors another action.

A similar context is analyzed by Bikhchandani et al. (1991). An important implication of their work is that the introduction of new pieces of information can cause radical, discontinuous shifts in the actions of the agents. This may explain, they argue, fads and bubbles. Mayzlin (2001) focuses specifically on WOM on-line and the potential that it presents for the firm to pose as a consumer and create firm-to-consumer communications that "looks like" consumer-to-consumer communication. She finds that, even when this is possible, rational consumers still pay attention to anonymous on-line posts. As a result, posing as a customer on line may be a profitable equilibrium strategy for the firm.

There have also been numerous experimental and empirical attempts to provide support for this role of word-of-mouth, with mixed success. Reingen et al. (1984) conduct a survey of the members of a sorority in which they measure brand preference congruity as a function of whether they lived in the sorority house or not. Those that lived together had more congruent brand preferences than those that didn't. Presumably, those that lived together had more opportunities for interaction and thus WOM was more prevalent. A similar study, in a very different context, was performed by Foster and Rosenzweig (1995). They look at the adoption of high-yield varieties (HYV) of seeds among Indian farmers. They find that the profitability of farmers employing the HYV's was higher as the adoption rate of the village increased. They interpret this as a learning spillover in that the more experienced one's neighbors become with a new technology, the better one is at employing it. Again, the presumption here is that there is significant WOM at the village level which facilitates the flow of information regarding the new technology. They also present evidence that WOM has a positive but small effect on the farmers' rate of adoption of the new HYV's.

Van den Bulte and Lilien (2001a) call into question the general primacy of WOM communication as a sales driver. They revisit an analysis performed by Coleman et al. (1966), arguing that the latter erred in their conclusion that social contagion was the driving factor behind physicians' adoption of the new product. By specifying the information available to the physicians as well as their social networks, the authors show that marketing effort, and not interpersonal communication, dominated. In Van den Bulte and Lilien (2001b), the same authors decompose the adoption process into an awareness phase and an evaluation/adoption phase. In this model, they find evidence of social contagion.

2.2 The Impact of Social Structure

While there exist many reasons to believe that WOM is often important in driving future sales, it's less clear *which* aspects of WOM are especially important. Existing literature has demonstrated that "not all WOM is created equal." Depending on who is talking to whom, the conversation can have more or less impact. Granovetter (1973) characterizes relationships as being either "strong ties" or "weak ties." Moreover, he assumes that if A and B are connected by a strong tie and B and C are connected by a strong tie, then A and C must *also* be connected by a strong tie. He defines the case where this condition is violated as the "forbidden triad." We might make the further assumption that communities or groups are characterized by relatively strong ties among their members. Then, a direct implication of this model is that the only connections between communities are those made along weak ties. Granovetter characterizes these weak tie connections as "bridges." This highlights the critical role played by weak ties in the diffusion of WOM: any piece of information that traverses a weak, as opposed to a strong, tie has the opportunity to reach more people. This has the important implication that *information moves quickly within communities but slowly across them.*

In a similar vein, Kaplan et al.'s work in mathematical bioscience (Kaplan et al., 1989) shows that different patterns of contact between groups with different incidence of HIV/AIDS have a different impact on the spread of the disease. This modeling approach has been utilized in the marketing literature by Putsis et al. (1997). There, the authors find heterogeneity in "mixing" behavior across 10 EC nations. Importantly for this study, Putsis et al find that there is greater interaction within a population of a country than between populations of different countries (with heterogeneity in the propensity to interact across the different countries).

2.3 WOM as an Outcome

Part of the difficulty in measuring WOM is the fact that it is not only a precursor to, but also an outcome of, sales. There have also been many papers that provide evidence of the latter. Richins (1983) looks at the moderating factors that determine whether one talks about his or her negative experience. Anderson (1998) looks at the entire spectrum of WOM communication, from negative to positive. He proposes a utility-based model that gives rise to a U-shaped function: very dissatisfied customers and very satisfied customers are most likely to engage in WOM. He finds support for these hypotheses using a panel data set on customer satisfaction.

Bowman and Narayandas (2001) investigate the firm's disposition of customer-based inquiries (CICs). Two outcomes of this process are market share and WOM behavior. An intermediate outcome is customer satisfaction. They measure WOM via a survey, capturing both the incidence of WOM (whether the customer told someone else of their experience) and the breadth of referral (how many people they told). They find additional support for the U-shaped model put forth in Anderson (1998). Moreover, they find that WOM is increasing in customer loyalty: those customers that described themselves as loyal customers of the brand were significantly more likely to engage in WOM. However, these customers were less likely to engage in WOM the higher their satisfaction with the outcome of their inquiry. The authors suggest that this indicates that loyal customers engage only in negative WOM and only when they are dissatisfied.

2.4 Measurement Techniques

WOM activity has typically been analyzed using two methodologies: inference and/or surveys. Examples of the former include Foster and Rosenzweig (1995) in which the farmers in the dataset were never explicitly asked about their WOM behavior. Instead, by comparing across villages, the researchers assume that "learning spillovers" take place *within* villages at a higher rate than they do *across* villages. Similarly, Reingen et al. (1984) infer the presence of interpersonal communication by comparing women who live in the same house with those that do not. The presumption is that those that live in closer proximity are more likely to exchange information with each other. Finally, Bass (1969) and those that have extended his model also infer WOM from other data. In these models, the "coefficient of imitation" is estimated using aggregate-level sales data.

Surveys remain the most popular method to study WOM. Bowman and Narayandas (2001); Brown and Reingen (1987); Reingen and Kernan (1986); Richins (1983) all base their analyses on proprietary surveys designed to test a specific hypothesis. Van den Bulte and Lilien (2001a) and Anderson (1998) draw on the existence of survey-based data that were prepared for other, more general, purposes. The attraction of the survey in this context is precisely that one is able to ask the direct question, "did you tell somebody about X?" In some cases, like Bowman and Narayandas (2001), one might even ask, "How many did you tell?" Additionally, some researchers have found it useful to design and use surveys to map out social networks. For example, Reingen and Kernan (1986) used surveys to map out the entire social network comprised of the customers of a piano tuner. With this, they were able to understand which people played particularly important roles in the referral process. Brown and Reingen (1987) did so for piano teachers. Similarly, the dataset used by Van den Bulte and Lilien (2001a) contained data for each physician about the other physicians with whom he or she discussed medical practices and from whom he or she sought advice.

One purpose of this paper is to offer an alternative method for the firm to measure word of mouth. On-line conversations may offer the firm an attractive opportunity to learn about its environment by *directly observing* the flow of interpersonal communication. By looking at activity across different online communities, we will argue that we are able to obtain measures of social structure. As compared with the survey method, direct observation is potentially lower-cost and eliminates any reliance on recall. The downside of our method, however, is that we are not able to control for certain individual-level factors that a survey is capable of doing. So, for example, we wouldn't be able to identify "loyal users" as Bowman and Narayandas (2001) do.

3 Research Objectives

Our goal in this paper is to begin the decomposition of the *construct* "word of mouth" into pieces that are informative to, and potentially manageable by, the firm. We investigate two distinct dimensions of WOM, volume and dispersion. These measures are attractive in that they are implementable by the firm at low cost and effort. The first and most obvious dimension of WOM is its *volume*: how much WOM is there? This is essentially what has been measured by Bowman and Narayandas (2001), Reingen and Kernan (1986), Richins (1983), Anderson (1998), Van den Bulte and Lilien (2001a,b), the Yahoo! Buzz Index and others. The more conversations there are about this paper, for example, the higher the likelihood that someone who hasn't been informed will become informed. Since awareness is a necessary condition for purchase, we expect that higher volumes of WOM will be associated with higher future sales.

As Mohr and Nevin (1990) did in inter-firm communication, we investigate two distinct dimensions of inter-personal communications. Using assumptions similar to those made in Granovetter (1973) and supported by Putsis et al. (1997), we expect WOM to spread quickly within communities and slowly across them. Members of the same community have frequent interaction with each other and thus are more likely to learn from each other than from members of other communities. If this is true, then conditional on a certain volume of word-of-mouth, more people will become informed about something the more "dispersed" this information is between communities. This motivates us to explore the relationship between WOM dispersion and future sales. We expect that this relationship will be positive.

Finally, we explore the dynamics in the relationship between WOM and sales. We want to understand not only *which* aspects of WOM are informative but also *when* the informativeness is particularly high. This is important for managers as it identifies when in the process they may want to make the biggest investment in information gathering and in influencing the flow of information. We expect that the magnitude of the effect of both the dispersion and the volume of WOM on future sales will decrease over time. This is primarily because as people become better informed about their preferences for different products, there is likely to be less added value from a recommendation.

Three comments are in order with respect to our proposed measures. First, we can draw an analogy between these measures and those used in advertising: reach and frequency. Traditionally, people have focused on counts or volume to measure WOM. This is an analog of frequency: how often are people talking about the product? We hypothesize that a measure like reach would also be useful: how many different people are talking about it?

Second, note that both of these ignore potentially valuable information contained in the content of the conversations themselves. In particular, the volume of WOM may have a very different effect depending on the valence of comments it contains. The downside of collecting this content data is that doing so is a costly and noisy process, as we demonstrate in Section 6. Nonetheless, it is interesting to compare the informativeness of these deeper, but more costly, measures to the simpler and more efficient measures.

Finally, note that we're interested in the informativeness of these measures conditional on past sales data. We expect a lot of variance in current sales to be explained by past sales. Since past sales drive current WOM activity and since the manager observes the sales, it is essential to account for this in our model. We want to see how much extra information exists in the WOM data.

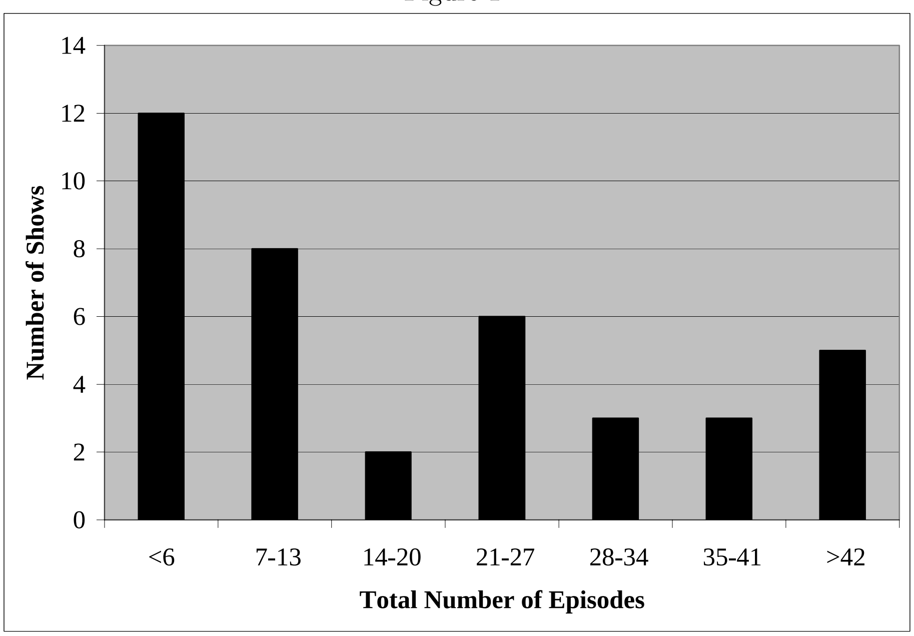
4 Data

We study the 44 TV shows that premiered in the US market during the 1999/2000 season by combining two publicly-available datasets. For "sales" data, we use Nielsen ratings (reported weekly in *Broadcasting & Cable* magazine) and for WOM we use conversations observed in Usenet newsgroups.

4.1 Ratings Data

Our sample includes only the shows aired on the six major networks: ABC, CBS, NBC, FOX, UPN, and WB. Only 14 shows survived into the 2000/2001 season. A few of the shows were cancelled quickly: four shows were cancelled after only 2 episodes. Half of the shows were shown fewer than 17 times.

Figure 1



In Figure 1, we present the distribution of total episodes of a new show. The "rating" reflects the percentage of households who watched the show that week. Table 13 in the Appendix lists the shows. Table 14 in the Appendix summarizes the data by network.

Table 1

Top Five Premieres					
Show	Network	Day of Week	Date	Nielsen Rating	TV Homes (millions)
Judging Amy	CBS	Sun	9/19/1999	13.5	13.4
Stark Raving Mad	NBC	Thur	9/23/1999	12.3	12.2
Once and Again	ABC	Tue	9/21/1999	12.3	12.2
Malcolm in the Middle	FOX	Sun	1/9/2000	12.1	12.2
West Wing	NBC	Wed	9/22/1999	12.1	12.0

Table 2

Worst Five Premieres					
Show	Network	Day of Week	Date	Nielsen Rating	TV Homes (millions)
DC	WB	Sun	4/2/2000	1.6	1.6
Mission Hill	WB	Tue	9/21/1999	1.8	1.8
The Beat	UPN	Tue	3/21/2000	2.2	2.2
The Strip	UPN	Tue	10/12/1999	2.3	2.3
Popular	WB	Thur	9/30/1999	2.5	2.5

The variance in the ratings is very high. Tables 1 and 2 present the most and least successful premieres, respectively. While 13.4 million households watched the premiere of *Judging Amy*, only 1.6 million households watched the premiere of *DC*. Note that while most of the shows premiered in late September or early October 2000 following the Sydney Summer Olympics, some shows were mid-season replacements.

4.2 WOM Data

Our WOM data are drawn from Usenet newsgroups. These are attractive sources of data for several reasons. First, a historical archive of Usenet newsgroups – some reaching as far back as twenty years – is currently publicly available at http:\\groups.google.com.[1] In comparison to the social network mapping procedures carried out by other researchers, this dataset offers a painless and affordable alternative. Moreover, there is a wide breadth of topics covered on Usenet, from rec.autos.sport.nascar to alt.fan.noam-chomsky. Thus, it seems that this is a fertile area for both managers and academics to conduct research on WOM. These benefits do not come without costs; there is a potential for bias at two levels. First, on-line conversations may not necessarily be representative of all conversations. Moreover, the subset of on-line conversations that are being

[1] At the time of our data collection, the Usenet data were archived by deja.com. The archive has since been purchased by Google.

held on Usenet may not be a representative sample of all on-line conversations. It would seem that both of these potential biases would, if anything, decrease the estimated relationship between WOM and future sales.

A Usenet posting contains the author's nickname, a subject line, the name of the newsgroup to which the post was sent, the date of the post, and the text of the message. The archive is searchable by subject, author, group, etc. Within each newsgroup, posts are organized into "threads" which contain posts on roughly the same topic. One might think of a thread as the on-line analog of a "conversation." Very often, all posts in a thread contain the same subject line. For an example of a complete thread, see Appendix A.1.

We restrict our analysis to newsgroups with names beginning with either "alt.tv" or "rec.arts.tv." To identify a post as being "about" a show, we looked for the name of the show in the subject line. This is a conservative approach as there are likely to be a fair number of posts about shows which do not include the show's name in the subject line. We found 169 different groups that contained messages about the shows in our sample. The groups' focus varies from television in general (for example, "rec.arts.tv") to specific shows (for example, alt.tv.x-files is visited by fans of *The X-Files.*). Those who visit alt.tv.x-files often chat about other shows that they find interesting. Appendix A.1 presents a thread that deals with the show *Roswell* that takes place in alt.tv.x-files. This is not particularly surprising since both are science fiction shows. It takes time for fans to assemble a newsgroup devoted to a new show like *Roswell*. In the initial period following the debut of a show, the conversations are dispersed among groups that are devoted to other shows. Table 3 presents the 20 newsgroups that had the most postings about the shows in our sample. None of these groups is specifically devoted to conversations about any show in the sample.

We excluded three of the 44 shows from the sample: *Angel*, *Harsh Realm* and *Grownups*. We exclude *Angel* because we found too many posts – over 3,000 – that contained the word "angel" in the subject line. From a simple reading of the entire subject line (not the content of the message), it was clear that most of the posts were obviously unrelated to the show. On the other hand, there were no posts that contained the words "grownups" or "harsh realm" in the subject line. This demonstrates that our technique for extracting posts is imperfect. This is especially the case when shows' names contain common words such as "angel" or involve shows that generated very little buzz. We emphasize again that, for our main analysis, we do not analyze the post's *content*. We revisit this issue below in Section 6.

4.3 Variables

From these conversations, we construct measures corresponding to "volume" and "dispersion" discussed in Section 3. Let $n = 1, ..., N$ index the newsgroups. We define $POST_{it}^{n}$ as the number

Table 3

20 Top Newsgroups in the Sample	
Group	Number of posts
rec.arts.tv	9,649
alt.tv.game-shows	2,892
alt.tv.law-and-order	1,621
alt.tv.party-of-five	1,013
alt.tv.homicide	932
alt.tv.buffy-v-slayer	764
rec.arts.tv.mst3k.mis	578
alt.tv.simpsons	533
alt.tv.star-trek.voya	527
alt.tv.dawsons-creek	498
alt.tv.x-files	440
alt.tv.er	391
alt.tv.emergency	326
alt.tv.millennium	311
alt.tv.newsradio	258
alt.tv.real-world	236
alt.tv.highlander	176
alt.tv.3rd-rock	162
alt.tv.twin-peaks	153

of posts in newsgroup n about show i between episodes t and $t+1$. So, the "volume" of WOM is:

$$POST_{it} = \sum_{n=1}^{N} POST_{it}^{n} \tag{4.1}$$

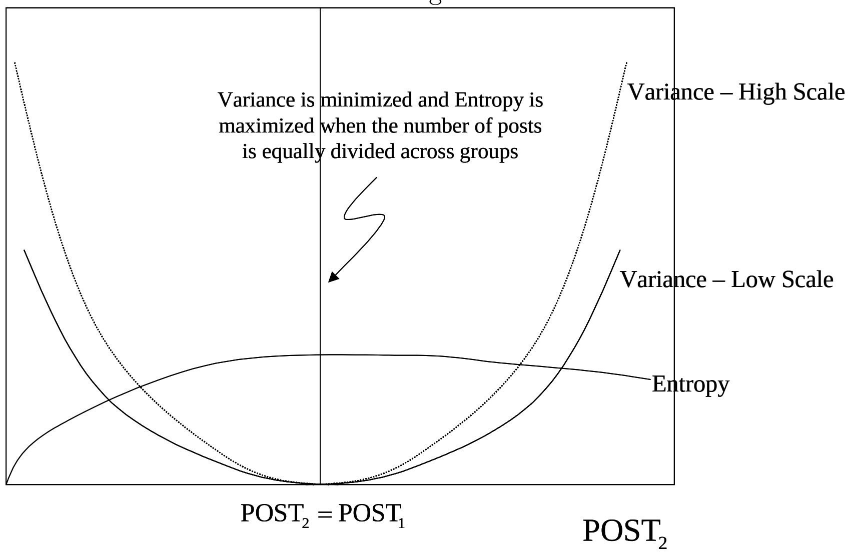
We operationalize dispersion as the "entropy" of conversations across newsgroups. This is a fairly common measure in the information theory literature (Zwillinger, 1996). Here entropy is defined as follows:

$$ENTROPY_{it} = \begin{cases} -\sum_{n=1}^{N} \frac{POST_{it}^{n}}{POST_{it}} Log\left(\frac{POST_{it}^{n}}{POST_{it}}\right) & if\ POST_{it} > 0 \\ 0 & if\ POST_{it} = 0 \end{cases} \tag{4.2}$$

We prefer entropy to variance because the former is independent of the total volume of posts. Variance is maximized (and entropy minimized) if the posts are all concentrated in one newsgroup. Entropy is maximized (and variance minimized) when posts are evenly distributed across all of the groups in which there is at least one post. Figure 2 presents a comparison of variance and entropy.[2]

[2] This figure depicts variance and entropy in a context in which there are two newsgroups. The number of posts in the first newsgroup, $POST_1$, is fixed; the x-axis captures the number of posts in the second group, $POST_2$. Two

Figure 2



We also calculate an alternative measure of dispersion that simply counts the number of newsgroups in which posts appear about show i after episode t:

$$NUMGROUPS_{it} = \sum_{n=1}^{N} 1\left(POST_{it}^{n} > 0\right) \tag{4.3}$$

where $1(\cdot)$ is the indicator function.

While most shows air at the same time every week, this is not always the case. Some have episodes separated by more than a week, perhaps due to special programming. Others run more than once a week, particularly early on. The results that we present below do not control for these factors. We have estimated alternative specifications that do so, however. Since the results are qualitatively equivalent, we do not present them here. We do control for the fact that sometimes two episodes of the same show run on the same day, which is a relatively rare occurrence. In this scenario, we are not able to infer which episode generated the WOM that occurs in the following days. In this case, we use the ratings from the first episode that day and exclude the second.

Our dependent variable is $RATING_{it}$, the rating of episode t for show i. Since there is also a strong time trend in TV ratings, we also include a time variable $EPISODE_{it} \equiv t$. Finally, in order to investigate WOM, we define the "early" period to be the first τ episodes of a show. That is, we define a dummy variable,

$$EARLY_{it} \equiv 1\left(t \leq \tau\right) \tag{4.4}$$

variance curves are provided. The "high scale" curve depicts variance when the number of posts in *both* groups is multiplied by a constant greater than one. Note that entropy is not affected by this scaling.

Thus, τ represents the number of episodes that comprise the "early" period in a show's life. We estimate our models across a range of τ values.

Table 4 provides summary statistics for the variables used and Table 5 provides pairwise correlations.

Table 4

Summary Statistics

Variable	Mean	Std. Dev.	Min	Max
$RATING_t$	5.48	2.97	0.70	14.10
$POST_{t-1}$	27.76	41.21	0.00	261.00
$ENTROPY_{t-1}$	0.49	0.66	0.00	3.00
$NUMGROUPS_{t-1}$	1.96	2.24	0.00	20.00

Table 5

Pairwise Correlations

	$RATING_t$	$RATING_{t-1}$	$POST_{t-1}$	$ENTROPY_{t-1}$	$NUMGROUPS_{t-1}$	EPISODE
$RATING_t$	1					
$RATING_{t-1}$	0.9109	1				
$POST_{t-1}$	0.0825	0.1240	1			
$ENTROPY_{t-1}$	-0.1366	-0.1158	0.4536	1		
$NUMGROUPS_{t-1}$	-0.0946	-0.071	0.6629	0.8798	1	
EPISODE	-0.1031	-0.1279	-0.0762	-0.1072	-0.0839	1

5 Main Results

There are (at least) two ways to investigate the role of WOM "early" in a show's life. One approach would be to restrict, or truncate, the dataset to only those episodes that were aired early on. Another approach would be to use all of the data but estimate separate coefficients for those observations that occurred early and those that occurred late. The advantage of the truncated approach is that it is conceptually appealing. It matches closely the context being faced by the manager: after, say, five episodes, she wants to understand how good her show may be, for example. The advantage of the latter approach is that we have more data and it allows us to compare directly the role of WOM early and late. Taking the best of both worlds, we present our main findings using the conceptually appealing truncated approach but investigate dynamics using all of the data. As we show below, the results for the effect of early word of mouth are similar across these two approaches.

5.1 Model with Early Data Only

We estimate the following model:

$$RATING_{it} = \lambda \cdot RATING_{i,t-1} + \pi \cdot POST_{i,t-1} + \delta \cdot ENTROPY_{i,t-1} + \beta \cdot EPISODE_{it} + u_i + \varepsilon_{it} \text{ for t} \leq \tau \quad (5.1)$$

We include a fixed effect for each show: u_i. This captures a combination of scheduling influences – the network, the day of week, the previous show – as well as each show's intrinsic "quality."[3] It is well known that the estimation of a fixed effects model with a lagged endogenous variable is subject to potential finite sample bias (Nerlove, 1967, 1971; Nickell, 1981). We note here that we wouldn't necessarily expect the bias to be substantial in our estimates since the number of observations per show is not extremely low here (mean = 15). Nonetheless, we estimate in Appendix A.2 a model that isn't subject to this bias and show that our results are qualitatively equivalent.

Table 6

Truncated Sample Fixed Effects Model
(t- statistics beneath)

	$\tau = 4$		$\tau = 5$		$\tau = 6$		$\tau = 7$	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$RATING_{i,t-1}$	-0.5484 ***	-0.5557 ***	-0.4607 ***	-0.4557 ***	-0.4234 ***	-0.4235 ***	-0.2997 ***	-0.2997 ***
	-6.40	-6.53	-5.85	-5.86	-5.89	-5.87	-4.24	-4.23
$POST_{i,t-1}$	0.0027	0.0039	0.0031	0.0046	0.0043	0.0043	0.0051 *	0.0046
	0.71	1.02	0.87	1.28	1.48	1.35	1.73	1.41
$ENTROPY_{i,t-1}$	0.5769 **	1.0738 **	0.3819 **	0.9658 ***	0.2975 *	0.2945	0.2063	0.1018
	2.42	2.63	2.07	2.64	1.87	0.94	1.29	0.36
$NUMGROUPS_{i,t-1}$		-0.2531		-0.2765 *		0.0014		0.0500
		-1.49		-1.85		0.01		0.45
$EPISODE_{it}$	-0.3445 ***	-0.3699 ***	-0.2329 ***	-0.2495 ***	-0.1869 ***	-0.1870 ***	-0.1636 ***	-0.1637 ***
	-2.95	-3.16	-2.92	-3.15	-3.40	-3.37	-3.52	-3.51
N	109	109	138	138	168	168	195	195
R^2	0.45	0.47	0.31	0.33	0.27	0.27	0.18	0.18
F Test: All Coefficients = 0	13.14	11.16	10.27	9.11	11.45	9.09	8.20	6.56
Pr > F	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
*** = p<.01								
** = p<.05								
* =p<.10								

The estimation of 5.1 is presented in columns (1), (3), (5) and (7) of Table 6. More dispersed WOM seems to associated with higher future ratings early in the show's life.[4] The coefficient

[3] A random effects model would be preferable but show "quality," we'd expect, is correlated with, $RATING_{i,t-1}$. A specification test confirmed this.

[4] Note that our analysis occurs at the post level, not at the individual poster level. Thus, we do not capture the fact that some posters may be participating in several newsgroups. Moreover, the interpretation of the results may be different depending on the extent to which entropy is caused by different people in different communities or by the same people participating across many communities. We implicitly assume the former but one should consider

on $ENTROPY_{i,t-1}$ is positive and significant at the $p < .05$ level when $\tau = 4$ and when $\tau = 5$ and at the $p < .10$ level when $\tau = 6$. Thus, it seems that *more dispersed early conversations are associated with higher future ratings*. To illustrate the magnitude of this effect, consider a show that has 15 posts on one newsgroup and 5 posts on another, yielding an entropy of 0.562. The coefficient on entropy of 0.577 implies that a change in the distribution of posts to an even split between the two newsgroups (10 posts in each) would yield an entropy of 0.693 and would be associated with an estimated increase of approximately 75,000 viewers for the next episode. The coefficient loses significance as later episodes are included in the sample (i.e., τ gets higher). This finding is consistent with the expected decrease of impact of WOM over time. Surprisingly, we find somewhat less support for the effect of volume. The coefficient on $POST_{i,t-1}$ reaches only marginal significance when $\tau = 7$. Still, both of these measures appear to have some explanatory power in the specification and thus warrant further investigation. Nonetheless, a strategy of simply "counting" WOM appears to be less informative than also modeling and measuring the spread of WOM across communities.[5]

Since the dispersion measure captures both the *number* of communities and *distribution* of conversations across these communities, it is interesting to explore to what extent a measure of the number of groups alone can explain our results. To investigate this, we include the number of groups in addition to the entropy variable. The results of the regressions that include this variable are presented in columns (2), (4), (6) and (8) of Table 6. The key difference between $NUMGROUPS_{it}$ and $ENTROPY_{it}$ is that high values of the former may result when there are several communities covered but the preponderance of the activity occurs in just a few. As Table 6 shows, $NUMGROUPS_{i,t-1}$ is only marginally significant when we include $ENTROPY_{i,t-1}$. Most important, $ENTROPY_{i,t-1}$ retains its explanatory power in the early periods when $NUMGROUPS_{i,t-1}$ is added.[6] This is notwithstanding significant pairwise correlation (.88) between $NUMGROUPS_{it}$ and $ENTROPY_{it}$.

A few other interesting results also emerge. The negative coefficients on $RATING_{i,t-1}$ are evidence for mean reversion in ratings early in a show's life. We might infer that there's a lot of "sampling" of early episodes. Finally, there is a negative time trend in ratings. At the mean rating of 5.5, the coefficient of -0.3445 on $EPISODE_{i,t}$ implies a decrease in ratings of about 6% from episode to episode. While the findings in this section allow us some investigation of dynamics over time (i.e., the observation that the t-statistics on $ENTROPY_{it}$ decrease as τ is increased), our insights in this regard are constrained by our use of only the first τ episodes. A more detailed

that the latter could be at least partially at work. We are indebted to an anonymous referee for pointing this out.

[5] Since the analysis in this paper is somewhat exploratory, we estimated several variants of this specification in order to test the robustness of the main results. We estimated the equation taking logs of the RHS variables to capture possible decreasing marginal returns. We also estimated a model that included a non-linear episode variable. None of these estimations yielded significantly different results from the ones presented.

[6] In a specification in which $NUMGROUPS_{it}$ appears without $ENTROPY_{it}$, the former is never significant at the .10 level.

exploration of dynamics is best carried out by estimating the model on all the data.

5.2 Dynamics

First, we build on the results of the previous section by allowing a differential impact of all variables over the early and later episodes by estimating two different set of coefficients for these periods.[7] The model we estimate is:

$$\begin{aligned} RATING_{it} = {} & \lambda^E \cdot RATING_{i,t-1} \times EARLY_{it} + \lambda^L \cdot RATING_{i,t-1} \times (1 - EARLY_{it}) + \\ & \pi^E \cdot POST_{i,t-1} \times EARLY_{it} + \pi^L \cdot POST_{i,t-1} \times (1 - EARLY_{it}) + \\ & \delta^E \cdot ENTROPY_{i,t-1} \times EARLY_{it} + \delta^L \cdot ENTROPY_{i,t-1} \times (1 - EARLY_{it}) + \\ & \beta^E \cdot EPISODE_{it} \times EARLY_{it} + \beta^L \cdot EPISODE_{it} \times (1 - EARLY_{it}) + u_i^E + u_i^L + \varepsilon_{it} \end{aligned} \quad (5.2)$$

Note that the specification above essentially replicates 5.1 since all variables (including the fixed effects) are allowed to vary over the early and late periods. Indeed, as Table 7 demonstrates, the coefficients on the variables interacted with $EARLY_{it}$ are numerically identical to the coefficients in Table 6. The only difference is that the t-statistics are slightly lower in Table 7 since the variance-covariance matrix is re-estimated using all data in the model. We again find support for the idea that this association is strong early in a show's life but not later on since the coefficient on $ENTROPY_{i,t-1} \times (1 - EARLY_{it})$ is not significant for any of the values of τ. This does not mean that dispersion is unrelated to sales later in the product's life. On the contrary, given the dynamic nature of the process, dispersion in period five, for example, is strongly associated with sales in period six which drive sales in period 7 and so on. Thus, dispersion is likely to have a lasting indirect association with future sales even though the direct association seems to wane.[8] Again, we find less support for an association between the volume of WOM and sales (early WOM is significant at the .10 level only when $\tau = 7$).

One unattractive aspect of the approach above is that τ exogenously imposes a discrete change in the regime, while it is more likely that the change is continuous. We have estimated the model on several different values of τ to show the sensitivity to this exogenous assumption. Another approach – which would allow us to sidestep this decision – is to specify the WOM dynamics in terms of a continuous functional form. We investigate the following specification:

[7] Note that the specification in (5.2) is equivalent to one which specifies volume as $\theta_1 \cdot POST_{i,t-1} + \theta_2 \cdot EARLY_{it} \cdot POST_{i,t-1} = \theta_1 \cdot [POST_{i,t-1} \cdot EARLY_{it} + POST_{i,t-1} \cdot (1 - EARLY_{it})] + \theta_2 \cdot EARLY_{it} \cdot POST_{i,t-1} = (\theta_1 + \theta_2) \cdot POST_{i,t-1} \cdot EARLY_{it} + \theta_1 \cdot POST_{i,t-1} \cdot (1 - EARLY_{it})$.

[8] We are grateful to an anonymous referee for pointing this out.

Table 7

Full Sample Fixed Effects Model

(t- statistics beneath)

	$\tau = 4$	$\tau = 5$	$\tau = 6$	$\tau = 7$
$RATING_{i,t-1}$ x $EARLY_{it}$	-0.5484 ***	-0.4607 ***	-0.4234 ***	-0.2997 ***
	-5.55	-5.37	-5.47	-4.14
$RATING_{i,t-1}$ x (1-$EARLY_{it}$)	0.2068 ***	0.2135 ***	0.2386 ***	0.2315 ***
	4.87	4.98	5.37	5.02
$POST_{i,t-1}$ x $EARLY_{it}$	0.0027	0.0031	0.0043	0.0051 *
	0.62	0.80	1.37	1.69
$POST_{i,t-1}$ x (1-$EARLY_{it}$)	-0.0012	0.0031	-0.0019	-0.0020
	-0.75	0.80	-1.16	-1.20
$ENTROPY_{i,t-1}$ x $EARLY_{it}$	0.5769 **	0.3819 *	0.2975 *	0.2063
	2.10	1.90	1.73	1.26
$ENTROPY_{i,t-1}$ x (1-EARLYit)	0.0081	-0.0600	-0.0204	0.0416
	0.09	-0.63	-0.20	0.40
$EPISODE_{i,t-1}$ x $EARLY_{it}$	-0.3445 **	-0.2329 ***	-0.1869 ***	-0.1636 ***
	-2.56	-2.68	-3.15	-3.43
$EPISODE_{i,t-1}$ x (1-$EARLY_{it}$)	-0.02 ***	-0.02 ***	-0.02 ***	-0.02 ***
	-5.21	-4.51	-4.28	-4.37
N	688	688	688	688
R^2	0.15	0.13	0.14	0.12
F Test: All Coefficients = 0	13.34	11.82	12.37	10.81
Pr > F	.000	.000	.000	.000

*** = p<.01
** = p<.05
* =p<.10

$$RATING_{it} = \lambda_1 \cdot RATING_{i,t-1} + \pi_1 \cdot POST_{i,t-1} + \delta_1 \cdot ENTROPY_{i,t-1} + \theta \exp(-r \cdot EPISODE_{i,t}) \cdot ENTROPY_{i,t-1} + \beta_1 \cdot EPISODE_{it} + u_i + \varepsilon_{it} \quad (5.3)$$

In (5.3), we interact entropy with a decreasing function of the time trend: $\exp(-r \cdot EPISODE_{i,t})$. This allows us to vary the effects of entropy over time in a continuous manner. An increase in r implies a faster rate at which the impact of entropy is changing. For example, for episode 3, $\exp(-rEPISODE_{i,t}) = 0.687$ when $r = 0.125$ and $\exp(-rEPISODE_{i,t}) = 0.050$ when $r = 1$. This convex decline is important to capture since the results in Tables 6 and 7 indicate that the effect if dispersion declines fairly quickly. The results of our estimation of (5.3) are presented in Table 8. The coefficient on $ENTROPY_{i,t-1}$ (δ_1) is not significant, while the coefficient on the interaction term (θ) is significant for $r = 0.125, 0.25, 0.50$ and 0.75. These results are consistent with our earlier finding that the impact of entropy decreases over time. For example, according to our estimates for $r = 0.125$, the marginal effect of entropy on future ratings $(\frac{\partial RATING_{it}}{\partial\ ENTROPY_{i,t-1}} = \delta_1 + \theta \exp(-0.125 \cdot EPISODE_{i,t}))$ is 0.232 for episode 3, and 0.192 for episode 4. The assumed velocity of the decline (the value of r) does make a difference. The results are not significant when the assumed decline is too steep (for example, $r = 1$). We expect the steepness of the decline to vary by category. The effect of volume is not significant.[9] The positive coefficient on $RATING_{i,t-1}$ is not surprising here in light of Table 7. It is clear that, for most of the (latter part of the) lives of most shows, viewership has strong persistence.

Table 8

Episode - Entropy Interactions: Full Sample

Dependent variable: $RATING_{it}$	r = 0.125		r = 0.25		r = 0.50		r = 0.75		r = 1	
$RATING_{i,t-1}$	0.1454	***	0.1415	***	0.1413	***	0.1439	***	0.1465	***
	3.82		3.70		3.67		3.73		3.80	
$POST_{i,t-1}$	-0.0006		-0.0008		-0.0009		-0.0009		-0.0009	
	-0.44		-0.55		-0.64		-0.65		-0.64	
$ENTROPY_{i,t-1}$	-0.1045		-0.0528		-0.0132		0.0062		0.0174	
	-0.94		-0.55		-0.15		0.07		0.20	
exp (- r EPISODE) X $ENTROPY_{i,t-1}$	0.4891	**	0.6287	**	0.9436	**	1.3119	*	1.8059	
	2.13		2.29		2.08		1.74		1.45	
$EPISODE_{it}$	-0.0221	***	-0.0230	***	-0.0244	***	-0.0252	***	-0.0256	***
	-4.46		-4.86		-5.31		-5.54		-5.66	
N	688		688		688		688		688	
R^2	0.12		0.12		0.12		0.11		0.11	
F Test: All Coefficients = 0	16.84		17.00		16.79		16.50		16.29	
Pr > F	0.00		0.00		0.00		0.00		0.00	

*** = p<.01
** = p<.05
* =p<.10

6 Investigating the Role of Volume

The results in Section 5 suggest that *dispersion* is an important aspect of WOM worthy of the manager's attention However, these results do not provide consistent support for the importance of the *volume* of WOM. There are several potential reasons for this null result, some of which we investigate in this Section. On one hand, it may be an artifact of our data collection and analysis

[9] We have also estimated a specification where we interact $POST$ with the same function of episode. We do not find that either of the volume variables is ever significant, while the results for entropy are qualitatively similar to the ones presented (albeit the significance of θ is slightly reduced, especially for higher r).

approach. In particular, our focus on cost-effective data collection precluded our adoption of content analysis to assess what the posts actually said. This certainly decreases the amount of information available in our data. Negative and positive volumes may have offsetting relationships with future sales which may cancel each other out in our estimates. We investigate this in Section 6.1 by performing content analysis on a sample of the posts. In Section 6.2, we check whether volume and dispersion differ in terms of the amount of information they contain conditional on the other RHS variables. In Section 6.3, we discuss other possible explanations.

6.1 Valence Data Results

To investigate positive and negative WOM separately, we collected content data for a sample of posts in our database. The secondary benefit of analyzing the content of the discussions is that it allows us to gauge the difficulty of obtaining these labor-intensive measures compared to the measures discussed so far. Due to the volume of posts in our dataset (over 20,000 posts overall), it is infeasible for us to analyze the content of each of them. We initiated a sampling scheme in which we sample 10% of each shows' posts each week, rounded up. We employed two independent raters who were unaware of our research objectives. After reading the entire post, each rater was asked to classify the post into one of six categories:

$$1 - Positive$$
$$2 - Negative$$
$$3 - Neutral$$
$$4 - Mixed$$
$$5 - Irrelevant$$
$$6 - Not\ Sure$$

Of the 2,398 total posts that were evaluated in this manner, only 1,356 of them (57%) received identical categorizations from the two raters. This highlights the point that truly accurate content analysis in this domain is extremely difficult due to the subjective nature of the content.

For example, here is a post relates to the show *Movie Stars* that was given different ratings: 4 and 2:

> "I wonder if they're actually going to come up with a few jokes next season. I know, no-one listened to me when I was saying this about Animaniacs, but name-dropping is not in and of itself funny. The show's got a decent cast and a reasonably nifty premise, but it's no 'Thanks'."

To resolve differences, we employed a third rater to evaluate all posts on which the initial two raters disagreed. When this third rater agreed with one of the previous two, we used that evaluation. If the third rater didn't agree with either of the previous two, we assigned it to a seventh category of "disagreed posts." This method resulted in a net of 2,023 usable sample posts. See Table 9 for the distribution of the evaluations of the posts collected. A large proportion of the conversations (42%) were deemed not relevant to the shows under considerations. These conversations were either mistakenly included in our sample since the subject name coincidentally fit our criteria or the posts included the name of the show in the subject name, but, in fact, proceeded to discuss other issues.[10] The other notable result is that within the categories that stated an opinion on the show (that is, within a sample of positive, negative or mixed posts), almost three out of four of the recommendations were either positive or mixed. Moreover, there's nearly twice as much positive WOM as negative WOM.

Table 9

Distribution of Evaluations of Sample Posts

	Total Sample		Only Relevant and Valenced	
	Number	Percentage	Number	Percentage
Positive	326	14%	326	51%
Negative	176	8%	176	27%
Neutral	415	18%		
Mixed	139	6%	139	22%
Irrelevant	950	42%		
Not Sure	17	1%		
No Agreement	252	11%		
Total	2275	100%	641	100%

We define $SAMP\_POS\%_{it}$ as the percentage of sampled posts for week t for show i that were rated as positive. We similarly define $SAMP\_NEG\%_{it}$, $SAMP\_NEU\%_{it}$, $SAMP\_MIX\%_{it}$, $SAMP\_IRR\%_{it}$, $SAMP\_NS\%_{it}$, $SAMP\_DIS\%_{it}$, We then apply these sample proportions to the underlying number of posts for show i in week t to form the new independent variables that represent the inferred volume of posts broken down by valence based on our sampling technique. For example:

$$POS\_POSTS_{it} \equiv SAMP\_POS\%_{it} \cdot POST_{it} \tag{6.1}$$

In addition to reading the posts, we measured the length of the conversations – number of words in a post – since this may indicate either passion or the quality of the posts[11]. We obtain several measures, including, AVG_LENGTH_{it}, for all sampled posts for week t for show i , as well as

[10] This latter case, which was quite common, highlights the fact that while the content of the post itself was deemed irrelevant, it is not necessarily the case that the impact of that post was zero in terms of future sales. The fact that the name of the show was in the subject line and/or the contribution of the post to the overall impression of the volume of conversations suggests that even these "irrelevant" posts may have a marginal impact on a potential viewer's decision to sample the show.

[11] Note that this measure excludes text that is copied from the post to which the author may be replying.

AVG_LENGTH_{it} calculated separately for positive and negative posts.

Table 10

Pairwise Correlations: Valence Data				
	$POS\_POSTS_{i,t-1}$	$NEG\_POSTS_{i,t-1}$	$MIX\_POSTS_{i,t-1}$	$AVGLENGTH_{i,t-1}$
$POS\_POSTS_{i,t-1}$	1			
$NEG\_POSTS_{i,t-1}$	0.1827	1		
$MIX\_POSTS_{i,t-1}$	0.3208	0.2611	1	
$AVGLENGTH_{i,t-1}$	0.2313	0.1344	0.2867	1
$RATING_{t}$	0.0758	0.1097	0.0783	0.1129
$RATING_{t-1}$	0.1189	0.1603	0.1121	0.1106
$POST_{t-1}$	0.6669	0.4552	0.5161	0.2780
$ENTROPY_{t-1}$	0.3349	0.2203	0.2097	0.2380
EPISODE	-0.0719	-0.1611	-0.0878	-0.0858

We re-estimate specification (5.1), where the volume measure, $POST_{i,t-1}$, has been replaced by $POS\_POSTS_{i,t-1}$, $NEG\_POSTS_{i,t-1}$ and $MIX\_POSTS_{i,t-1}$. In addition, we also include $AVG\_LENGTH_{i,t-1}$. Table 10 presents the pairwise correlation matrix with the new variables included. Table 11 presents the estimation results. First, note that including the valence information does not weaken the inferred relationship between dispersion and future sales. As is clear from the coefficients on $ENTROPY_{i,t-1}$, this relationship appears even stronger. This is not surprising since we have eliminated some of the "noise" associated with the truly irrelevant posts. However, it seems further clear that – even with valence data – the volume of word of mouth does not have a strong relationship with future sales. We also do not find the length variables to significantly impact future sales.

6.2 Residual Information

As discussed in Section 2.3, we know that past sales are likely to at least partially drive current WOM activity. Thus, it may be the case that, conditional on knowing past sales, current WOM volume data is superfluous. Higher sales in time t should generally mean more positive conversations about the product in time t+1, all else equal[12]. However, the same argument cannot necessarily be made of dispersion. To check this intuition, we estimate and compare the following models:

$$POS\_POSTS_{i,t-1} = \beta_1 \cdot RATING_{i,t-1} + \beta_2 \cdot EPISODE_{it} + \eta_i + \psi_{it} \tag{6.2}$$

[12] Note that the same relationship may not necessarily hold for negative conversations, which is why we concentrate on positive conversations only here. Indeed, we find that the relationship between negative posts and lagged ratings is not significant.

Table 11

Truncated Sample Fixed Effects Model - Post Valence (t- statistics beneath)								
	τ = 4		τ = 5		τ = 6		τ = 7	
$RATING_{i,t-1}$	-0.5455	***	-0.4512	***	-0.4116	***	-0.3029	***
	-5.93		-5.57		-5.55		-4.17	
$ENTROPY_{i,t-1}$	0.6246	**	0.4097	**	0.3588	**	0.2344	
	2.46		2.2		2.22		1.46	
$POS\_POSTS_{i,t-1}$	0.0019		-0.0078		-0.0041		0.0025	
	0.14		-0.66		-0.39		0.24	
$NEG\_POSTS_{i,t-1}$	0.0009		0.0107		0.0063		0.0041	
	0.06		0.83		0.56		0.35	
$MIX\_POSTS_{i,t-1}$	-0.0070		0.0056		0.0074		0.0081	
	-0.33		0.31		0.48		0.54	
$AVGLENGTH_{i,t-1}$	0.0004		-0.0013		-0.0006		-0.0020	
	0.21		-0.98		-0.53		-1.79	
$EPISODE_{it}$	0.3745	***	-0.2363	***	-0.1963	***	-0.1920	***
	-2.91		-2.79		-3.37		-3.99	
N	109		138		168		195	
R^2	0.45		0.32		0.26		0.18	
F Test: All Coefficients = 0	7.07		5.95		6.15		4.67	
Pr > F	0.00		0.00		0.00		0.00	

*** = p<.01
** = p<.05
* =p<.10

$$ENTROPY_{i,t-1} = \alpha_1 \cdot RATING_{i,t-1} + \alpha_2 \cdot EPISODE_{it} + \varsigma_i + \varphi_{it} \tag{6.3}$$

It is important to note the difference between these models and (5.1). In (6.2) and (6.3), we are regressing the WOM variables on ratings of the show from the *prior* week, which are included alongside the WOM variables in (5.1). In (5.1), we are regressing *future* ratings on the WOM variables. Note also that this is different from the straight pairwise correlations presented in Table 5. The results of these estimations are partial correlations, which hold constant the other variables, including the show fixed effects. This is not true of the pairwise correlations. The results for a fixed effects regression with τ=4 are presented in Table 12.

These results show a striking asymmetry between the positive volume model, on one hand, and the dispersion model, on the other. In the former, we see that higher ratings for an episode are associated with more positive WOM (the coefficient on $RATING_{i,t-1}$ is positive and significant in the POS_POSTS_{it} model). This is consistent with our belief that a simple volume measure – at least in the current context – captures information about past sales. Conditional on the

Table 12

Residual Information Analysis, τ=4 (t-statistics beneath)		
Dependent Variable:	**ENTROPY$_{i,t-1}$**	**POSPOSTS$_{i,t-1}$**
RATING$_{i,t-1}$	-0.0086	1.8969 **
	-0.19	2.16
EPISODE$_{it}$	-0.1822 ***	-0.4125
	-3.18	-0.38
N	109	109
R^2	0.15	0.09
F Test: All Coefficients = 0	5.97	3.42
Pr > F	0.00 ***	0.04 **
*** = p<.01		
** = p<.05		
* =p<.10		

manager already knowing this, however, the measure may not be at all informative. This is not true of dispersion. It is less intuitively clear why the dispersion of conversations would necessarily be either higher or lower as a product's sales grow. This offers a partial explanation for the difference in the informativeness of dispersion and volume: *dispersion seems to offer more incremental information than does volume of positive posts.*

6.3 Additional Explanations

There are at least three additional reasons why we may not find a consistent relationship between the volume of WOM and future sales. First, as shown in Table 5, we see that the pairwise correlation between $POST_{it}$ and $ENTROPY_{it}$ is non-negligible. This collinearity could partially explain our null result. Second, it might be the case that we haven't captured the exact form of the relationship with our simple linear model. Finally, it may simply be true that there exists no systematic relationship between these quantities in this particular context. Perhaps the linear form we specify isn't quite rich enough. Note that we have estimated models with the more obvious non-linear transformations of $POST_{it}$ including logs and quadratic forms. None of these have yielded qualitatively different results. Additional future research is required to discriminate among these, and potentially other, explanations.

7 Discussion and Conclusion

The objective of this paper has been to investigate the measurement of WOM communications. We've addressed the measurement issue from three perspectives: data collection, construct decomposition and dynamics. Each of these represents a potentially significant contribution to managerial practice. The existence of a vast, publicly accessible reservoir of observable person-to-person communications is unprecedented. Our analysis demonstrates that there is meaningful information in these communications and that this information can be accessed at minimal cost. Compared with the more costly survey-based methods typically employed, this data source is significantly more efficient. We have also specified a dimension of WOM that is critical for the manager to measure: dispersion. Regardless of the source of WOM data being accessed, simple counts are clearly not sufficient. There is valuable information to be gleaned from the extent to which the conversations are taking place across heterogeneous communities as opposed to simply within them. Finally, we have highlighted for managers that a WOM measurement strategy should be enacted early in a product's life cycle.

Throughout the body of paper, we have been careful to discuss the "relationship" between WOM and future sales and to avoid any suggestions of causality. This is in keeping with the methods employed; it is, of course, very difficult to draw clean inferences of causality with traditional econometrics. Nonetheless, it would seem that our results are suggestive of potential causal implications as well. In particular, they suggest that firms interested in adopting "buzz management" – the proactive creation of WOM – as an element of their promotional mix should recognize that, all else equal, more-dispersed buzz may be better than concentrated buzz. This raises several interesting managerial issues worthy of future research. First and foremost, more work is needed to identify the causal link between WOM and future sales. In particular, the differential links between volume and dispersion, on one hand, and sales, on the other, should be investigated. Moreover, assuming that this link exists, the operationalization of dispersion in a managerial context is an interesting question. While the structure of online communities offered us a convenient framework for thinking about dispersion, the off-line world is unlikely to offer such low-hanging fruit. Future research is needed to develop a more-generally implementable basis for the calculation of dispersion.

This leads to another important issue in terms of the management of WOM: the relationship between the on-line and the off-line worlds. In this paper, we investigated the usefulness of information contained in online communities for recovering the underlying sales process occurring offline. This suggests that (a) people make off-line decisions based on on-line information, and/or that (b) online conversations may be a proxy for off-line conversations. While (a) is not surprising, the suggestion that the impact of WOM crosses "worlds" would imply that the manager has the option of creating on-line WOM – for example, through newsgroups and/or websites – or off-line

WOM. Future research to understand better the relationships between WOM and sales across these worlds would be of great value. A more general analysis of the implications of (b) would also be of great value: to what extent is online WOM "similar to" offline WOM? How do they differ? This understanding would help, for example, to create WOM strategies and to drive data collection decisions.

The present study raises several important ethical issues. Most saliently, consumers' decisions to participate in online communities is undoubtedly made without the consideration that firms may be observing these conversations and drawing inferences from them. This differs from traditional market research measurement techniques in which the consumer gives her approval for such use of the data. When one makes the further step to consider the proactive management of WOM, the potential for ethical debate expands further. Is it "right" for the firm to take advantage of the most credible form of information, the personal recommendation? What about actually posing as a consumer and offering recommendations that appear credible but are simply advertising? We offer no answers to these questions here. See Kozinets (2002); King (1996). In particular, the latter argues that one "litmus test" to consider is whether and to what extent the research makes public particularly private information such as the identity of the participants and/or the verbatim of their conversations.

While we believe we have taken an important first step in several directions, we acknowledge nonetheless that our approach is burdened with several limitations. We have focused on a single product category, television shows. While we believe the results to be relatively general, it would be important to replicate these results in other categories. More interesting, it would be important to identify the underlying category factors that make, say, dispersion more important than volume or, say, the decline in the effect of WOM to be particularly steep. This understanding would have an impact on both measurement and management strategies. Econometrically, our approach leaves open the question of sample selection bias. One benefit of the truncated sample approach we focus on is that it minimizes the potential for such a problem since most – though not all – shows survive at least four or five episodes. Our investigation of dynamics which uses all of the data is potentially prone to such a problem. Finally, we have not been able to control for potential important factors in the model. For example, we cannot rule out that at least some of the WOM and/or ratings we observe may be generated due to advertising or positive critical acclaim. To demonstrate causality between WOM and subsequent sales, future research will either need to include advertising data or to control for such exogenous factors in other ways.

References

Anderson, E. W. (1998, August). Customer satisfaction and word of mouth. *Journal of Service Research 1*(1), 5–17.

Anderson, T. W. and C. Hsiao (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econometrics 18*, 47–82.

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies 58*, 277–97.

Banerjee, A. (1992). A simple model of herd behavior. *Quarterly Journal of Economics 110*, 797–817.

Banerjee, A. (1993). The economics of rumours. *Review of Economic Studies 60*, 309–27.

Bass, F. (1969, January). A new product growth model for consumer durables. *Management Science 15*, 215–27.

Bikhchandani, S. D., D. Hirshleifer, and I. Welch (1991). A theory of fads, fashions, custom and cultural change as information cascades. *Journal of Political Economy 100*, 992–1026.

Bowman, D. and D. Narayandas (2001, August). Managing customer-initiated contacts with manufacturers: The impact on share of category requirements and word of mouth behavior. *Journal of Marketing Research 38*, 291–297.

Brown, J. J. and P. Reingen (1987, December). Social ties and word of mouth referral behavior. *Journal of Consumer Research 14*, 350–62.

Coleman, J. S., E. Katz, and H. Menzel (1966). *Medical Innovation: A Diffusion Study*. Indianapolis, Indiana: Bobbs-Merrill.

Forrester Research (2000). *Entertaining Young Net Surfers*. Forrester Research.

Foster, A. and M. Rosenzweig (1995, December). Learning by doing and learning from others: Human capital and technical change in agriculture. *Journal of Political Economy 103*(6), 1176–1210.

Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology 78*(6), 1360–80.

Kaplan, E., P. Cramton, and A. Paltiel (1989). Nonrandom mixing models of hv transmission. *Lecture Notes in Mathematics 83*, 218–241.

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

King, S. A. (1996). Researching internet communities: Proposed ethical guidelines for the reporting of results. *Information Society 12*(2), 119–127.

Kotler, P. (2000). *Marketing Management: The Millenium Edition*. Upper Saddle River, New Jersey: Prentice Hall. p. 560.

Kozinets, R. V. (2002, February). The field behind the screen: Using netnography for marketing research in online communities. *Journal of Marketing Research 39*, 61–72.

Mayzlin, D. (2001). *Promotional Chat on the Internet*. PhD dissertation, Massachusetts Institute of Technology, Sloan School of Management.

Mohr, J. and J. R. Nevin (1990, October). Communication strategies in marketing channels: A theoretical perspective. *Journal of Marketing 54*(4), 36–51.

Nerlove, M. (1967). Experimental evidence on the estimation of dynamic economic relations from a time series of cross sections. *Economic Studies Quarterly 18*, 42–74.

Nerlove, M. (1971). Further evidence on the estimation of dynamic economic relations from a time series of cross sections. *Econometrica 39*, 359–87.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica 49*, 1417–26.

Putsis, W., S. Balasubramanian, E. Kaplan, and S. Sen (1997). Mixing behavior in cross-country diffusion. *Marketing Science 16*, 354–369.

Reingen, P., B. Foster, J. J. Brown, and S. Seidman (1984, January). Brand congruence in interpersonal relations: A social network analysis. *Journal of Consumer Research 23*, 1–26.

Reingen, P. and J. Kernan (1986). Analysis of referral networks in marketing: Methods and illustration. *Journal of Marketing Research 23*, 370–8.

Richins, M. L. (1983). Negative word-of-mouth by dissatisfied consumers: A pilot study. *Journal of Marketing 47*, 68–78.

Van den Bulte, C. and G. Lilien (2001a). *Medical Innovation* revisited: Social contagion versus marketing effort. *American Journal of Sociology 106*(5), 1409–35.

Van den Bulte, C. and G. Lilien (2001b). Two-stage partial observability models of innovation adoption.

Zwillinger, D. (Ed.) (1996). *Standard Mathematical Tables and Formulae* (30th ed.). Boca Raton, Florida: CRC Press.

A Appendix

Table 13

Showname	Type	Network	Runtime (mins)	Number of times aired
Battery Park	Comedy/ Crime	NBC	30	4
Action	Comedy	FOX	30	12
Love & Money	Comedy	CBS	30	3
Get Real	Comedy/ Drama	FOX	60	21
Greed	Game Show	FOX	60	46
Stark Raving Mad	Comedy	NBC	30	21
Once and Again	Drama	ABC	60	26
Work with me	Comedy	CBS	30	4
Caulfield	Drama	FOX	60	2
Movie Stars	Comedy	WB	30	27
Mission Hill	Comedy/	WB	30	2
Malcom in the Middle	Comedy	FOX	30	27
Ladies Man	Comedy	CBS	30	27
City of Angels	Drama	CBS	60	12
Cold Feet	Drama	NBC	60	4
DC	Drama	WB	60	4
Family Law	Drama	CBS	60	24
Freaks and Geeks	Drama/ Comedy	NBC	60	12
God, Devil & Bob	Comedy/	NBC	30	3
WWF Smackdown	Action/	UPN	120	44
Wonderland	Drama	ABC	60	2
West Wing	Drama	NBC	60	32
Judging Amy	Drama	CBS	60	35
Now and Again	Action/ SciFi	CBS	60	25
Odd Man Out	Comedy	ABC	30	13
Oh Grow Up	Comedy	ABC	30	11
The Mike O'Malley Show	Comedy	NBC	30	2
The Parkers	Comedy	UPN	30	43
Popular	Comedy/ Drama	WB	60	44
Roswell	Drama/ SciFi	WB	60	35
Safe Harbor	Drama	WB	60	17
Shasta McNasty	Comedy	UPN	30	26
Snoops	Drama/ Crime	ABC	60	10
Law and Order: Special Victims	Drama/ Crime	NBC	60	34
The Beat	Drama	UPN	60	6
Talk to me	Comedy	ABC	30	3
Then came you	Comedy	ABC	30	6
The others	SciFi	NBC	60	14
The Strip	Drama	UPN	60	16
Third Watch	Drama	NBC	60	32
Time of your life	Drama	FOX	60	13
Angel	Action/ Drama	WB	60	41
Harsh Realm	Drama/ SciFi	FOX	60	3
Grown-Ups	Comedy	UPN	30	43

Table 14

Network	Number of new shows	Min airings	Max airings	Mean airings (per show)
ABC	7	2	26	10.1
CBS	7	3	35	18.6
NBC	10	2	34	15.8
FOX	7	2	46	17.7
UPN	6	6	44	29.7
WB	7	2	44	24.3
Total	44	2	46	18.9

A.1 A thread on Usenet dealing with a WB show "Roswell"

(Note: we have deleted the signatures to shorten the posts, but all else, including the grammar, is unaltered)

From: Spooky Alex (mflulder@mindspring.com)
Subject: OT: Roswell on the WB
Newsgroups: alt.tv.x-files
Date: 1999/10/06
did anyone see this show? it was like a cross between 'dawsons creek' and
'3rd rock from the sun'. so what do you guys think of it?

From: Steven Weller (az941@lafn.org)
Subject: Re: OT: Roswell on the WB
Newsgroups: alt.tv.x-files
Date: 1999/10/07
In another thread, I dubbed it Dawson's Crash, so I think we probably
agree on it.

From: Jeff Burden (jeff922@aol.com.x-files)
Subject: Re: OT: Roswell on the WB
Newsgroups: alt.tv.x-files
Date: 1999/10/07
I watching Dawsons Creek (yes I like, so :P) and saw the previews for it, it
looked good, but I was to geeked to watch "The West Wing" which was very good
tonight.

From: Phil R. (TwoCentsWorth_@webtv.net)
Subject: Re: OT: Roswell on the WB
Newsgroups: alt.tv.x-files
Date: 1999/10/07

What a great little show. I was so impressed. The acting, the settings
and the characters were all great. And the music was fitting as well. I
loved the song 'Crash' by The Dave Mathews Band at the end of the show.
I will definitely be catching this one weekly. I like the blonde alien
<raising eyebrows up and down>.
Phil

From: C. Morgan (mako1@herald.infi.net)
Subject: Re: OT: Roswell on the WB
Newsgroups: alt.tv.x-files
Date: 1999/10/08

I forgot to mention the music! I liked the use of Crash and the Garbage song, but especially Sarah McLachlan's "Fear." That was extremely effective.

Connie

A.2 Arellano and Bond Estimation

As shown first by Nerlove (1967, 1971) and Nickell (1981), when the panel has relatively few time periods per individual (or show, in this case), the specifications presented here may suffer from finite sample bias. Since then, there has been a great deal of work addressing the question of how one might go about obtaining consistent estimates of these equations. Most of the approaches have focused on the use of instrumental variables GMM estimation. Beginning with Anderson and Hsiao (1982), this approach has achieved relative success at the expense of restrictions that one must place on the data. The approach taken by Anderson and Hsiao (1982) is based on the assumption that all of the regressors are strictly exogenous: an assumption that may not be reasonable in our case. A particularly attractive approach – in the sense that the restrictions are relatively benign here – is offered by Arellano and Bond (1991). They instead make the assumption that the errors are not serially correlated. So, in terms of 5.1, for example, they assume that:

$$E\left[\varepsilon_{it} \cdot \varepsilon_{i,t-1}\right] = 0 \tag{A.1}$$

Moreover, they provide a test of this assumption which we will utilize as well. With this assumption, they do not require strict exogeneity of the regressors. Instead, they allow some of the regressors to be "predetermined" and others to be exogenous. Essentially, the method uses as instruments the levels of the ratings lagged by two or more periods. Since the method requires at least 3 periods of data, we are forced to drop the four shows in the sample that were canceled after two episodes. This leaves us with 37 shows in the sample.

We estimate Equation (5.1) with valence posts data as is discussed in Section 6.1. The results of the estimation are below in Table 15. Note that, in each of these estimations, each of the volume variables is assumed to be predetermined. In summary, we find that the results are qualitatively identical to the fixed effects estimates. In addition, we find no evidence that A.1 is violated.

Table 15

Arellano & Bond (z- statistics beneath)	$\tau = 4$	$\tau = 5$	$\tau = 6$	$\tau = 7$
$RATING_{i,t-1}$	-0.4342 **	-0.4795 **	-0.4861 **	-0.3817 ***
	-2.00	-2.22	-2.48	-2.58
$ENTROPY_{i,t-1}$	0.4575	0.3350 *	0.3312 **	0.2253
	1.47	1.92	2.18	1.12
$POS\_POSTS_{i,t-1}$	0.0005	-0.0032	-0.0015	0.0042
	0.05	-0.43	-0.27	0.71
$NEG\_POSTS_{i,t-1}$	0.0017	0.0140	0.0128	0.0097
	0.12	0.94	1.07	1.02
$MIX\_POSTS_{i,t-1}$	-0.0075	0.0027	0.0133	0.0115
	-0.36	0.18	1.01	0.68
$EPISODE_{it}$	-0.4024 **	-0.3094 **	-0.2683 ***	-0.2506 ***
	-2.06	-2.29	-2.98	-3.01
N	69	98	128	155
Sargan Test p	0.99	0.95	0.96	1.00
Serial Correlation p	NA	0.14	0.21	0.27
F Test: All Coefficients = 0	14.46 **	11.62 *	14.16 **	65.10 ***
*** = p<.01				
** = p<.05				
* =p<.10				