MARKETING SCIENCE
Vol. 23, No. 4, Fall 2004, pp. 545–560
ISSN 0732-2399 | EISSN 1526-548X | 04 | 2304 | 0545

DOI 10.1287/mksc.1040.0071

Using Online Conversations to Study Word-of-Mouth Communication

David Godes
Graduate School of Business Administration, Harvard University, Soldiers Field, Boston, Massachusetts 02163,
dgodes@hbs.edu

Dina Mayzlin
School of Management, Yale University, New Haven, Connecticut 06520,
dina.mayzlin@yale.edu

Managers are very interested in word-of-mouth communication because they believe that a 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. First, how does one gather the data? Because the information is exchanged in private conversations, direct observation traditionally has been difficult. Second, what aspect of these conversations 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 an outcome of past sales. Our primary objective is to address these challenges. As a context for our study, we have chosen new television (TV) shows during the 1999–2000 seasons. Our source of word-of-mouth conversations is Usenet, a collection of thousands of newsgroups with diverse topics. We find that online conversations may offer an easy and cost-effective opportunity to measure word of mouth. We show that a measure of the dispersion of conversations across communities has explanatory power in a dynamic model of TV ratings.

Key words: word of mouth; diffusion of innovations; measurement; networks and marketing; new product research; Internet marketing
History: This paper was received June 3, 2002, and was with the authors 13 months for 3 revisions; processed by Gary Lilien.

1. Introduction

Among the many and varied channels through which a person may receive information, it is hard to imagine any that carry the credibility and, thus, the importance of interpersonal communication, or word of mouth (WOM). There is little debate as to whether WOM matters to the firm. There is good reason to believe that it has more potential impact than any other communication channel. Katz and Lazarsfeld (1955) showed that WOM was the most important source of information for certain household items. More recently, a study by Jupiter Communications (1999) found that 57% of people visiting a new Web site did so based on a personal recommendation; this is higher than any other source of influence. As these studies suggest, managers are interested in WOM because it is often an important driver of consumer behavior such as the adoption of a new technology, the decision to watch a TV show, or the choice of which laptop to purchase. It might affect awareness in some cases, or preferences in others. Alternatively, WOM may simply serve as a leading indicator of a product's success. Whatever the specific mechanism, there is empirical evidence, as well as an intuitive justification, for the hypothesized link between WOM and consumer behavior.

One implication of this relationship is that the firm should measure WOM. As a leading indicator, WOM measurement would be important for market research. As a driver of behavior, WOM measurement would be a prerequisite to effective "buzz management." To paraphrase Edward Deming, "You can't manage what you can't measure." However, there are at least three challenges associated with measuring WOM. First, how does one gather the data? Because the information is exchanged in private conversations, direct observation has traditionally been difficult. As a result, most marketers and researchers either have relied on consumer recall or have inferred the process of information exchange from aggregate data. An important implication of the rise of online communities is the enablings of observation of consumer-to-consumer conversations. Here 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? How does one measure a set of statements between people?

Which of the infinite transformations of a conversation are meaningful and managerially useful? The most common approach is to use simple counts. This approach is similar to news-clipping services that monitor how many times a firm's products are mentioned. We investigate the informativeness of this naïve measure. We also investigate another dimension of WOM: dispersion. We define this construct as the extent to which product-related conversations 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 broadly dispersed WOM.

The third challenge comes from the fact that WOM is not exogenous. While the mapping from WOM to future behavior is of interest to the firm, we must recognize that WOM is also an outcome of past behavior. This has important implications for the measurement of WOM. For example, high WOM today does not necessarily mean higher sales tomorrow. It may just mean that the firm had high sales yesterday. Thus, to understand the nature of the link, we need to understand the dynamic relationship between consumer behavior and WOM. Further, we must allow for the fact that the role and the impact of WOM may change over a product's life.

Our primary objective in this paper is to address these challenges. In so doing, we will evaluate the informativeness of two measures—volume and dispersion—to the manager. Specifically, we envision a manager attempting to learn from aggregate data the underlying process governing her customers' behavior. If she had the opportunity to measure WOM, we offer unique insight into which aspects of it she should measure. Given this focus, we are seeking measures that are practical to implement at reasonable cost. We make no claim concerning the optimality of the investigated measures. Another objective of the paper is to investigate the usefulness of online conversations in the study of WOM. The context we study is characterized by a decision made offline, yet we measure WOM online. 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 online WOM are proxies for overall WOM. Given the operational advantages of measuring WOM online, we hope to spur a significant increase in focus on the Web as a laboratory for WOM research.

As a context for our inquiry, we study the relationship between TV viewership behavior and WOM. Specifically, we analyze the ratings for 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 Internet surfers rely on WOM recommendations to purchase CDs, movies, videos or DVDs, and games (Forrester Research 2000). Note that the decision to view a TV show is made repeatedly. This is interesting because the consumer's purchase experience in period t will affect her decision to talk about it as well as her consumption decision in period $t+1$. Our source of WOM information is Usenet, a collection of thousands of newsgroups with very diverse topics.

The paper proceeds as follows. After reviewing the relevant literature in §2, we discuss our research objectives in §3. In §4, we describe the two sources of data used in the study: Nielsen ratings and Usenet. In §5, we present the main empirical results. We find that higher WOM dispersion is related to higher future ratings. We also find that the impact of dispersion declines over time. This argues for measuring WOM early in a product's life. Surprisingly, we find that volume is not consistently associated with higher future ratings. We discuss this result in §6. One potential explanation for the null result could be the fact that positive and negative volume have offsetting associations with future ratings. Because the valence of the post is unobserved in our main analysis, these effects may cancel each other out. To test this, we collect valence information for a sample of the data. Nonetheless, this more costly analysis does not yield the expected association between the volume of WOM and future ratings. Another explanation might be that there is less additional information from a volume measure—as compared with dispersion—conditional on past ratings. Our three equation estimations (dispersion, volume, and ratings) provides some support for this. We conclude in §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 buyer behavior, (2) the importance of social structure in the flow of WOM, and (3) WOM as an outcome of consumer behavior in the past. In addition, we discuss the traditional approaches that have been used to measure WOM.

2.1. WOM as a Driver

There exists ample theoretical support for the idea that WOM impacts consumers' actions. Banerjee (1992, 1993) presents two models that suggest that people are influenced by others' opinions. In fact, rational agents may ignore their own private information in favor of information inferred from others' actions. This may lead to "herding" in which all agents select the same action, which at times may be suboptimal. A similar context is analyzed by Bikhchandani et al. (1991). An important implication of the latter group's work is that the introduction of new information can cause discontinuous shifts in the actions of the agents. This may explain fads and bubbles. Mayzlin (2004) focuses specifically on WOM online and the potential that it presents for the firm to pose as a consumer and create firm-to-consumer communications that look like consumer-to-consumer communications. She finds that, even when this is possible, rational consumers still pay attention to anonymous online posts. As a result, posing as a customer online 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 WOM, 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. Those who lived together had more congruent brand preferences than those who did not. Presumably, living together provides for more opportunities for interaction and communication. Of course, because of the nature of the study, the authors cannot definitively rule out an alternative explanation that women with similar tastes choose to live together. A similar study, in a 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 HYVs was higher as the adoption rate of the village increased. They interpret this as a learning spillover. 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 small positive effect on the farmers' adoption rate of the new HYVs.

Van den Bulte and Lilien (2001) question the primacy of WOM communication as a driver of product adoption. They revisit the Coleman et al. (1966) analysis, arguing that the latter erred in concluding that social contagion drove the physicians' adoption of tetracycline. By adding the information available to the physicians, the authors show that marketing effort was the dominating factor. In Van den Bulte and Lilien (2003), the same authors decompose the adoption process into an awareness phase and an evaluation and adoption phase. In this model, they find evidence of social contagion.

2.2. The Impact of Social Structure

While there are many reasons to believe that WOM is often important in driving consumer actions, it is less clear which aspects of WOM are especially important. Existing literature has demonstrated that not all WOM is created equal. WOM's impact depends on who is talking to whom. Granovetter (1973) characterizes relationships as being either strong ties or weak ties. 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. 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. 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 is likely to reach more people. This has the important implication that information moves quickly within communities but slowly across them.

In a similar vein, the work by Kaplan et al. (1989) in mathematical bioscience shows that different patterns of contact between groups with different incidences of HIV/AIDS have different impacts on the spread of the disease. This modeling approach has been utilized in the marketing literature by Putsis et al. (1997). They find heterogeneity in mixing behavior across 10 nations. Importantly for the present study, they find greater interaction within the population of a country than between populations of different countries.

2.3. WOM as an Outcome

Part of the difficulty in measuring WOM is the fact that it is a precursor as well as an outcome of consumer actions. Numerous papers provide evidence of the latter point. Richins (1983) looks at the moderating factors that determine whether one talks about her negative experience. Anderson (1998) looks at negative and positive WOM communication. 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 customer satisfaction data.

Bowman and Narayandas (2001) investigate the firm's disposition of customer-initiated contacts (CICs). Two outcomes of this process are market share and WOM behavior. Bowman and Narayandas measure WOM via a survey, capturing both the incidence of WOM and the breadth of referral. 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 who described themselves as loyal 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 typically has been analyzed using two methodologies: inference or surveys, or both. Examples of the former include Foster and Rosenzweig (1995) in which the farmers 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 who do not. The presumption is that those who live in closer proximity are more likely to exchange information. Finally, Bass (1969) and those who 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), and Richins (1983) all base their analyses on proprietary surveys designed to test a specific hypothesis. Van den Bulte and Lilien (2001, 2003) and Anderson (1998) draw on the existence of survey-based data that were prepared for other purposes. The attraction of surveys in this context is that one can directly ask, "Did you tell somebody about X?" In some cases, such as Bowman and Narayandas (2001), one might even ask, "How many people did you tell?" Additionally, some researchers use surveys to map out social networks. For example, Reingen and Kernan (1986) use surveys to map out the entire social network comprising a piano tuner's customers. With this, they were able to understand which people were important in the referral process. Brown and Reingen (1987) used a similar methodology for piano teachers. Similarly, the dataset used by Van den Bulte and Lilien (2001) contains data for each physician about the other physicians with whom he or she discussed medical practices.

One purpose of this paper is to offer an alternative method to measure WOM. Online conversations offer the firm an attractive opportunity to learn about its environment by directly observing the flow of interpersonal communication.[1] By looking at activity across different online communities, we are able to infer 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. So, for example, we are not able to identify loyal users, as Bowman and Narayandas (2001) do.

3. Research Objectives

Our goal 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 (2001, 2003), the Yahoo! Buzz Index (http://buzz.yahoo.com), and others. The more conversations there are about this paper, for example, the more people will become informed about it. Because awareness is a necessary condition for viewing a TV show, we expect that higher volumes of WOM will be associated with higher future ratings.

As Mohr and Nevin (1990) do in interfirm communication, we investigate two distinct dimensions of interpersonal communications. Using assumptions similar to those 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 interact frequently with each other and thus are more likely to learn from each other than from members of other communities. Thus, conditional on a certain volume of WOM, more people will become informed about a new TV show the more dispersed this information is between communities. This motivates us to explore the relationship between WOM dispersion and future ratings. We expect that this relationship will be positive.

Finally, we explore the dynamics in the relationship between WOM and ratings. We want to understand not only which aspects of WOM are informative but also when the informativeness is particularly high. This is managerially important because it affects the timing of investment in information gathering and in influencing the flow of information. We expect that the magnitude of the effect of dispersion and volume of WOM on future ratings will decrease over time. This is because as people become better informed about their preferences for different shows, a recommendation is less likely to impact decisions.

Three comments are in order concerning 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

[1] A number of marketing researchers have recently identified the Web as an interesting and valuable research context. For example, Kozinets (2002) investigates the Web as a source of ethnographic data, Danaher et al. (2003) compare online and offline brand loyalty, Park and Fader (2004) and Chatterjee et al. (2003) model online browsing and click behavior. Other researchers address the impact of new online institutions on competition. See, for example, Iyer and Pazgal (2003) and Chen et al. (2002).

talking about the show? We hypothesize that a measure such as reach would also be useful: How many different people are talking about it?

Second, note that both of these measures ignore the potentially valuable content of the conversations. In particular, the volume of WOM may have a very different effect depending on the valence of comments. The downside of collecting these data is that doing so is a costly and noisy process, as we demonstrate in §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 explore the informativeness of these measures conditional on past ratings. We expect a lot of variance in current ratings to be explained by past ratings. Because past ratings drive current WOM activity and the manager observes the ratings, 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 U.S. market during the 1999–2000 season by combining two publicly available datasets. For viewership data, we use Nielsen ratings (reported weekly in *Broadcasting & Cable* magazine), and for WOM we use Usenet newsgroup conversations.

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 two episodes each. Half of the shows were shown fewer than 17 times. 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 1 lists the shows and Table 2 summarizes the data by network.

Figure 1 **Distribution of New Shows by Episodes Aired**

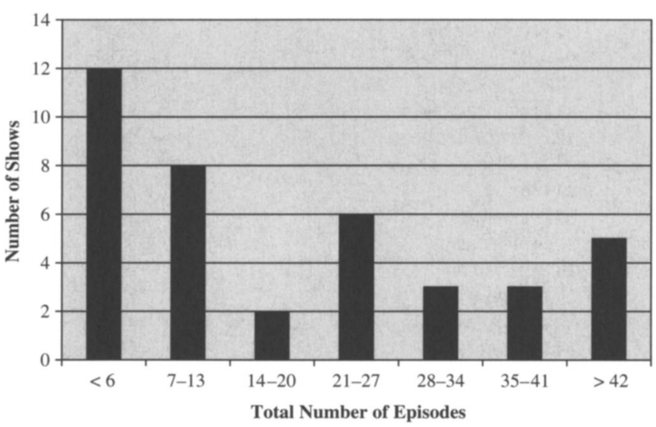


Table 1 **Shows in the Sample**

Show name	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, the Devil, and 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 2 **Summary of Shows by Network**

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

Table 3 Five Highest-Rated 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	Tues	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 4 Five Lowest-Rated 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	Tues	9/21/1999	1.8	1.8
The Beat	UPN	Tues	3/21/2000	2.2	2.2
The Strip	UPN	Tues	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 3 and 4 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 midseason 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 is currently publicly available at http://groups.google.com.[2] In comparison to the social network mapping procedures, this dataset offers an easy and affordable alternative. Moreover, Usenet covers a wide breadth of topics, from rec.autos.sport.nascar to alt.fan.noam-chomsky. Thus, this appears to be a fertile area for managerial and academic research on WOM. These benefits do not come without costs: There is a potential for bias at two levels. First, online conversations may not be representative of all conversations. Moreover, the subset of Usenet conversations may not be a representative sample of all online conversations. However, these potential biases would, if anything, decrease the estimated relationship between WOM and future ratings.

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, and so on. Posts are organized into threads that contain posts on the same topic. One might think of a thread as an analog of a conversation. Often, all posts in a thread contain the same subject line. For an example of a partial thread, see appendix.

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 a fair number of posts about shows which do not include the show's name in the subject line. We found 169 groups that contained messages about the shows in our sample. The groups' focuses range from TV in general (rec.arts.tv) to specific shows (alt.tv.x-files, which is devoted to *The X-Files*.). Those who visit alt.tv.x-files often chat about other shows that they find interesting. The appendix presents a thread about the show *Roswell* that takes place in alt.tv.x-files. This is not particularly surprising because both are science fiction shows. It takes time for fans to assemble a newsgroup devoted to a new show such as *Roswell*. In the initial period following the show's debut, the conversations are dispersed among groups that are devoted to other shows. Table 5 presents the 20 newsgroups that had the most postings about the shows in our sample. None of these groups is specifically devoted to any show in the sample.

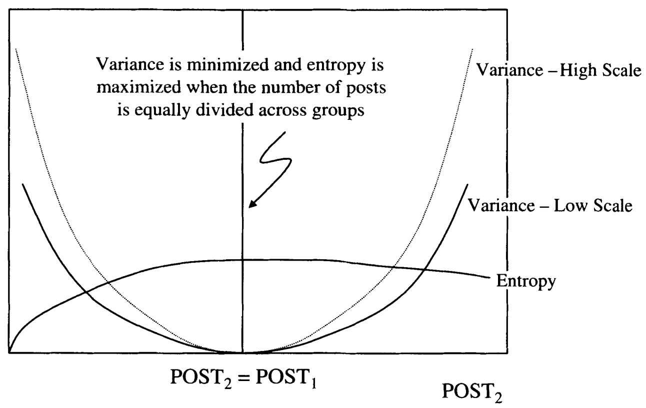
We excluded three of the 44 shows from the sample: *Angel*, *Harsh Realm*, and *Grownups*. We exclude *Angel* because we found too many posts—more than 3,000—that contained the word "angel" in the subject line. On the one hand, from a simple reading of the entire subject line alone, it was clear that most of the posts were unrelated to the show. On the other hand, there

Table 5 Twenty 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
alt.tv.ally-mcbeal	144

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

Figure 2 A Comparison of Entropy and Variance



were no posts with the words "grownups" or "harsh realm" in the subject line. This demonstrates that our technique for extracting posts is imperfect. This is especially so when shows' names contain common words such as "angel" or involve shows that generate little buzz. We emphasize again that, for our main analysis, we do not analyze the post's content. We revisit this issue below in §6.

4.3. Variables

From these conversations, we construct volume and dispersion measures as discussed in §3. Let $n = 1, \ldots, N$ index the newsgroups. We define $POST^n_{it}$ as the number of posts in newsgroup n about show i between episodes t and $t+1$. So, the volume of WOM is

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

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

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

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 the groups in which there is at least one post. Figure 2 presents a comparison of variance and entropy.[3]

[3] 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 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.

Table 6 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

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

$$NUMGROUPS_{it} = \sum_{n=1}^{N} 1(POST^n_{it} > 0). \tag{4.3}$$

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 in the show's life. The results that we present below do not control for these factors. We have estimated alternative specifications that control for the effect of interepisode length on the WOM measures. Because 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, although this is a relatively rare occurrence. In this scenario, it would seem that ratings of the second show that day are driven by a different process. Hence, we use the ratings from the first episode that day and exclude the second. The results, however, are nearly identical if we do not exclude these episodes.

Our dependent variable is $RATING_{it}$, the rating of episode t for show i. To control for a time trend in ratings, we include a time variable $EPISODE_{it} \equiv t$. Finally, we define the *early* period to be the first τ episodes of a show. That is, we define a dummy variable,

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

We estimate our models across a range of τ values.

Table 6 provides summary statistics for the variables used and Table 7 provides pairwise correlations.

5. Main Results

There are (at least) two ways to investigate the role of WOM early in a show's life. One approach is to truncate the dataset to only the early episodes. Another approach is to use all the data but estimate

Table 7 Correlation Matrix

	$RATING_t$	$RATING_{t-1}$	$POST_{t-1}$	$ENTROPY_{t-1}$	$NUMGROUPS_{t-1}$	$EPISODE$
$RATING_t$	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

separate coefficients for the early and late observations. The advantage of the truncated approach is that it is conceptually appealing. It matches the context faced by the manager: After, say, five episodes, she wants to understand how good her show is. 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 the data.

5.1. Model with Early Data Only

We estimate the following model:

$$\begin{aligned} RATING_{it} &= \lambda \cdot RATING_{i,t-1} + \pi \cdot POST_{i,t-1} \\ &\quad + \delta \cdot ENTROPY_{i,t-1} + \beta \cdot EPISODE_{it} \\ &\quad + u_i + \varepsilon_{it} \quad \text{for } t \leq \tau. \end{aligned} \tag{5.1}$$

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.[4, 5]

The estimation of (5.1) is presented in columns (1), (3), (5), and (7) of Table 8.[6] More dispersed WOM is associated with higher future ratings early in the show's life.[7] The coefficient on $ENTROPY_{i,t-1}$ is positive and significant at the $p < 0.05$ level when $\tau = 4$ and when $\tau = 5$ and at the $p < 0.10$ level when $\tau = 6$. Thus, it seems that more dispersed early conversations are associated with higher future ratings. Returning to the advertising analogy, the reach of WOM appears to be significantly related to the TV show's next week's ratings. Higher entropy implies that information about the show—its existence, its premise, its potential quality—is finding its way into a more diverse set of communities. In this sense, it is likely that more uninformed people become informed by the WOM the higher the entropy. Rather than the same people reading more posts about the show, the word is being spread across communities. This is analogous to traditional advertising where the campaign reach is considered an important driver of its effectiveness.

To illustrate the magnitude of this effect, consider a show that has 15 posts in one newsgroup and 5 posts in 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 would yield an entropy of 0.693 and would be associated with an 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 less support for the effect of volume. The coefficient on $POST_{i,t-1}$ reaches only marginal significance when $\tau = 7$. Still, both measures appear to have explanatory power in the specification and thus warrant further investigation. Nonetheless, a strategy of counting WOM appears to be less informative than also modeling and measuring the spread of WOM across communities.[8]

[4] A random effects model would be preferable but show quality, we would expect, is correlated with $RATING_{i,t-1}$. A specification test confirmed this.

[5] 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). In our sample, the bias is not expected to be substantial since the number of observations per show is not very low (mean = 15). Arellano and Bond (1991) offer a GMM-based method as a solution to this problem. We estimate a model according to this method and find qualitatively equivalent results. Details are available from the authors.

[6] In this paper, we calculate the R^2 statistic for the differenced model. That is, the R^2 we report estimates the percentage of variance explained by the model beyond the show fixed effects.

[7] Note that our analysis occurs at the post level, not at the poster level. Thus, we do not capture the fact that posters may participate in several newsgroups. Moreover, the interpretation of the results may differ depending on the extent to which entropy is caused by different people in different communities or by the same people participating across communities. We implicitly assume the former but one should consider that the latter could also be at work. We thank an anonymous referee for pointing this out.

[8] Because the analysis here is somewhat exploratory, we estimated several variants of this specification to test the robustness of the main results. We estimated the equation taking logs of the right-hand side (RHS) variables to capture possible decreasing marginal returns. We also estimated a model that included a nonlinear episode variable. None of these estimations yielded significantly different results from the ones presented.

Table 8 Estimation Results: Truncated Sample Fixed-Effects Model

	$\tau=4$		$\tau=5$		$\tau=6$		$\tau=7$	
	(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

Note. t-statistics are shown beneath parameter estimates.
$* = p < 0.10$.
$** = p < 0.05$.
$*** = p < 0.01$.

Because the dispersion measure captures both the number of communities and distribution of conversations across communities, we explore to what extent the number of groups can explain our results: We include the number of groups in addition to the entropy variable. The results of these regressions are presented in columns 2, 4, 6, and 8 of Table 8. The key difference between $NUMGROUPS_{it}$ and $ENTROPY_{it}$ is that high values of the former may result when there are several communities but the preponderance of the activity occurs in just a few. As Table 8 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, despite significant correlation (0.88) between $NUMGROUPS_{it}$ and $ENTROPY_{it}$.[9]

Some other interesting results also emerge. We see in Table 8 that the coefficient on lagged ratings in the early period is negative and less than one in absolute value. The latter suggests that the impact of a shock to ratings dissipates over time, as expected. The negative coefficient suggests that, early in a show's life, there is oscillation in ratings around the mean. One would not expect this instability to persist as the uncertainty surrounding the show becomes resolved. Indeed, we show below (see Table 9) that the coefficient becomes positive in later periods. We also find 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 here 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 early data. A more detailed exploration is best carried out by estimating the model on all the data.

[9] In a specification in which $NUMGROUPS_{it}$ appears without $ENTROPY_{it}$, the former is never significant at the 0.10 level.

5.2. Late WOM vs. Early WOM

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.[10]
The model we estimate is

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

[10] Note that the specification in (5.2) is equivalent to one which specifies the effect of volume as

$$\begin{aligned}
&\theta_1 \cdot POST_{i,t-1} + \theta_2 \cdot EARLY_{it} \times POST_{i,t-1} \\
&\quad = \theta_1 \cdot [POST_{i,t-1} \times EARLY_{it} + POST_{i,t-1} \times (1 - EARLY_{it})] \\
&\qquad + \theta_2 \cdot EARLY_{it} \times POST_{i,t-1} \\
&\quad = (\theta_1 + \theta_2) \cdot POST_{i,t-1} \times EARLY_{it} + \theta_1 \cdot POST_{i,t-1} \times (1 - EARLY_{it}).
\end{aligned}$$

Table 9 Estimation Results: Full Sample Fixed-Effects Model

	$\tau=4$	$\tau=5$	$\tau=6$	$\tau=7$
$\text{RATING}_{i,t-1} \times \text{EARLY}_{it}$	−0.5484***	−0.4607***	−0.4234***	−0.2997***
	−5.55	−5.37	−5.47	−4.14
$\text{RATING}_{i,t-1} \times (1-\text{EARLY}_{it})$	0.2068***	0.2135***	0.2386***	0.2315***
	4.87	4.98	5.37	5.02
$\text{POST}_{i,t-1} \times \text{EARLY}_{it}$	0.0027	0.0031	0.0043	0.0051*
	0.62	0.80	1.37	1.69
$\text{POST}_{i,t-1} \times (1-\text{EARLY}_{it})$	−0.0012	−0.0012	−0.0019	−0.0020
	−0.75	−0.78	−1.16	−1.20
$\text{ENTROPY}_{i,t-1} \times \text{EARLY}_{it}$	0.5769**	0.3819*	0.2975*	0.2063
	2.10	1.90	1.73	1.26
$\text{ENTROPY}_{i,t-1} \times (1-\text{EARLY}_{it})$	0.0081	−0.0600	−0.0204	0.0416
	0.09	−0.63	−0.20	0.40
$\text{EPISODE}_{i,t-1} \times \text{EARLY}_{it}$	−0.3445**	−0.2329***	−0.1869***	−0.1636***
	−2.56	−2.68	−3.15	−3.43
$\text{EPISODE}_{i,t-1} \times (1-\text{EARLY}_{it})$	−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
$\Pr > F$	0.000	0.000	0.000	0.000

Note. t-statistics are shown beneath parameter estimates.
$^{*} = p < 0.10$.
$^{**} = p < 0.05$.
$^{***} = p < 0.01$.

The specification above essentially replicates (5.1) for the two periods because all variables (including the fixed effects) are allowed to vary over the early and late periods. Indeed, as Table 9 demonstrates, the coefficients on the variables interacted with $EARLY_{it}$ are numerically identical to the coefficients in Table 8. The only difference is that the t-statistics are slightly lower in Table 9 because the variance-covariance matrix is reestimated using all data in the model. We again find support for the idea that this association is strong early 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 ratings later in the show's life. On the contrary, given the dynamic nature of the process, dispersion in the fifth period, for example, is associated with ratings in the sixth period, which drive ratings in the seventh period, and so on. Thus, dispersion is likely to have a lasting indirect association with future ratings even though the direct association seems to wane.[11] Again, we find less support for an association between the volume of WOM and ratings (early WOM is significant at the 0.10 level only when $\tau=7$).

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

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

In (5.3), we interact entropy with a decreasing function of the time trend: $\exp(-r \cdot EPISODE_{i,t})$. This allows us to continuously vary the effects of entropy over time. An increase in r implies that the impact of entropy changes at a faster rate. For example, for Episode 3 $\exp(-r \cdot EPISODE_{it}) = 0.687$ when $r = 0.125$ and $\exp(-r \cdot EPISODE_{i,t}) = 0.050$ when $r = 1$. This convex decline is important to capture because the results in Tables 8 and 9 indicate that the effect of dispersion declines quickly. The estimation results of (5.3) are presented in Table 10. The coefficient on $ENTROPY_{i,t-1}(\delta_1)$ is not significant, while the coefficient on the interaction term (θ) is significant for $r = 0.05$, 0.125, 0.25, 0.50, and 0.75. This is 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 $\left(\frac{\partial RATING_{it}}{\partial ENTROPY_{i,t-1}} = \delta_1 + \theta \exp(-0.125 \cdot EPISODE_{it})\right)$ is 0.232 for Episode 3 and 0.192 for Episode 4. Note that the velocity of decline (r) makes a difference. The R^2 seems to peak at

[11] We thank an anonymous referee for pointing this out.

Table 10 Estimation Results: Full Sample with Episode-Entropy Interaction

	$r = 0.05$	$r = 0.125$	$r = 0.25$	$r = 0.50$	$r = 0.75$	$r = 1$
$RATING_{i,t-1}$	0.1497***	0.1454***	0.1415***	0.1413***	0.1439***	0.1465***
	3.95	3.82	3.70	3.67	3.73	3.80
$POST_{i,t-1}$	−0.0005	−0.0006	−0.0008	−0.0009	−0.0009	−0.0009
	−0.38	−0.44	−0.55	−0.64	−0.65	−0.64
$ENTROPY_{i,t-1}$	−0.2312	−0.1045	−0.0528	−0.0132	0.0062	0.0174
	−1.35	−0.94	−0.55	−0.15	0.07	0.20
$\exp(-r\ EPISODE) \times ENTROPY_{i,t-1}$	0.4957*	0.4891**	0.6287**	0.9436**	1.3119*	1.8059
	1.88	2.13	2.29	2.08	1.74	1.45
$EPISODE_{it}$	−0.0217***	−0.0221***	−0.0230***	−0.0244***	−0.0252***	−0.0256***
	−4.18	−4.46	−4.86	−5.31	−5.54	−5.66
N	688	688	688	688	688	688
R^2	0.115	0.116	0.117	0.116	0.114	0.113
F Test: All coefficients = 0	16.61	16.84	17.00	16.79	16.50	16.29
$\Pr > F$	0.00	0.00	0.00	0.00	0.00	0.00

Note. t-statistics are shown beneath parameter estimates.
* $= p < 0.10$.
** $= p < 0.05$.
*** $= p < 0.01$.

around $r = 0.25$. The results are not significant when the assumed decline is too steep (for example, $r = 1$). As before, the effect of volume is not significant.[12] The positive coefficient on $RATING_{i,t-1}$ is not surprising here, in light of Table 9. It is clear that, for the latter part of the shows' lives, viewership has strong persistence.

6. Investigating the Role of Volume

The results in §5 suggest that dispersion is an important aspect of WOM. 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. It may be an artifact of our data collection and analysis. In particular, our focus on cost-effective data collection precluded our adoption of content analysis. This decreases the amount of information in our data. Negative and positive volumes may have offsetting relationships with future ratings that cancel each other out in our estimates. We investigate this in §6.1 by performing content analysis on a sample of the posts. In §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 §6.3, we discuss other possible explanations.

6.1. Valence Data Results

To investigate positive and negative WOM, we collected content data for a sample of posts. Specifically, we sampled 10% of each show's posts each week, rounded up. We employed two independent raters who were unaware of our research objectives. After reading the post, each rater was asked to classify it into one of six categories:

1. Positive
2. Negative
3. Neutral
4. Mixed
5. Irrelevant
6. Not sure

Of the 2,398 posts that were evaluated, 1,356 (57%) received identical categorizations from each rater. Accurate content analysis is extremely difficult because of its subjective nature. To resolve differences, we employed a third rater to evaluate all posts on which the others disagreed. When this third rater agreed with one of the previous two, we used that evaluation. Otherwise, we assigned it to a seventh category of disagreed posts. This yielded 2,023 usable posts. See Table 11 for their distribution. Many (42%) were deemed not relevant to the show under consideration. These were either mistakenly included in our sample because the subject name matched our

Table 11 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	2,275	100%	641	100%

[12] We 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).

Table 12 Correlation Matrix with Valence Data Included

	$POS\_POSTS_{i,t-1}$	$NEG\_POSTS_{i,t-1}$	$MIX\_POSTS_{i,t-1}$	$AVGLENGTH_{i,t-1}$
$POS\_POSTS_{i,t-1}$	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

criteria, or the posts included the name of the show in the subject name but then proceeded to discuss other issues.[13] Of the relevant posts (either positive, negative, or mixed), almost three out of four were either positive or mixed. Moreover, there was nearly twice as much positive WOM as negative WOM.

We define $SAMP\_POS\%_{it}$ as the percentage of sampled posts in period t for show i that were rated as positive and $POS\_POSTS_{it} \equiv SAMP\_POS\%_{it} \cdot POST_{it}$ as the expected number of positive posts about show i in period t in the entire dataset. We similarly define NEG_POSTS_{it}, NEU_POSTS_{it}, MIX_POSTS_{it}, IRR_POSTS_{it}, NS_POSTS_{it}, and DIS_POSTS_{it}. We also measured the number of words in each post. This variable—$AVGLENGTH_{it}$—might indicate either the passion or quality of the post.[14]

We reestimate (5.1) with these new variables. Table 12 presents the pairwise correlation matrix with the new variables included. Table 13 presents the estimation results. First, note that including the valence information does not weaken the inferred relationship between dispersion and ratings. This relationship appears even stronger. This is not surprising because we have eliminated some of the noise associated with irrelevant posts. Moreover, even with valence data, neither the volume of WOM nor the post's length demonstrate a strong relationship with ratings.

6.2. System Estimation

As shown by Richins (1983), Anderson (1998), and Bowman and Narayandas (2001), we know that past ratings are likely to impact current WOM. Thus, it may be that conditional on past ratings current WOM volume data are superfluous. Higher ratings in time t should imply more conversations about the product in time $t+1$, all else equal. However, the same argument may not be true for dispersion. To investigate this, we are interested in the following equations, which we estimate as a system along with (5.1):

$$POST_{i,t-1} = \gamma_1 \cdot RATING_{i,t-1} + \gamma_2 \cdot EPISODE_{it} + \eta_i + \psi_{it}. \quad (6.1)$$

$$ENTROPY_{i,t-1} = \alpha_1 \cdot RATING_{i,t-1} + \alpha_2 \cdot EPISODE_{it} + \varsigma_i + \varphi_{it}. \quad (6.2)$$

Note that these three equations ((5.1), (6.1), and (6.2)) form a triangular system as defined by Lahiri and Schmidt (1978), and therefore, estimation by generalized least squares (GLS) in the manner of the seemingly unrelated regression (SUR) model is consistent (Lahiri and Schmidt 1978).

The results of this estimation are shown in Table 14.[15] Note that while the estimated coefficients on the $RATING_{it}$ equation are unchanged, the standard errors decrease.[16] Qualitatively, the results are similar to the ones in Table 8: $ENTROPY$ is significant early on, while $POST$ becomes significant later on. However, the coefficients on the $ENTROPY$ and $POST$ variables are now more significant. For example, $ENTROPY$ is now significant at the 5% level and $POST$ is significant at the 10% level for $\tau = 6$.

The estimation of (6.1) and (6.2) demonstrates an interesting asymmetry. In the former, we see that higher ratings for an episode are associated with more WOM (the coefficient on $RATING_{i,t-1}$ is positive and significant in the $POST$ model for $\tau > 4$). This is consistent with the view that a simple volume measure captures information about past ratings.

[13] 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 may contribute to the overall impression of a large volume of conversations, which in turn suggests that even these presumably irrelevant posts may have a marginal impact on a potential viewer's decision to sample the show.

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

[15] Because R^2 is not a well-defined concept for GLS, we do not report it in this table.

[16] It is straightforward to show that the equivalence of the two estimates of (5.1) is due to two factors: (1) The explanatory variables in Equations (6.2) and (6.1) are each a subset of those in (5.1), and (2) the dependent variables in Equations (6.2) and (6.1) are explanatory variables in (5.1).

Table 13 Estimation Results: Truncated Sample Fixed-Effects Model with Valence Data

	$\tau = 4$	$\tau = 5$	$\tau = 6$	$\tau = 7$
$RATING_{i,t-1}$	−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

Note. t-statistics are shown beneath parameter estimates.
* = $p < 0.10$.
** = $p < 0.05$.
*** = $p < 0.01$.

Conditional on the manager already knowing this, however, the measure may not be informative. This is not true of dispersion: $RATING_{i,t-1}$ is never significant in the *ENTROPY* model. This is not surprising because it is less clear why the dispersion would necessarily be either higher or lower as a show's ratings 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.

Table 14 Estimation Results: Truncated Sample Seemingly Unrelated Regression Model

	$t = 4$	$t = 5$	$t = 6$	$t = 7$
Dependent variable = $RATING_{i,t}$				
$RATING_{i,t-1}$	−0.5484***	−0.4607***	−0.4234***	−0.2997***
	−8.35	−7.13	−6.89	−4.84
$POST_{i,t-1}$	0.0027	0.0031	0.0043*	0.0051**
	0.93	1.06	1.73	1.97
$ENTROPY_{i,t-1}$	0.5769***	0.3819**	0.2975**	0.2063
	3.16	2.52	2.18	1.47
$EPISODE_{it}$	−0.3445***	−0.2329***	−0.1869***	−0.1636***
	−3.85	−3.56	−3.97	−4.01
Wald Test: all coeffs = 0	89.51***	60.94***	62.55***	42.61***
Dependent variable = $POST_{i,t}$				
$RATING_{i,t-1}$	2.7136	3.4010*	3.7645*	3.3152*
	1.22	1.71	1.88	1.88
$EPISODE_{it}$	−9.3037***	−6.6726***	−4.8619***	−4.6525***
	−3.34	−3.45	−3.27	−4.18
Wald Test: all coeffs = 0	20.02***	26.03***	23.12***	33.66***
Dependent variable = $ENTROPY_{i,t}$				
$RATING_{i,t-1}$	−0.0086	0.0208	0.0990	0.0104
	−0.24	0.55	0.27	0.32
$EPISODE_{it}$	−0.1822***	−0.1023***	−0.0787***	−0.0693***
	−4.09	−2.77	−2.91	−3.35
Wald Test: all coeffs = 0	19.71***	12.01***	11.02***	14.91***
N	109	138	168	195

Note. z-statistics are shown beneath parameter estimates.
* = $p < 0.10$.
** = $p < 0.05$.
*** = $p < 0.01$.

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 ratings. First, as shown in Table 7, we see that the pairwise correlation between $POST_{it}$ and $ENTROPY_{it}$ is nonnegligible. This collinearity could partially explain our null result. Second, we may not have captured the exact form of the relationship with our model. Perhaps the linear form we specify is not quite rich enough. Note that we have estimated models with the obvious nonlinear transformations of $POST_{it}$ including logs and quadratic forms. None of these has yielded qualitatively different results. Finally, no systematic relationship may exist between these quantities in this context. 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 have addressed this issue from three perspectives: data collection, construct decomposition, and dynamics. Each perspective represents a potentially significant contribution to managerial practice. The existence of a publicly accessible reservoir of observable person-to-person communications is unprecedented. Our analysis demonstrates that there is information in these communications and that it can be accessed at minimal cost. Compared with the costly 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, simple counts are not sufficient. There is valuable information in the extent to which the conversations are taking place across heterogeneous communities as opposed to simply within them. Finally, we have highlighted the point that a WOM measurement strategy should be enacted early in a product's life-cycle.

Throughout the paper, we have been careful to discuss the relationship between WOM and future ratings and to avoid any suggestions of causality. This is in keeping with the methods employed; it is very difficult to draw clean inferences of causality with traditional econometrics. Nonetheless, it would seem that our results are also suggestive of causal implications. 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 more-dispersed buzz may be better than concentrated buzz. This raises several interesting managerial issues. 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 question of how to operationalize dispersion is an interesting one. While online communities offered us a convenient framework for thinking about dispersion, the offline 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 online and the offline worlds. In this paper, we investigated the usefulness of online communities in recovering the underlying sales process occurring offline. This suggests that (a) people make offline decisions based on online information, or that (b) online conversations may be a proxy for offline conversations. While (a) is not surprising, the suggestion that the impact of WOM crosses worlds implies that the manager has the option of creating online WOM—for example, through newsgroups or Web sites, or both—or offline WOM. Future research to understand better the relationships between WOM and sales across these worlds would be valuable. A more general analysis of the implications of (b) would also be of great value: To what extent is online WOM similar to or different from offline WOM? This would help, for example, to create WOM strategies and to drive data collection decisions.

This study raises several important ethical issues. 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 approval for use of the data. In the case of proactive management of WOM, the potential for ethical debate expands further. Is it right for the firm to take advantage of personal recommendations? 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 and King 1996.) In particular, the latter argues that one litmus test to consider is whether the research makes public particularly private information such as the identity of the participants or the verbatim of their conversations, or both.

While we have taken an important first step in several directions, we acknowledge that our approach is burdened with several limitations. We have focused on a single product category, TV shows. While we believe the results to be relatively general, it would be important to replicate these results in other categories characterized by different types of consumer actions. The decision to watch a new TV show is a relatively low-cost and low-risk decision. It would be interesting to investigate the role of WOM on the adoption of new technologies or the purchase of higher ticket items, for example.

It would also be important to identify the underlying category factors that make dispersion more important than volume or the decline in the effect of WOM to be particularly steep. This 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 because most—though not all—shows survive at least four or five episodes. Our investigation of dynamics that uses all the data is potentially prone to sample selection bias. 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 or ratings we observe may be generated because of 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.

Acknowledgments

The authors 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. The authors would also like to thank participants in the Massachusetts Institute of Technology Marketing Lunch; the Wiesbaden 2001 Marketing Science Conference; Harvard Business School Marketing Seminar; Yale Marketing Seminar; 2002 New York University/Columbia University/Wharton School/Yale University Consortium; University of California–Berkeley Marketing Seminar; Stanford Institute for Theoretical Economics Summer Conference; and the Young and Rubicam Brand Buzz Summer Lunch Workshop. This research, in the early phases of development, also benefited from conversations with Sendhil Mullainathan. The authors contributed equally, and their names are listed in alphabetical order.

Appendix. A Partial 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.

References

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

Arellano, M., S. Bond. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Rev. Econom. Stud.* **58** 277–297.

Banerjee, A. 1992. A simple model of herd behavior. *Quart. J. Econom.* **110** 797–817.

Banerjee, A. 1993. The economics of rumours. *Rev. Econom. Stud.* **60** 309–327.

Bass, F. 1969. A new product growth model for consumer durables. *Management Sci.* **15** 215–227.

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

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

Brown, J. J., P. Reingen. 1987. Social ties and word-of-mouth referral behavior. *J. Consumer Res.* **14** 350–362.

Chatterjee, P., D. L. Hoffman, T. P. Novak. 2003. Modeling the clickstream: Implications for web-based advertising efforts. *Marketing Sci.* **22**(4) 520–541.

Chen, Y., G. Iyer, V. Padmanabhan. 2002. Referral infomediaries. *Marketing Sci.* **21**(4) 412–434.

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

Danaher, P. J., I. W. Wilson, R. Davis. 2003. A comparison of online and offline consumer brand loyalty. *Marketing Sci.* **22**(4) 461–476.

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

Granovetter, M. 1973. The strength of weak ties. *Amer. J. Sociology* **78**(6) 1360–1380.

Iyer, G., A. Pazgal. 2003. Internet shopping agents: Virtual co-location and competition. *Marketing Sci.* **22**(1) 85–106.

Kaplan, E., P. Cramton, A. Paltiel. 1989. Nonrandom mixing models of hv transmission. *Lecture Notes Math.* **83** 218–241.

Katz, E., P. F. Lazarsfeld. 1955. *Personal Influence.* Free Press, Glencoe, IL.

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

Kozinets, R. V. 2002. The field behind the screen: Using netnography for marketing research in online communities. *J. Marketing Res.* **39** 61–72.

Lahiri, K., P. Schmidt. 1978. On the estimation of triangular structural systems. *Econometrica* **46**(5) 1217–1221.

Mayzlin, D. 2004. *Promotional Chat on the Internet.* Working paper, Yale School of Management, New Haven, CT.

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

Nerlove, M. 1967. Experimental evidence on the estimation of dynamic economic relations from a time series of cross sections. *Econom. Stud. Quart.* **18** 42–74.

Nerlove, M. 1971. Further evidence on the estimation of dynamic economic relations from a time series of cross sections. *Econometrica* **39** 359–387.

Nickell, S. 1981. Biases in dynamic models with fixed effects. *Econometrica* **49** 1417–1426.

Park, Y.-H., P. S. Fader. 2004. Modeling browsing behavior at multiple websites. *Marketing Sci.* **23**(3) 280–303.

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

Reingen, P., J. Kernan. 1986. Analysis of referral networks in marketing: Methods and illustration. *J. Marketing Res.* **23** 370–378.

Reingen, P., B. Foster, J. J. Brown, S. Seidman. 1984. Brand congruence in interpersonal relations: A social network analysis. *J. Consumer Res.* **11** 1–26.

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

Slack, M. 1999. Guerilla marketing breaking through the clutter with word-of-mouth. Jupiter Research, Darien, CT.

Van den Bulte, C., G. Lilien. 2001. *Medical Innovation* revisited: Social contagion versus marketing effort. *Amer. J. Sociology* **106**(5) 1409–1435.

Van den Bulte, C., G. Lilien. 2003. Two-stage partial observability models of innovation adoption. Working paper, Wharton School of Business, Philadelphia, PA.

Walsh, Ekaterina. 2000. *Entertaining Young Net Surfers.* Forrester Research, Cambridge, MA.

Zwillinger, D., ed. 2004. *Standard Mathematical Tables and Formulae*, 30th ed. CRC Press, Boca Raton, FL.