

A Longitudinal Content Analysis of Weblogs: 2003-2004

Susan C. Herring, Lois Ann Scheidt, Inna Kouper, and Elijah Wright
School of Library and Information Science
Indiana University, Bloomington

Introduction

In the past several years, weblogs—frequently modified web pages containing dated entries listed in reverse chronological sequence—have gone from relative obscurity to immense popularity. Weblogs are popular in part because they enable easy, inexpensive self-publication of content for a potentially vast audience on the World Wide Web, while being more flexible and interactive than previous publication formats, print or digital (Herring, Scheidt, Bonus, & Wright, 2005). The rapid efflorescence of the blogosphere—the universe of available weblogs—has also been fertilized by a series of external events over the past few years that have inspired blogging activity: terrorist acts, political events, and natural disasters. Blogging about these events has attracted the attention of the mainstream news media, further contributing to the popularization and adoption of this new mode of computer-mediated communication.

Scholarly research about blogging has also blossomed in recent years, much of it focused on blogs concerned with external events (known as 'filters,' in that they 'filter' information from other sources on the web; Blood, 2002), and the impact bloggers can have as 'citizen journalists' (Gillmor, 2003; Lasica, 2002, 2003), 'public intellectuals' (Park, 2003) and 'opinion leaders' (Delwiche, 2004). The most linked-to, hence presumed influential, bloggers are known as the 'A-list' (Trammell & Keshelashvili, 2005). Other research has examined weblogs as a new communication genre, using content analysis (Herring, Scheidt, et al., 2004, 2005; Papacharissi, 2004), rhetorical analysis (Miller & Shepherd, 2004), and ethnographic interviews (Nardi, Schiano, & Gumbrecht, 2004) to characterize the forms, functions, and audiences of blogs, as well as people's motivations for blogging. The blogging phenomenon is represented in all of this research as dynamic and evolving. As yet, however, no research has empirically investigated how weblogs or blogging practices change over time, or how external events shape blogging practice.

This chapter begins to address this lack by presenting the results of a longitudinal content analysis of random blogs sampled between March 2003 and April 2004. This was a troubled year in world politics: The U.S. began pre-emptive bombing of Iraq amid global protests on March 19, 2003, closely followed by the "shock and awe" blitz on Baghdad by the U.S. and the U.K. One year later, terrorist bombings in Madrid led another ally, Spain, to withdraw its troops from Iraq, around the time of the first outbreak of civil war in Najaf. At the same time, closely watched by the rest of the world, the 2004 U.S. presidential race was heating up after Howard Dean, the leading challenger of George W. Bush's presidency and the first candidate to make effective use of blogging (Kerbel & Bloom, 2005), ruined his chances for election by screaming at a campaign rally on January 29, 2004, and John Kerry cinched the Democratic party nomination on March 2, 2004.

During this event-filled year, the number of active weblogs more than quadrupled. Mass media coverage of weblogs also increased dramatically, bringing blogs into the mainstream consciousness (Drezner & Farrell, 2004). Against this backdrop, we collected and analyzed three random samples at six-month intervals, totaling 457 blogs. Our primary goal was to assess the extent to which the characteristics of blogs themselves, as described in previous research, remained stable or changed during this active period. Our findings reveal longitudinal trends that support earlier observations, as well as variation that suggests a responsiveness to external events.

This chapter is organized as follows. The following section reviews literature on growth of the blogosphere, blogging in response to external events, and structural and functional characteristics of weblogs, as background for the presentation of our research question and methodology. Three categories of results of our longitudinal analysis are then presented and discussed: change over time, stability over time, and variability. In concluding, we identify challenges associated with longitudinal weblog analysis, and suggest directions for future research.

Background

Growth of the Blogosphere

Weblogs have come a long way since Turnbull (1999) first introduced them to *The Scotsman's* readership as "the latest internet craze." Earlier that year two software products that automated blog creation, Blogger and Pitas, were released to the public (Blood, 2002). In the first few months after Blogger's release, the site gained 10-20 new users per day, ending the year with 2,300 registered users. By the end of January 2001, Blogger had registered 117,970 users, a 5,029% increase in 13 months (Turnbull, 2001).

The number of blogs continued to grow at an exponential rate, although estimates vary depending on the methods used to count blogs (see Herring, Kouper, Paolillo, Scheidt, Welsch, Tyworth, Wright, & Yu, 2005). At the beginning of our data collection period in March 2003, the blog-tracking site blo.gs reported 350,000 active blogs (Herring, Scheidt, et al., 2004); by late September 2003, that number had doubled (Herring, Kouper, Scheidt, & Wright, 2004). In early 2004, surveys found that between 2% and 7% of American adult Internet users (between 1.06 and 3.71 million people) had created blogs (Lenhart, Horrigan, & Fallows, 2004). According to blo.gs and Technorati, another blog tracking service, the number of active blogs around the end of our data collection period in April 2004 was between 1.5 and 2 million.

Estimates that included blog hosting services such as LiveJournal, DiaryLand, and Xanga were much higher during the same period. In October 2003, the Perseus Corporation estimated that 4.12 million blogs had been created (Henning, 2003), although many of them were inactive. By the end of the first quarter of 2004, this estimate had risen to 31.6 million blogs, most of them personal journals (Henning, 2005).¹

¹ Since the time of the data collection for this study, the blogosphere continues to grow. By February 2006, Technorati was estimating that the blogosphere was over 60 times the size it was in 2003 (Sifry, 2006).

Response to External Events

As the number of active blogs has grown, so too has their ability to respond to the world around them. After the events of 9/11, traditional media outlets could not keep up with the original reporting and man-on-the-street commentary found in blogs (Stone, 2004). On September 11, 2001, veteran blogger Dave Winer reported that "[m]ost of the major news sites are inaccessible, but news and pictures are reaching us through email, webcams and the weblog community" (Winer, 2001). Blog posting increased in amount and frequency in the week after 9/11, as did the number of comments per post and the number of posts per individual blogger (Krishnamurthy, 2002).

Bloggers have also shown that they can ignite the traditional media's interest in a topic. For example, when in 2002 traditional media did not respond with significant outcry over U.S. Senator Trent Lott's racist remarks at Strom Thurmon's 100th birthday party, bloggers took over the story with such vigor that traditional media outlets were forced to renew their coverage (Shachtman, 2002). Lott's eventual resignation has been attributed to the influence exerted by bloggers. Similarly, blogs have shown a light onto stories that were missed by the traditional media, thereby acting as agenda setters for the mass media (Bloom, 2003; Delwiche, 2004).

Bloggers as Journalists

Media attention to weblogs has grown in proportion to the increase in the number of weblogs published. In their study of weblog influence, Drezner and Farrell (2004) found that only 11 news articles published between 1995 and 1999 used the term "weblog;" by 2003 that number had grown to 647 articles. Part of the interest of the mainstream media stems from the perception of blogging as an alternative form of journalism (Blood, 2003; Bloom, 2003; Gillmor, 2003; Lasica, 2002, 2003; Welch, 2003).

However, although some of the most-read A-list bloggers are professional journalists (Park, 2003), most bloggers would not call themselves journalists (Lasica, 2002) and do not even dream of being journalists (Cook, 2005). Nor would their writing qualify as journalism, since most "blogs focus on narrow subject matter of interest to a select but circumscribed niche. And the blogs that do contain bona fide news are largely derivative, posting links to other blogs and, in many cases, print journalism" (Andrews, 2003).

Warbloggers

Among the blogs that have attracted the most media attention are the so-called warblogs. Warbloggers are filter bloggers who express their views on current political events. In recent years, they address post-9/11 issues, in particular the wars in Afghanistan and Iraq. Early warbloggers were labeled as "a hawkish bunch" (Cavanaugh, 2002) while more recent accounts have credited warbloggers with the rise of anti-war sentiments (Chesher, 2005).

Warblogging differs from journalistic blogging in that most warbloggers write their commentary at a safe distance from the action on which they comment (Cavanaugh,

2002). In publishing their commentaries, they are dependent on the mainstream media for the details that populate their writings. Other warblogs are written by those in the field, in particular by members of the armed forces, embedded reporters, Iraqis, or others who write about their first-hand experiences (Thompson, 2003). Whether first-hand accounts or commentary on published news, warbloggers differ from mainstream news media in their open inclusion of personal outlook in their writing (Wall, 2004).

The period of our data collection and analysis coincides with the onset of U.S. hostilities and the official declaration of war against Iraq, a highly controversial topic in the blogosphere. The Iraq war was also made a central issue by bloggers in the 2004 U.S. presidential election campaign (for the role played by blogging in the campaign, see, e.g., Bloom, 2003; Kerbel & Bloom, 2005).

Content Analysis of Weblogs

Content analysis has been used to analyze the structure, purpose, and themes found in high-profile blogs. Lawson-Borders and Kirk (2005) analyzed campaign weblogs of political candidates during the 2004 U.S. presidential election. They found that the blogs were primarily social diaries and organizing tools for the candidates. Trammell and Keshelashvili (2005) conducted a content analysis of 209 single-authored A-list blogs drawn from Popdex. The list included blogs written by both men and women, where the men produced primarily filter blogs and the women wrote more diary-type blogs. They found that all of the A-list blogs included elements of self-revelation, and metadiscourse about blogging was a common theme.

Content analysis methods have also been used to analyze ordinary blogs. Herring, Scheidt, et al. (2004, 2005) employed content analysis as a means to characterize weblogs as a genre. In a random sample of 203 blogs from spring 2003, we found that the average English-language blog was single-authored, focused on personal events in the blogger's life, contained relatively few links, and received few comments. In a second study that added 154 blogs to the sample (Herring, Kouper, et al., 2004), ordinary bloggers were found to be female nearly as often as male, and young (teens or young adults) as often as adult. However, gender and age of bloggers varied according to blog type, with adult males writing almost all filter blogs, and young females writing the largest proportion of personal journals or diary-type blogs. Papacharissi (2004) conducted a quantitative content analysis of a random sample of 260 blogs hosted by Blogger.com, and found results similar to those of Herring, Scheidt, et al. (2004, 2005). Papacharissi characterizes blogs as having more in common with diaries than with independent journalism.

Recent research suggests that there are cultural differences in blogging practices. Trammell, Tarkowski, and Hofmokl (2006), in their content analysis of 358 Polish-language blogs drawn from blog.pl, found that Polish bloggers are more often female (nearly 75%) and younger than English-language bloggers, often elementary and middle-school ages. Similar to Scheidt (2006), Trammell et al. found that the themes of young Polish female bloggers' posts tend to focus on the emotional impact of the life events they are relaying.

The present study goes beyond this research by incorporating a longitudinal perspective. Our primary research question concerns the extent to which the

characteristics of random English-language blogs and blog authors, e.g., as identified by Herring, Scheidt, et al. (2004, 2005), Herring, Kouper, et al. (2004), and Papacharissi (2004), remained stable or changed between 2003 and 2004. The answer to this question has implications for understanding the nature of emergence and change in new media, and for predicting future trends in blogging practice.

Methodology

Data

This study is based on content analysis of three samples of random weblogs collected at roughly six-month intervals. 203 blogs were collected between March and April, 2003, 154 blogs were collected September 29-30, 2003, and 100 blogs were collected April 12-13, 2004, for a total of 457. All three samples were obtained by using the random blog selection feature of the blog tracking website blo.gs. At the time, the blo.gs site tracked the largest number of blogs from the most diverse sources (the site has since been bought by Yahoo! and the random feature disabled). Individuals could contact blo.gs to have their blogs tracked, and lists of updated weblogs were imported by blo.gs every hour from antville.org, blogger.com, pitas.com, and weblogs.com. Later in the study period, blo.gs also began tracking weblogs from LiveJournal and other sites.

To create coherent corpora that would lead to optimally interpretable results, we excluded non-English blogs, photo and audio blogs, and instances of blog software being used for non-blog purposes (e.g., news aggregators, community center events calendars, retail/advertising). We also excluded blogs containing fewer than two entries, to ensure that the blogs were not one-time experiments. Thus any generalizations from this study apply only to active, English-language, text-based weblogs. These constituted the majority of available blogs during the period under investigation.²

Analytical Methods

Content analysis (Bauer, 2000; see Bates & Lu, 1997; Ha & James, 1998; McMillan, 1999 for applications of content analysis methodology to websites) was conducted to identify and quantify structural and functional properties of the blogs in the three corpora. We coded for 22 categories including blogger characteristics, blog type, blog software used, and textual and interactive features of the first (most recent) entry. These codes are a reduced set of those developed and described in greater detail in Herring, Scheidt, et al. (2004, 2005).

Blogger characteristics include gender, age, occupation (if indicated), and identity indicators (e.g., choice of name, inclusion of photo). We searched for this information throughout the entire blog. *Blog type* codes include filter, personal journal, and k(nowledge)-log. A blog was coded as one of the first three types if a clear majority

² We did not draw random blogs from blog hosting services such as LiveJournal and Xanga (other than those tracked through blo.gs), because the infrastructure and the blogging practices of those services, especially in 2003, seemed sufficiently different to warrant separate treatment. The results of the present analysis may not apply to blogs hosted on these services.

of its entries exhibited the labeled content; blogs with a roughly balanced mix of content were coded as "mixed," and blogs whose content indicated a purpose other than those described above were coded as "other." Blog type was usually identifiable from the title, blog description (if any), and entries on the blog's homepage. Finally, the *first blog entry* was coded for number and nature of links (e.g., to other blogs or news sources), number of comments received, and presence of images. We also measured text length by counting the number of words and sentences in the first entry of each blog.

The spring 2003 and fall 2003 samples were coded by three out of the four authors, who achieved better than 80% agreement on code assignment; the spring 2004 sample was coded by the third author; and the coding for all three samples was checked for consistency by the first author. The findings are reported in this chapter using descriptive statistics with a focus on comparison across the three samples. Because multi-authored blogs are infrequent in the samples (see Table 4 below) and raise separate issues about blog authorship, most of the findings are reported for single-authored blogs only.

Results

The results of our longitudinal content analysis can be grouped into three categories: change, stability, and variability. Change indicates that certain characteristics of weblogs show a clear pattern of increasing or decreasing over time. Stability refers to characteristics that do not change appreciably over time. Variability indicates results that do not show a clear directional pattern, but fluctuate from sample to sample.

Change Over Time

An examination of the structural characteristics of entries in single-authored blogs reveals that despite the general expectation that the Web will generate more and more multimedia content (cf. Herring, 2004), weblogs appear to be becoming increasingly text-based. Over time the bloggers in our corpus tended to post longer entries with more words and more sentences, and that included fewer images. Blogs also become less connected to each other (see Table 8 below) and to the rest of the Web, as the average number of links per entry decreased. The results for these structural characteristics are summarized in Table 1. Change over time in the average number of words per entry is also represented graphically in Figure 1.

Table 1. Structural characteristics of blog entries

<i>Sample</i>	<i>Number of words</i>	<i>Number of sentences</i>	<i>Number of images</i>	<i>Number of links</i>	<i>Number of comments</i>
March-April 2003	202.95	15.17	0.11	0.65	0.40
September 2003	224.51	17.87	0.04	0.43	0.64
April 2004	299.62	22.05	0.04	0.47	0.39

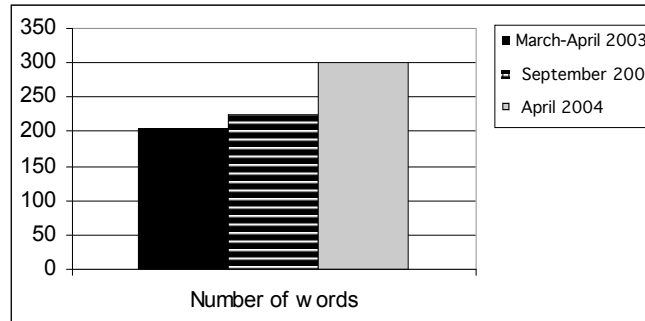


Figure 1. Average number of words in blog entries

Even in the first sample, the frequency of images per entry is very low (0.11). The average number of links per entry also starts out low (0.7), considering popular claims about the overall connectivity of the blogosphere (e.g., Blood, 2002). In light of this, although the amount of measurable change is slight, the fact that images and links decrease rather than hold constant or increase is noteworthy. The number of comments per entry fluctuates, but also remains low overall (0.4, 0.6, 0.4).³

It is often claimed that computer-mediated communication encourages anonymity. The data from the three samples do not support this claim, in that most of the blog authors provide some identifiable personal information, and this tendency increases over time: 65.9%, 77.3% and 78.6% of authors in the three samples, respectively, use either their first name or their full name. In particular, the use of first names increases over time; in contrast, pseudonyms and blogs in which no name is given decrease steadily. These results are summarized in Table 2 and represented graphically in Figure 2.

Table 2. Blogger name types (percentages)

Name type	March-April 2003	September 2003	April 2004
No name	14 (8%)	8 (5.7%)	4 (4.5%)
Pseudonym	43 (24.4%)	23 (16.3%)	13 (14.6%)
First name (or transparently derived nickname)	57 (32.4%)	74 (52.5%)	44 (49.4%)
Full name	59 (33.5%)	35 (24.8%)	26 (29.2%)
Other	1 (0.6%)	-	1 (1.1%)
First name & initial	2 (1.1%)	1 (0.7%)	-
Initial & last name	-	-	1 (1.1%)
Total	176 (100%)	141 (100%)	89 (100%)

³ About half of the blogs in all three samples did not allow comments on entries, either because that was the default setting of the blog software used, or because the author had deactivated the comment option. The average numbers of comments received for those blogs that allowed comments are still low, and show a similar pattern of fluctuation across the three samples (0.9, 1.3, 0.7).

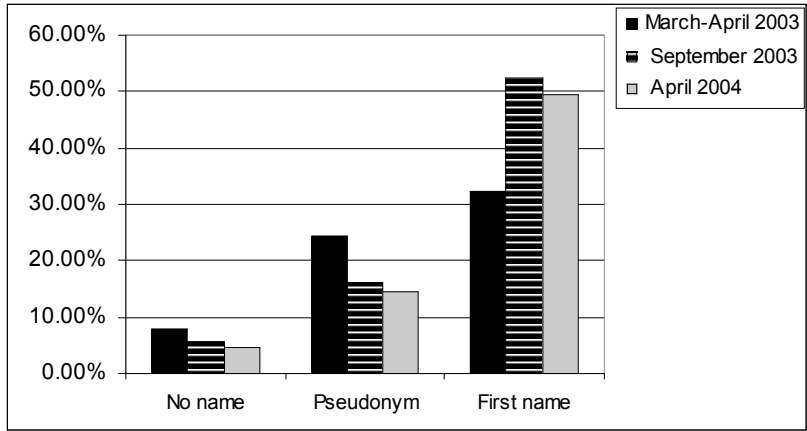


Figure 2. No names, pseudonyms, and first names

The tendency for first names to increase at the expense of all other name types suggests that first names were emerging as the conventionalized means of blogger self-reference during this period.

A majority of bloggers also gave some indication of their occupation, either explicitly or implicitly. Over half of the bloggers are students in all three samples (see Table 3), although the frequency fluctuates somewhat over time. The second most frequent occupation is information technology work (programmers, webmasters, system administrators, etc.); their frequency increases from 15.4% in spring 2003 to 16.2% in fall 2003 to 20.4% in spring 2004. Although the actual numbers involved are small, we include this finding because it is opposite what we expected based on the popular view that blogging is becoming increasingly adopted by "ordinary," less technically-sophisticated users.

Table 3. Occupation of blog authors for whom occupation could be determined

Occupation	March-April 2003	September 2003	April 2004
Student	53 (50.9%)	62 (59.0%)	30 (55.6%)
Teacher/Faculty	11 (10.6%)	2 (1.9%)	3 (5.6%)
IT worker	16 (15.4%)	17 (16.2%)	11 (20.4%)
Other ⁴	24 (23.1%)	24 (22.9%)	10 (18.5%)
Total	104 (100%)	105 (100%)	54 (100%)

Stability Over Time

Two main characteristics remain stable over time: the number of blog authors and the presence of an image (typically a photograph) of the author on the homepage of the blog. An overwhelming majority of blogs are created and maintained by one author (Table 4). In an almost equal number of cases, the blogs in our corpus lack an image of the blog author (Table 5). This last practice contrasts with the practice in personal home pages

⁴ Each occupation in this category was encountered fewer than 10 times in all three samples.

(Bates & Lu, 1997), and contributes to the impression that these weblogs are primarily textual documents.

Table 4. Number of blog authors

	<i>March-April 2003</i>	<i>September 2003</i>	<i>April 2004</i>
One author	176 (91.2%)	141 (92.8%)	89 (89.0%)
Multiple authors	17 (8.8%)	11 (7.2%)	11 (11.0%)
Total	193 (100%)	152 (100%)	100 (100%)

Table 5. Presence of author image on blog homepage

	<i>March-April 2003</i>	<i>September 2003</i>	<i>April 2004</i>
No image	128 (81%)	104 (80%)	60 (80%)
Image	30 (19%)	26 (20%)	15 (20%)
Total	158 (100%)	130 (100%)	75 (100%)

In addition, the blogging software used to host and publish blogs remains stable overall, despite some fluctuations. Blogger is the preferred publishing software in all three samples (67.9%, 77.5%, and 67.4%, respectively), while various other blog software types gain or lose users over time. Over the 13-month period, Movable Type and Pitas lost ground, and LiveJournal and Blogdrive gained in popularity. However, none of these fluctuations affected the overall dominance of Blogger. The frequencies for the most common types of blogging software used in the samples are shown in Table 6.

Table 6. Blogging software used

<i>Sample / Type</i>	<i>March-April 2003</i>	<i>September 2003</i>	<i>April 2004</i>
Blogger	106 (67.9%)	107 (77.5%)	58 (67.4%)
Movable Type	19 (12.2%)	14 (10.1%)	5 (5.8%)
Pitas	12 (7.7%)	-	3 (3.5%)
Radio	6 (3.8%)	-	-
Blogdrive	-	6 (4.3%)	8 (9.3%)
Live Journal	-	5 (3.6%)	3 (3.5%)
Typepad	-	3 (2.2%)	3 (3.5%)
Other	13 (8.4%)	3 (2.1%)	6 (7%)
Total	156 (100%)	138 (100%)	86 (100%)

Variability Over Time

The largest number of features displays a pattern of variability, according to which the first and third samples pattern together in contrast with the second sample. This variable pattern is evident for blogger gender, blogger age, blog type, and link type. It is also evident to some extent for number of comments, student bloggers, and use of Blogger software.

A majority of bloggers are male in samples 1 and 3, whereas in sample 2, males and females are represented nearly equally (with a slight female majority). Gender of blog authors is shown in Table 7, and represented graphically in Figure 3.

Table 7. Gender of blog authors

Sample / Gender	March-April 2003	September 2003	April 2004
Males	95 (54%)	68 (48.2%)	51 (57.3%)
Females	74 (42%)	70 (49.6%)	36 (40.4%)
Unknown	7 (4%)	3 (2.1%)	2 (2.2%)
Total	176 (100%)	141 (100%)	89 (100%)

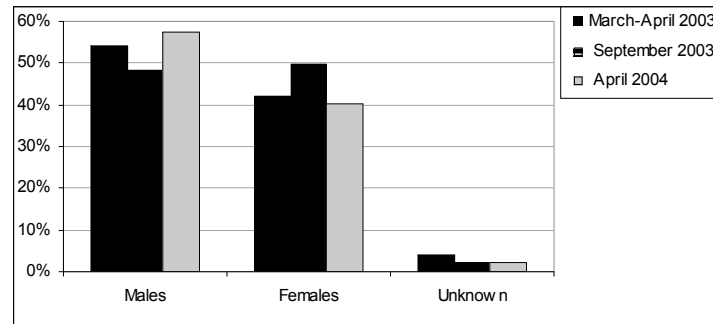


Figure 3. Gender of blog authors

Regarding age, adult bloggers predominate in samples 1 and 3, while sample 2 has equal numbers of adults and teenagers. The percentage of young adults (aged 20-25) increases across the three samples. These patterns are shown in Table 8 and Figure 4.

Table 8. Age of blog authors

Sample / Age	March-April 2003	September 2003	April 2004
Adults	105 (64.8%)	50 (38.2%)	42 (47.2%)
Teenagers	44 (27.2%)	50 (38.2%)	23 (25.8%)
Young adults (20-25 years old)	13 (8%)	31 (23.7%)	24 (27%)
Total	162 (100%)	131 (100%)	89 (100%)

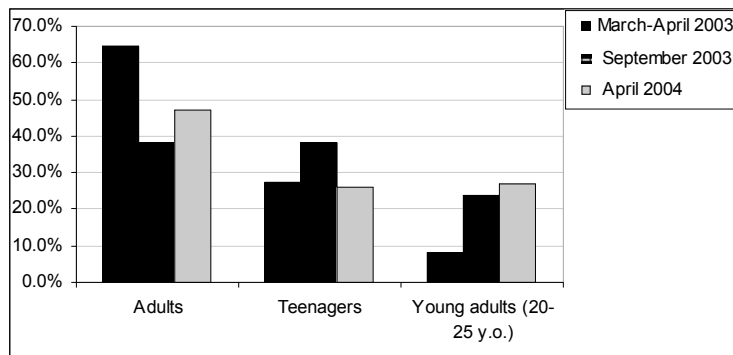


Figure 4. Age of blog authors

Blog type exhibits similar variation. Filter blogs, although not as common as personal journals overall, are more frequent in samples 1 and 3 than in sample 2, while personal journals show the inverse pattern. Mixed blogs pattern with personal journals; k-logs and blogs classified as 'other' were infrequent in all three samples. The frequencies of all blog types are shown in Table 9; Figure 5 graphically represents the distribution of personal journals, filters, and k-logs.

Table 9. Blog type

Sample / Blog type	March-April 2003	September 2003	April 2004
Personal journal	115 (65.3%)	105 (74.5%)	60 (67.4%)
Filter	25 (14.2%)	9 (6.4%)	14 (15.7%)
K-log	7 (4%)	3 (2.1%)	2 (2.2%)
Mixed	22 (12.5%)	21 (14.9%)	10 (11.2%)
Other	6 (3.4%)	3 (2.1%)	1 (1.3%)
Unknown	1 (0.6%)	-	2 (2.2%)
Total	176 (100%)	141 (100%)	89 (100%)

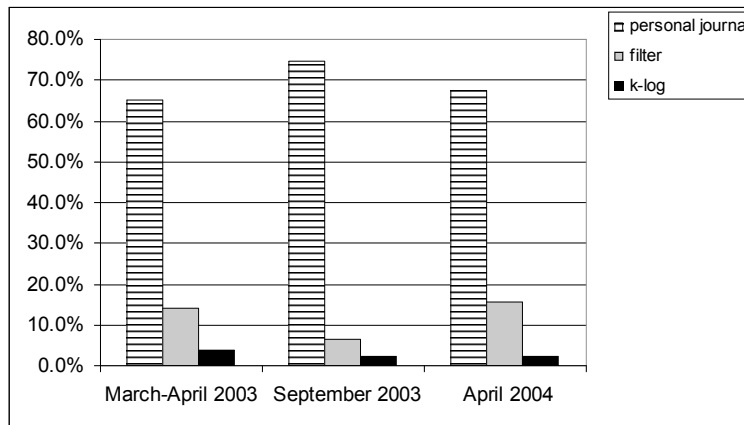


Figure 5. Blog type: personal journal, filter and k-log

Even the links in the first entries show this pattern of variation for some link types. Links to news sources and other websites are more frequent in samples 1 and 3 than in sample 2, while entries with no links are most frequent in sample 2. Links to other blogs, considered by Blood (2002) to be an essential characteristic of the blog genre, are surprisingly infrequent and tend to decrease over time. These findings are summarized in Table 10.

Table 10. Link types

<i>Sample / Link type</i>	<i>March-April 2003</i>	<i>September 2003</i>	<i>April 2004</i>
No links	120 (61.2%)	112 (76.7%)	65 (70.7%)
Links to other blogs	11 (5.6%)	8 (5.5%)	2 (2.2%)
Links to news sources	13 (6.6%)	6 (4.1%)	8 (8.7%)
Links to other web sites created by blogger	4 (2%)	2 (1.4%)	-
Links to other web sites created by others	47 (24%)	16 (11%)	17 (18.5%)
Internal blog links	1 (0.5%)	2 (1.4%)	-
Total	196 (100%)	146 (100%)	92 (100%)

Finally, the first and third samples also differ from the second in terms of number of comments received on entries (Table 1), student bloggers (Table 5), and use of Blogger software (Table 6). All of these are highest in sample 2.

Discussion

Despite the relatively brief time span covered in this study (13 months), some changes are evident in the characteristics of the weblogs sampled. These changes may be due to the fact that weblogs were still relatively young and emergent as a mode of computer-mediated communication at the time. During the period studied, blog entries became more textual, supporting the trend identified in Herring (2004) towards a renewed popularity of textual forms of CMC. Blog authors increasingly presented themselves by their first name and decreasingly by other means, suggesting convergence towards a convention of friendly, informal self-identification.

At the same time, many characteristics of the weblog genre remained more or less stable during this period. Consistent with earlier findings, the frequency of links and comments in blog entries remained low. This finding is contrary to the popular characterization of the blogosphere as interconnected and conversational (see also Herring, Kouper, et al., 2005). The longitudinal evidence also supports the earlier characterization of most blogs as single-authored, personal diaries (Herring, Scheidt, et al. 2004, 2005; Papacharissi, 2004). The overall predominance of personal over news content, together with the low incidence of links to news sources, argues against any simple equation of blogging with 'citizen journalism' (cf. Gillmor, 2003). Moreover, the stability of these findings suggests that the weblog genre had already taken on these characteristics by early 2003.

It remains to explain the pattern of variability according to which many characteristics of blog content and blog authors differed between samples 1 and 3 and sample 2. We suggest that these findings are interrelated, and that the variation reflects variation in external events, specifically, in the concentration and nature of the newsworthy events taking place at the time of the three samples.

Although the entire 13-month period was active in terms of news about the Iraq war, more important events took place just before the spring 2003 and spring 2004 samples, especially relating to Iraq, than before the fall 2003 sample. September was a

relatively quiet news month, with the exception of the stabbing in a department store of the Swedish foreign minister on September 10th, and Estonia's announcement that it would join the European Union on September 14th, which did not attract much attention from the blogosphere. In contrast, spring 2003 saw the outbreak of the Iraq war and the "shock and awe" bombing of Baghdad, and spring 2004 saw the Madrid bombings, the start of civil war in Iraq, and John Kerry's unseating of Howard Dean in the Democratic presidential primaries. Both periods were followed by a dramatic rise in new blog creation, as compared to the number of new blogs created after September 2003 (Sifry, 2004), as shown in Figure 6.

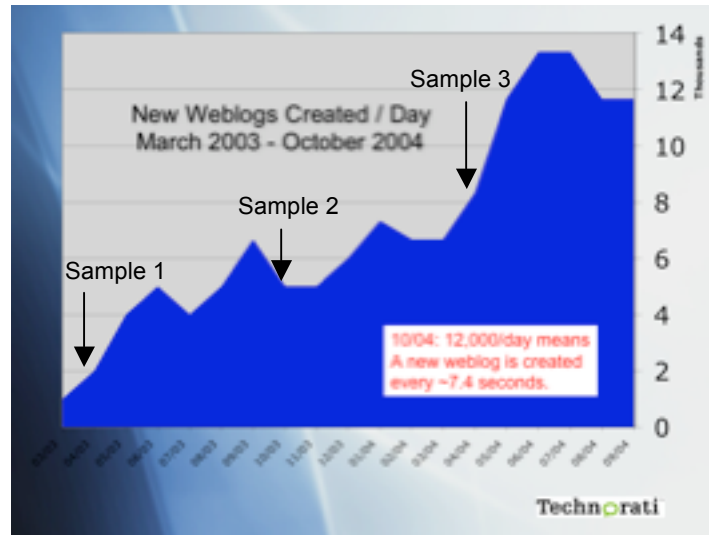


Figure 6. New weblogs created after each sample (modified from Sifry, 2004)

In addition to inspiring more people to start blogging, we suggest that the events around the time of sample 1 and sample 3 triggered more blogging about those events. Although we did not code for what people were blogging about in this study, indirect evidence from the features we did code supports this view. The number of filter blogs was highest in the first and third samples, and along with it, the number of male bloggers and adult bloggers. Filter blogs by definition comment on external events, are more likely to link to news sources and other web sites, and they are written overwhelmingly by adult males (Herring, Kouper, et al., 2004; Trammell & Keshelashvili, 2005). Conversely, the second sample included the highest proportion of personal journal blogs, fewer links, more students, and more equal numbers of women and young people, more of whom used easy-to-access Blogger software. This latter pattern suggests what blogs might look like in calmer times, and has implications for the future trajectory of the technology, since in a larger historical perspective, times of peace are more common than times of war.

Conclusion

This study presented the first longitudinal content analysis of blogs. More longitudinal analysis is needed as the weblog genre continues to evolve and expand its range of uses. In particular, research is needed on multimedia modes of blogging such as photoblogs

and audio and video blogging, which increased in popularity after our data collection period, as well as on the recent phenomenon of podcasting. It may be that the increase in textuality we observed from 2003 to 2004 was accompanied by an increasing specialization in multimedia elsewhere in the repertoire of weblog types. Understanding the forces operating in this process of differentiation would have broader implications for the evolution of new media formats.

This study was carried out at a historically significant moment, during the first year of the controversial U.S.-led war against Iraq. However, it did not analyze blog content specifically about the war or other news events. We do not know, for example, to what extent the filter bloggers in our samples were 'warbloggers.' Future longitudinal research should focus on the themes that people are blogging about, in order to probe in greater depth the relationship between blogging and the external world. The present study suggests that this relationship is gendered, in line with the public man / private woman dichotomy that has been previously identified in gender research (Elshtain, 1993). These dynamics deserve further exploration.

Perhaps the greatest challenge in analyzing weblogs longitudinally lies in identifying comparable samples at different points in time. In this study, we used blogs randomly selected by the blog-tracking site blo.gs, which more or less insured that we were drawing from the same sources, allowing that blo.gs expanded its coverage somewhat during the period of our data collection. The random selection service has since been discontinued, however, effectively limiting our ability to extend the methodology of this study to later samples. Stable tools for identifying, tracking, and searching blogs are essential to facilitate longitudinal analysis of the blogosphere, as well as to enhance the functionality of weblogs.

References

- Andrews, P. (2003). Is blogging journalism? *Nieman Reports*, 57(3), 63-64. Available from <http://www.nieman.harvard.edu/reports/03-3NRfall/V57N3.pdf>
- Bates, M. J., & Lu, S. (1997). An exploratory profile of personal home pages: Content, design, metaphors. *Online and CD Review*, 21, 331-340.
- Bauer, M. W. (2000). Classical content analysis: A review. In M. W. Bauer & G. Gaskell (Eds.), *Qualitative Researching with Text, Image, and Sound: A Practical Handbook* (pp. 131-151). London: Sage Publications.
- Blood, R. (2002). Introduction. In J. Rodzvilla (Ed.), *We've Got Blog: How Weblogs are Changing Our Culture* (pp. ix-xiii). Cambridge MA: Perseus Publishing.
- Blood, R. (2003). Weblogs and journalism: Do they connect? *Nieman Reports*, 57(3), 61-62. Available from <http://www.nieman.harvard.edu/reports/03-3NRfall/V57N3.pdf>
- Bloom, J. (2003, August). The Blogosphere: How a once-humble medium came to drive elite media discourse and influence public policy and elections. *Proceedings of the Annual Meeting of the American Political Science Association*. Philadelphia: American Political Science Association. Retrieved November 30, 2004 from http://archive.allacademic.com/publication/getfile.php?file=docs/apsa_proceeding/2003-08-22/3001/apsa_proceeding_3001.PDF&PHPSESSID=d99a266b9d5b3aa8aad8c907a2cdbf1e
- Cavanaugh, T. (2002). Let slip the blogs of war. In J. Rodzvilla (Ed.), *We've Got Blog: How Weblogs are Changing Our Culture* (pp. 188-197). Cambridge, MA: Perseus. Retrieved March 24, 2006 from <http://www.ojr.org/ojr/workplace/1017770789.php>
- Chesher, C. (2005, May). *Blogs and the crisis of authorship*. Presented at the meeting of the Blogtalk Downunder, Sydney. Retrieved May 24, 2005 from http://incsub.org/blogtalk/?page_id=40
- Cook, T. (2005, May). *Up against reality: Blogging and the cost of content*. Presented at the meeting of the Blogtalk Downunder, Sydney. Retrieved May 24, 2005 from http://incsub.org/blogtalk/?page_id=42
- Delwiche, A. (2004, May). Agenda-setting, opinion leadership, and the world of web logs. Paper presented at the Annual Conference of the International Communication Association, New Orleans, LA.
- Drezner, D., & Farrell, H. (2004, September). The power and politics of blogs. *Proceedings of the annual meeting of the American Political Science Association*. Chicago IL: American Political Science Association. Retrieved November 30, 2004 from http://archive.allacademic.com/publication/getfile.php?file=docs/apsa_proceeding/2004-08-18/30340/apsa_proceeding_30340.PDF&PHPSESSID=d99a266b9d5b3aa8aad8c907a2cdbf1e
- Elshtain, J. B. (1993). *Public Man, Private Woman: Women in Social and Political Thought*, 2nd Edition. Princeton: Princeton University Press.

- Gillmor, D. (2003). Moving toward participatory journalism. *Nieman Reports*, 57(3), 79-80. Available from <http://www.nieman.harvard.edu/reports/03-3NRfall/V57N3.pdf>
- Ha, L., & James, E. L. (1998). Interactivity reexamined: A baseline analysis of early business web sites. *Journal of Broadcasting and Electronic Media*, 42, 457-474.
- Henning, J. (2003). The blogging iceberg - Of 4.12 million hosted weblogs, most little seen, quickly abandoned. *Perseus Development Corporation White Papers*. Retrieved October 4, 2003 from <http://www.perseus.com/blogsurvey/>
- Henning, J. (2005, April 12). The blogging geyser. *Perseus Development Corporation White Papers*. Retrieved April 13, 2005 from <http://www.perseus.com/blogsurvey/geyser.html>
- Herring, S. C. (2004). Slouching toward the ordinary: Current trends in computer-mediated communication. *New Media & Society*, 6(1), 26-36.
- Herring, S. C., Kouper, I., Scheidt, L. A., & Wright, E. (2004). Women and children last: The discursive construction of weblogs. In: L. Gurak, S. Antonijevic, L. Johnson, C. Ratliff, & J. Reyman (Eds.), *Into the Blogosphere: Rhetoric, Community, and Culture of Weblogs*. University of Minnesota. Retrieved March 24, 2006 from http://blog.lib.umn.edu/blogosphere/women_and_children.html
- Herring, S. C., Scheidt, L. A., Bonus, S., & Wright, E. (2004). Bridging the gap: A genre analysis of weblogs. *Proceedings of the 37th Hawai'i International Conference on System Sciences (HICSS-37)*. Los Alamitos: IEEE Computer Society Press. <http://www.blogninja.com/DDGDD04.doc>
- Herring, S. C., Scheidt, L. A., Bonus, S., & Wright, E. (2005). Weblogs as a bridging genre. *Information, Technology & People*, 18(2), 142-171.
- Kerbel, M. R., & Bloom, J. D. (2005). Blog for America and civic involvement. *The Harvard International Journal of Press/Politics*, 10(4), 3-27. Retrieved March 24, 2006 from <http://www.uoregon.edu/~jbloom/BlogforAmerica.pdf>
- Krishnamurthy, S. (2002, October). *The multidimensionality of blog conversations: The Virtual Enactment of September 11*. Presented at the meeting of the Internet Research 3.0: NET / WORK / THEORY, Association of Internet Researchers (AoIR), Maastricht, The Netherlands.
- Lasica, J. D. (2002, April 18). Blogging as a form of journalism. *USC Annenberg Online Journalism Review*. Retrieved March 24, 2006 from <http://www.ojr.org/ojr/lasica/1019166956.php>
- Lasica, J. D. (2003). Blogs and journalism need each other. *Nieman Reports*, 57(3), 70-73. Available from <http://www.nieman.harvard.edu/reports/03-3NRfall/V57N3.pdf>
- Lawson-Borders, G., & Kirk, R. (2005). Blogs in campaign communication. *American Behavioral Scientist*, 49(4), 548-559. Abstract retrieved March 24, 2006 from <http://www.google.com/search?client=safari&rls=en&q=lawson+borders+kirk+blogs+campaign&ie=UTF-8&oe=UTF-8>

- Lenhart, A., Horrigan, J. B., & Fallows, D. (2004, June). Content creation online. *PEW Internet & American Life Project*. Available from http://www.pewinternet.org/pdfs/PIP_Content_Creation_Report.pdf
- McMillan, S. J. (2000). The microscope and the moving target: the challenge of applying content analysis to the World Wide Web. *Journalism and Mass Communication Quarterly*, 77(1), 80-98. Retrieved March 28, 2006 from <http://web.utk.edu/~sjmcmill/Research/research.htm>
- Miller, C. R., & Shepherd, D. (2004). Blogging as social action: A genre analysis of the weblog. In L. Gurak, S. Antonijevic, L. Johnson, C. Ratliff, & J. Reyman (eds.), *Into the Blogosphere: Rhetoric, Community, and Culture of Weblogs*. University of Minnesota. Retrieved March 24, 2006 from http://blog.lib.umn.edu/blogosphere/blogging_as_social_action_a_genre_analysis_of_the_weblog.html
- Nardi, B., Schiano, D., & Gumbrecht, M. (2004). Blogging as social activity, or, would you let 900 million people read your diary? *Proceedings of Computer Supported Cooperative Work 2004*. Retrieved March 24, 2006 from <http://home.comcast.net/%7Ediane.schiano/CSCW04.Blog.pdf>
- Papacharissi, Z. (2004, May). The blogger revolution? Audiences as media producers. Paper presented at the Annual Conference of the International Communication Association, New Orleans.
- Park, D. (2003, October). Bloggers and warbloggers as public intellectuals: Charging the authoritative space of the weblog. Paper presented at Internet Research 4.0, Toronto, Canada.
- Scheidt, L. A. (2006). Adolescent diary weblogs and the unseen audience. In D. Buckingham & R. Willett (Eds.), *Digital Generations: Children, Young People and New Media*. London: Lawrence Erlbaum. Retrieved March 28, 2006 from http://loisscheidt.com/linked/2006/Adolescent_Diary_Weblogs_and_the_Unseen_Audience.pdf
- Shachtman, N. (2002, December 23). Blogs make the headlines. *Wired News*. Retrieved March 24, 2006 from <http://www.wired.com/news/culture/0,1284,56978,00.html>
- Sifry, D. (2004, October 10). State of the Blogosphere, October 2004 (Part 1). *Sifry's Alerts*. Retrieved March 25, 2006 from <http://www.sifry.com/alerts/archives/000387.html>
- Sifry, D. (2006, February 6). State of the Blogosphere, February 2006. Part 1: On Blogosphere Growth. *Sifry's Alerts*. Retrieved February 16, 2006 from <http://www.sifry.com/alerts/archives/000419.html>
- Stone, B. (2004). *Who Let the Blogs Out?: A Hyperconnected Peek at the World of Weblogs*. New York: St. Martin's Press.
- Thompson, G. (2003). Weblogs, warblogs, the public sphere, and bubbles. Retrieved March 24, 2006 from http://www.svsu.edu/~glt/Transformations_piecerev.pdf

- Trammell, K. D., & Keshelashvili, A. (2005). *Examining the new influencers: A self-presentation study of A-List blogs*. *Journalism & Mass Communication Quarterly*, 82 (4), 968-982.
- Trammell, K. D., Tarkowski, A., & Hofmokl, J. (2006). Rzeczpospolita blogów [Republic of Blog]: Examining the motivations of Polish bloggers through content analysis. *Journal of Computer-Mediated Communication*, 11(3).
<http://jcmc.indiana.edu/vol11/issue3/trammell.html>
- Turnbull, G. (1999, August 30). Log on for the latest Internet craze. *The Scotsman*. Retrieved through EBSCO.
- Turnbull, G. (2001, February 26). The state of the blog: Part 1: Blogger past. *WriteTheWeb*. Retrieved November 28, 2004 from
<http://writetheweb.com/Members/gilest/old/106/>
- Wall, M. A. (2004). *Blogs of war: The changing nature of news in the 21st century*. Presented at the meeting of the Communication Research in the Public Interest, International Communication Association (ICA), New Orleans, LA. Retrieved June 30, 2004 from
http://convention.allacademic.com/ica_schedule/view_paper_info.html?pub_id=889
- Welch, M. (2003). Blogworld and its gravity. *Columbia Journalism Review*, 42(3), 21-27.
- Winer, D. (2001, September 11). Bulletin: Terrorist attack in NY. *DaveNet*. Retrieved March 25, 2006 from
<http://davenet.scripting.com/2001/09/11/bulletinTerroristAttackInNy>