The Liblog Landscape 2007-2008: A Lateral Look

What can you say about liblogs and how they're changing?

The library field has many blogs, both liblogs (blogs by library people) and library blogs (blogs officially published by libraries). Liblogs offer some of the most current and interesting discussions on professional topics—and help us know some of the most interesting people in the library field.

There are certainly more than 1,000 liblogs in all, and possibly more than 1,000 English-language liblogs. People start new liblogs because they have something to say or because they're expected to do so for library school courses or continuing education—after all, it only takes a few minutes to start a free liblog. Some people explicitly shut down liblogs; more simply walk away—but most, I believe, continue to blog, at least once in a while.

Naïve Hypotheses

My naïve sense has been that *creation* of liblogs and posting to liblogs peaked around 2006: That we're seeing fewer new liblogs and existing bloggers are posting less often. I also have the sense that posts are getting longer—and we're seeing more comments per post.

In other words, my sense is that liblogs are maturing. No longer the shiny new toy, they've become an established mechanism that works well for people with something to say. Those who still use them may use them less often but to more effect.

I can provide reasons those naïve hypotheses might be true:

➤ **Fewer new liblogs**? They're not "hot" in 2008—and there are so many out there that people can't expect to make quick reputations by starting blogs.
➤ **Fewer posts**? With feeds and aggregators, there's less urgency to post frequently. For that matter, some bloggers have discovered they're likely to lose subscribers if they post *too* often.

➤ **Longer posts**? Twitter and other tools may replace blogs where someone only has a sentence or a link to publish.
➤ **More comments**? Many bloggers *have* become voices within various parts of the library community, and they may be creating larger conversations as library people view liblogs as fundamental to the field.

Evidence?
Those beliefs were based on anecdotes and conversations. I had no idea whether they reflected reality. The time seemed ripe for a quantitative survey of liblogs—and for a lateral view, seeing how liblogs are changing over time.

That's the basis for this book. Are my naïve conclusions wholly wrong, on the money or partly right? You'll find out as you go along.

Typical Liblogs: True Lies

What does a typical liblog look like? Here are two profiles—both of them "true" statistically, neither of them meaningful.

The average liblog
From a universe of 607 liblogs (all but a few in English), the average (mean) liblog looks like this:
➤ It's 36 months old as of June 2008—it started in June 2005.
➤ In the three-month period from March through May 2007, there were 38 posts totaling 8,457 words, with 45 comments and 11 figures (illustrations). The average post was 220 words long and had 1.1 comments and 0.3 figures.
➤ In the three-month period from March through May 2008, there were 32 posts totaling 7,386 words, with 38 comments and 10 figures. The average post was 244 words long and had 2.0 comments and 0.3 figures.
➤ Comparing March-May 2008 to March-May 2007, the blog had 110% more posts, 127%

more words, 158% more comments, 158% more figures, 173% more words per post, 230% more comments per post, and 158% more figures per post.

By now, even the most innumerate reader should be saying "But that's absurd." So it is. The averages (words per post, comments per post, figures per post) are all *averages of averages*—which may be meaningful, but don't work for a so-called "average liblog." The last bullet is nonsense because of simplifying spreadsheet assumptions. A more meaningful equivalent, albeit still meaningless:

➢ This hypothetical liblog had 15% fewer posts and comments in 2008 than in 2007, 13% fewer words, 5% fewer figures—and 11% longer posts, 73% more comments per post and 14% more figures per post.

The typical blog, second cut

Let's make the numbers a *little* more meaningful by eliminating liblogs that didn't *have* any posts in March-May 2007 (either because they didn't start until June 2007 or later or because they were inactive)—and, for total length and length per post, eliminating liblogs where I was unable to measure the length.

That reduces the universe to 523 blogs for most figures and 475 blogs for length and post length.

Let's also use median figures rather than averages (means), since medians tend to be more realistic in highly variable sets.

This "typical" blog is also 36 months old, beginning in June 2005.

➢ **March-May 2007**: 22 posts totaling 5,346 words, with 11 comments and three figures. The average post was 254 words long with 0.7 comments and 0.1 figures.

➢ **March-May 2008**: 15 posts totaling 3,618 words, with eight comments and two figures. The average post was 224 words long with 0.6 comments and 0.1 figures.

➢ From 2007 to 2008, the blog had 32% fewer posts and words (but 12% shorter posts), 27% fewer comments (17% fewer comments per post) and 33% fewer figures (but 7% more figures per post).

Lies, lies, lies

The average and typical liblogs aren't possible: If a blog has 32% fewer posts *and* 32% fewer words, the average post can't be 12% shorter. Neither does 5,346 divided by 22 equal 254: It equals 243.

But it's also true that no single liblog fits either of these cases, even if the cases are normalized to rid them of inconsistencies (e.g., if the final case showed 243 words per post and 0.5 comments per post for 2007, 241 words per post and 0.5 comments per post for 2008).

Consider *actual* liblogs that are "close" to these two examples—based on age of the blog and number of posts in March-May 2007.

➢ **Average, first cut:** The closest real example is 36 months old with 27 posts in March-May 2007 (27 may not be close to 38, but it's closer than 68, the next more frequent example). These posts total 9,922 words (368 per post), 29 comments (1.1 per post) and 10 figures (0.4 per post). In 2008, there were 15 posts totaling 1,930 words (129 per post), with one comment (0.1 per post) and no figures. That comes out to 44% fewer posts, 81% fewer words (65% shorter posts), 97% fewer comments (94% fewer comments per post) and 100% fewer figures.

➢ **Typical, second cut**: The closest real blog is 36 months old with 23 posts in March-May 2007 totaling 9,438 words (410 per post), 134 comments (5.8 per post) and 27 figures (1.2 per post). For March-May 2008, eight posts totaling 3,315 words (414 per post), 33 comments (4.1 per post) and five figures (0.6 per post). Those are changes of 65% fewer posts, 65% fewer words (1% longer posts—the "65%" figures are actually 65.2% and 64.9% respectively), 75% fewer comments (29% fewer per post) and 82% fewer figures (47% fewer per post).

Those two are *real* liblogs and real changes. Is either of them typical? Not really.

Abandoning the typical

That's two pages devoted to average and typical liblogs, which is probably two pages more than such nonsense terms deserve. Even stripped of spreadsheet issues, the fact is that **there is no such thing as a typical or average liblog**—the creature doesn't exist.

As a result, any statement about changes in typical or average liblogs is *guaranteed* to be wrong. You can discuss liblogs as a whole, but I'm not sure that's much more meaningful.

The first part of this book looks at *groups* of liblogs and individual metrics, but also looks for correlations between different metrics. Are there meaningful things to say about relationships between, say, the age of a blog and changes in blogging behavior? We shall see.

Metrics and Quintiles

The nonsense examples above introduced most of the metrics involved in this broad survey: Quantifiable measures of a blog that can be determined by an outside observer. Qualitative analysis might be interest-

ing, but I for one have not the vaguest idea how you could do such analysis on a universe of blogs…nor any intention of trying.

I use *quintiles* extensively in looking at metrics. Quintiles break a population down into five groups, ideally of equal size but realistically of nearly-equal size, based on a particular metric. So, for example, when looking at number of posts in March-May 2008, the first quintile is the (roughly) 20% of blogs with the most posts, the second quintile the 20% of blogs with more posts than average (60%-80%), the third quintile (40%-60%) blogs with roughly average number of posts, fourth quintile (20%-40%) blogs with fewer posts than average, and fifth quintile (0%-20%) blogs with the fewest posts. I say "roughly" because it would make no sense to split two blogs with the same number of posts between two quintiles—so quintile splits will always come between numbers. For most metrics, I eliminate zero counts before calculating quintiles.

Why quintiles? Because they provide a compact picture of a universe that's not *too* compact—and one that allows for a meaningful "average" range. When you split a population into four parts, the average is split at the median. No matter how little an individual differs from that median, it's forced into the second or third quartile. With quintiles, there's a broad "average group"—the third quintile.

Twenty-two quantitative metrics come into play in this book:
➢ Age of blog
➢ Number of posts in March-May 2007
➢ Total length of posts in March-May 2007
➢ Average post length in March-May 2007
➢ Number of comments in March-May 2007
➢ Comments per post in March-May 2007
➢ Number of illustrations/figures in March-May 2007
➢ Figures per post in March-May 2007
➢ The seven metrics above for March-May 2008
➢ The percentage changes in the seven metrics from March-May 2007 to March-May 2008.

I summarize a few other *descriptive* measures for liblogs in this chapter.

In attempting to see how liblogs have changed, we'll go from quintiles to triplets, but we'll discuss that later on.

Numbers make your head spin?

This survey of the liblog landscape is inevitably heavy on numbers. My apologies if you feel a bit overwhelmed at times. In some cases, I provide a graphic or textual equivalent that offers the same information in a manner you may find more meaningful.

Discovering Liblogs

The universe of liblogs analyzed and described here is *not* complete. I don't believe it is possible to gather the complete universe of liblogs, particularly since new liblogs appear each month (a few typically die each month as well).

I *do* claim this is a broadly representative sample including most liblogs that meet the set of criteria established for this project. Here's how I developed the universe and criteria followed along the way.

Initial inclusions

I began with those blogs included in the "top 60" study I did in 2005 and in the "great middle" study from 2006 that could be still be located. Some blogs from 2005 were also in the 2006 study. That yielded 43 blogs from 2005 (that weren't in the 2006 study) and a total of 227 blogs.

The only blogs that aren't in English come from the 2006 study. I've chosen not to delete those blogs, but this is fundamentally a study of the *English-language* liblog universe.

Criteria for additional blogs

The criteria for additional blogs on the first pass were:
➢ In English.
➢ Not clearly defined as an official library blog.
➢ Somehow related to libraries or librarianship.
➢ Reachable (on the web and not password-protected).
➢ **Established**: At least one post before January 1, 2008.
➢ **Visible**: Sum of Bloglines subscriptions and Technorati "Authority" at least nine (rounding to 1.0 on my Visibility scale) when tested in first two weeks of March 2008.

Why the Visible criterion? Quite a few liblogs seem to operate "under the radar"—they're designed for a small circle of friends with no apparent intention of growing. That criterion was intended to leave such blogs relatively private. It may be a mistake (and, as discussed in Chapter 10, it will definitely need to be replaced with a different criterion).

There was another criterion, "**Not defunct**: At least one post after August 31, 2007." That criterion has not been consistently applied.

First pass for additional blogs

In the first pass, I checked the blogs in the "Favorite blogs" survey done by Meredith Farkas in *Information wants to be free*. I also checked blogs on the LISWiki weblogs list, blogs in the LISZen source list and blogs in a "tag cloud" source list that was, for a while, showing liblog tags for recent posts.

I checked more than 450 distinct new blogs in that pass. Some were too new (established in 2008). Quite a few were "invisible," apparently defunct, or simply not reachable. At the end of the first pass, the study universe included 542 liblogs.

Then I did March-May 2007 metrics for all 542 liblogs. In the process, I deleted three blogs—one because the name and content were simply too offensive, one because the content appeared completely unrelated to libraries and one because it was a "project" blog that was largely hidden and designed to end at a certain point.

Second pass

I looked at the first 100 (or so) liblogs in the list, checking for blogrolls and finding new candidates from those blogrolls. I only looked at front-page blogrolls of plausible length with some evidence of library focus. This yielded another 46 blogs, in addition to more than a hundred candidates that failed one criterion or more.

The universe at this point was 585 liblogs, and this is the point at which I began 2008 metrics. During that process, some blogs disappeared—in one case because I'd inadvertently included the same blog twice, in 11 cases because the blog was no longer reachable. The universe was now down to 573 liblogs.

Final pass

In late August 2008, I tried another "blogroll pass"—this time working from the end of the alphabet backward, with the intention of stopping when I wasn't getting enough yield to make it worthwhile. (Did you know you can print 573 liblog names on one sheet of paper? You can, I did, and it made this process a *lot* easier.)

Unfortunately, I found that my tools for looking at Bloglines subscription counts no longer worked, so I had to use a simplified "Visibility" cutoff that may have been too high. I looked for a Technorati authority figure of 7 or higher, which means at least seven other blogs had linked to this blog in the last six months.

In the end, I looked through 240 liblogs, adding 14 more blogs—and concluded the process wasn't effective when I went through 75 blogs without *any* new candidates.

Then I posted an invitation on *Walt at random* and two widely distributed lists (PUBLIB and Web4LIB) to see what else was out there. That yielded 19 more blogs, for the final count of 607 liblogs.

Describing the Universe

This study involves 607 liblogs. I'm not listing them here. They all appear, in alphabetic order, in Chapter 12, Liblog Profiles (the "second half" of this book)—and there's a web page with URLs at the time of this study at http://walt.lishost.org/blogs-in-the-liblog-landscape-2007-2008/.

The only metric that includes all 607 blogs is age, since some of the blogs didn't have posts during either study period.

The rest of this chapter considers characteristics of the liblog landscape, the universe of this study, that don't play directly into most metric analysis. The first few items are entirely anecdotal. The rest may be significant for subgroup studies.

Descriptive items are all as of August-September 2008; blogs *do* change, as do bloggers.

Program and Typography

For 511 of the 607 blogs, the blogging software is easily identifiable and one of the "big players." Here's the summary:

Program	Blogs	Percentage
WordPress	230	37.9%
Blogger	222	36.6%
TypePad	35	5.8%
MovableType	18	3.0%
LiveJournal	6	1.0%

Table 1.1: Blogging software

In the 2006 study of 213 blogs, 106 (50%) used Blogger, 52 (24%) used WordPress, 19 (9%) used MovableType and 12 (6%) used TypePad. Has there been that much migration in two years? Yes and no. Of the 184 blogs *from that study* that are in this book, 79 (42%) now use Blogger, 60 (32%) use WordPress, 11 (6%) use TypePad and five (3%) use MovableType. There's been some migration toward WordPress from other platforms—but WordPress only holds a clear lead over Blogger among blogs *not* in that study.

Leaving out a handful of blogs that mix serif and sans serif type in posts, 428 blogs use left-aligned sans serif type (71%); 50 use justified sans serif type (8%); 118 use left-aligned serif type (19%); and four use justified serif type (1%). I should say "at least" 50 and 4 respectively for justified type; I may have overlooked some. That gives a total of 79% sans serif (478 blogs) and 20% serif (122 blogs), and a total of 9% justified (54 blogs), the rest left-aligned. Compared to 2006, the dominance of left-aligned sans serif type (the default in many Blogger and WordPress templates) is even more extreme.

A "large handful" of blogs use combinations of text and background colors that make reading difficult—perhaps two dozen out of the 607—but for feed users, that's irrelevant.

Authorship

Is the blog's author clearly identified by full name? I didn't look beyond the blog (e.g., one blogger mentions her blog in signed list posts, but doesn't use her name on her blog), but I did check pages linked directly from the blog itself—"about" pages, vitas, book announcements. Here's what I found.

Author Identification	Blogs	Percentage
Full name	401	66.1%
Group authorship	95	15.7%
Pseudonymous	55	9.1%
First name only	43	7.1%
Anonymous	13	2.1%

Table 1.2: Blog authorship

Adding the last three rows, you could say 111 blogs—18% of the total—have some mystery about their authorship. That overstates the reality. It's clear from comments that some pseudonymous authors are, in fact, known by name to their core readership. Several cases are less mysterious than they appear—e.g., there's a photo of the blogger or the place of employment and position are named.

Affiliation

Some bloggers and groups identify themselves as affiliated with a type of library or other organization. Some don't. I noted affiliations when they seemed to be clear, which turned out to be 443 blogs or 73% of the total. Here's the table, but read the comments that follow the table.

Affiliation	Blogs	Percentage
Academic library	176	29.0%
Public library	102	16.8%
Law library or librarian	26	4.3%
Vendor	19	3.1%
School library	17	2.8%
Medical library or librarian	16	2.6%
Association	14	2.3%
Consultant	13	2.1%
Special library	10	1.6%
National library (including LC)	7	1.2%
Publisher	7	1.2%
Science library or librarian	7	1.2%
Student	7	1.2%
Archives	5	0.8%
Consortium	5	0.8%
State library	4	0.7%
Museum	2	0.3%

Table 1.3: Affiliation

One each: Art, Foundation, News library, Patron, Researcher, Seminary library.

I gave precedence to law, medical, and science over broader library types—so, for example, the 26 "blawgs" (law-related blogs by library people) include people in law libraries in universities, at law firms and possibly elsewhere. "Vendor" includes employees of for-profit and nonprofit firms.

"Academic library" does *not* imply academic *librarian*—I didn't make those distinctions. Sometimes, the affiliation was fuzzy: If a blogger identified herself as both a library school student and an academic library employee, I probably marked the blog as "Academic" unless it was clearly student-focused.

Are academic libraries overrepresented? Yes, compared to the total number of libraries in the English-speaking world—but not, I believe, in comparison to the actual universe of liblogs. I'm less confident about school librarians, although this is a much better representation than I've seen in earlier studies.

I also looked at "special cases"—blogs with clear focus other than affiliation. The list is certainly incomplete, but I did note five blogs concerned with open access, four on cataloging, three each devoted to children's literature and marketing and two devoted to gaming.

Country

I suspect the table below is incomplete—some bloggers outside the United States don't make it clear where they are, possibly because they assume their readership already knows. I identified 24% of the blogs as coming from outside the United States (including the few non-English blogs), as follows—noting that I list Scotland separately from the United Kingdom in general:

Country	Blogs	Percentage
Canada	56	9.2%
United Kingdom	32	5.3%
Australia	21	3.5%
Netherlands	5	0.8%
Sweden	5	0.8%
Philippines	4	0.7%
Denmark	3	0.5%
France	3	0.5%
Germany	3	0.5%
Norway	3	0.5%
Argentina	2	0.3%
Austria	2	0.3%
New Zealand	2	0.3%
Scotland	2	0.3%

One each from Chile, Iran, Serbia, Singapore and Thailand.

Table 1.4: Country of blogs (other than United States)

Age

One primary criterion for inclusion in this universe was age: A blog had to be at least six months old (as of June 2008) to be included—that is, it had to have at least one post in December 2007 or earlier.

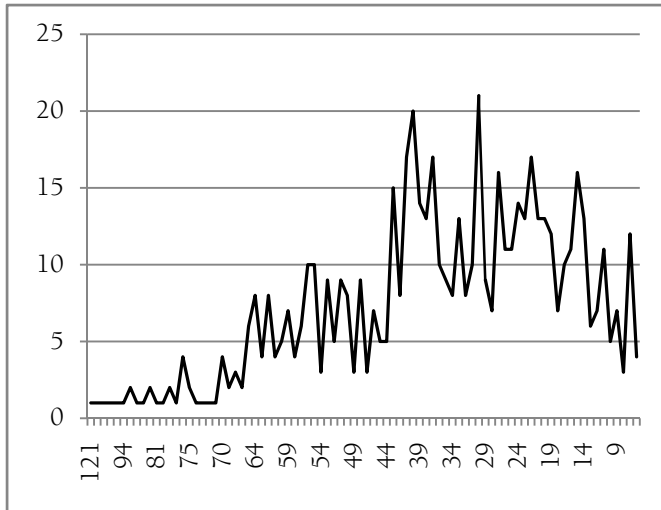Here's a chart showing the age of liblogs by month:

Figure 1.1: Blogs by age in months

The horizontal axis shows age in months (as of June 2008), the vertical axis the number of blogs. As you can see from the chart, the big growth in new liblogs begins in early 2005 and declines in 2007.

The monthly chart may show too much information. Here's a table showing blog start date by year:

Year	Blogs	Percentage
1998	3	0.5%
1999	2	0.3%
2000	1	0.2%
2001	7	1.2%
2002	25	4.1%
2003	75	12.4%
2004	86	14.2%
2005	160	26.4%
2006	144	23.7%
2007	104	17.1%

Table 1.5: Blog start date by year

One naïve conclusion does appear justified based on this universe: the *growth* of liblogs may have peaked in 2005, dropping slowly in 2006 and somewhat faster in 2007.

Here's a graph for the visually-oriented:
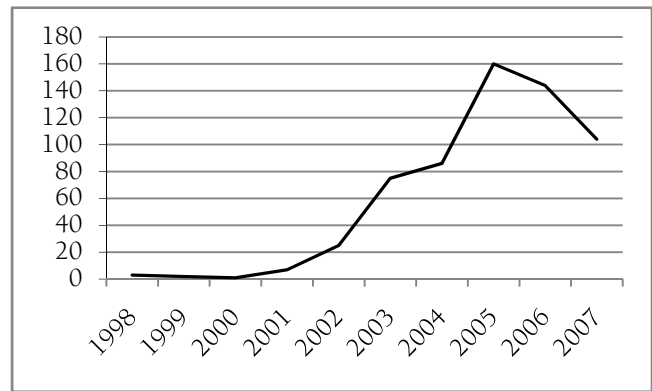
Figure 1.2: Blogs by year of origin

"Blogs by year of origin" and "Blogs by age in months" show exactly the same data in the same order—but months have been aggregated for the second graph.

That's the universe of liblogs explored in this book. The next few chapters explore activity within those blogs during two three-month periods: March, April and May 2007 and March, April and May 2008.

Why March-May?

When measuring blog activity, you need to strike a balance between using a long enough period to be meaningful—and having a short enough period so it's feasible to do the metrics. Three months seems to be a reasonably good period.

You want a three-month period when most bloggers are reasonably likely to be active, regardless of the type of library or activity. That argues against the summer months (June-September) and heavy vacation months (November and December).

January's problematic, both because it's somewhat of a vacation month and because of ALA Midwinter, one of two megaconferences that can skew blogging activity. That leaves two possible three-month slots: February through April or March through May. The latter period is a little longer, and in the U.S. each of those three-month periods has one three-day weekend, so I chose the longer quarter. (Yes, there are significant library conferences in the March-May period, but that's true for every quarter.)

The Missing Eighteen

Eighteen liblogs appear in the age notes and have profiles in the Liblog Profiles chapter—but don't factor into other chapters because they have no metrics other than age.

➢ **Archivalia**: Non-English but in 2006 survey. A *very* active blog, with 11 posts on the day I checked—but I could find no way to gather metrics from the archives.

- ➢ **Conservator**: In earlier survey. No longer available.
- ➢ **The Digital Librarian Redux**: Hidden posts, unable to prepare metrics.
- ➢ **Digital Reference**: Active, but with no apparent archives or way to page back.
- ➢ **Electric Forest, Endless Hybrids, information literacy by example, Keeping Current**: May be active, no posts in quarters.
- ➢ **EngLib, InfoCommuner's Info Harbor, The Library Diva, Theoretical Librarian**: In earlier survey, no post in quarters.
- ➢ **The Green Kangaroo, Librarians with Class**: Clearly active but with no posts in quarters.
- ➢ **Infogestores**: Non-English but in 2006 survey. Very active (perhaps eight posts per day), but with two posts per archive page, it was implausible to do metrics.
- ➢ **LawLibTech**: Active, but with no apparent provisions for paging back or going through the archive in chronological order.
- ➢ **LibraryCog**: In earlier survey, no visible archives.
- ➢ **Social Software in Libraries**: Posts appear to have been replaced by a rolling set of bookmarks.

The group includes two *very* active non-English blogs where I couldn't (or wouldn't) do the metrics, three other active blogs where I couldn't do the metrics, and 13 blogs that just didn't have posts in either quarter or have since disappeared.

How Many Posts?

How often do posts appear on a blog?

Until feeds and aggregators became common, that was an important question. If you didn't provide a reasonably steady stream of posts, people wouldn't have reason to come back to your blog or bookmark it. Few posts, few readers. Some people advised trying to do at least one post a day. Others offered slightly less strenuous advice.

These days, when most readers see posts indirectly, a steady stream of posts is only important for certain kinds of blogs. Indeed, too many posts can work *against* readership, particularly if posts appear to be there "for the sake of posting."

There's no question that posting frequency among the studied liblog universe is lower in March-May 2008 than in March-May 2007. That doesn't mean it's down for every blog.

Quintiles for 2007

In all, 523 liblogs had countable posts in March-May 2007, ranging from one post over the 13 weeks (92 days) to 1,161 posts.

Posts	High	Low	Blogs	Median	Total	%
All	1,161	1	523	22	22,969	
Q1	1,161	57	105	83	14,438	63%
Q2	56	31	103	42	4,397	19%
Q3	30	17	107	22	2,464	11%
Q4	16	8	106	12	1,242	5%
Q5	7	1	102	4	428	2%

Table 2.1: Post quintiles for 2007

If you prefer to look at frequency:
- ➢ Blogs with the most posts, Quintile 1 (Q1), ranged from 4.4 posts per week to 89 per week or 13 per day, with a median of 6.4 posts per week.
- ➢ Blogs with more posts, Quintile 2 (Q2), had anywhere from 2.4 posts per week to 4.3 per week, with a median of 3.2 per week.
- ➢ "Average blogs," Quintile 3 (Q3), ranged from 1.3 posts per week to 2.3 per week, with a median of 1.7 per week. Thus, "roughly twice a week" was indeed an average posting frequency for liblogs in 2007—a little above the median but well within the middle quintile.
- ➢ Blogs with fewer posts, Quintile 4 (Q4), ranged from 0.6 posts per week to 1.2 per week, with a median of just under one a week.
- ➢ Blogs with fewest posts, Quintile 5 (Q5), ranged from one post during the entire quarter to just over one post every two weeks, with a median of just under one post every *three* weeks.

Quintiles for 2008

In March-May 2008, 533 liblogs had countable posts.

Posts	High	Low	Blogs	Median	Total	%
All	919	1	533	18	19,616	
Q1	919	45	108	72	12,668	65%
Q2	44	26	108	33	3,682	19%
Q3	25	14	105	18	1,957	10%
Q4	13	6	106	10	937	5%
Q5	5	1	106	3	312	2%

Table 2.2: Post quintiles for 2008

- ➢ Q1 blogs ranged from 3.5 posts per week to 71 per week or 10 per day, with a median of 5.5 posts per week.
- ➢ Q2 blogs ranged from two posts per week to 3.4 per week, with a median of 2.5 posts per week.
- ➢ Q3 blogs ranged from just over one post per week to just under two per week, with a median of 1.4 posts per week.

- ➢ Q4 blogs ranged from just under one post every *other* week to one per week, with a median of 0.8 posts per week.
- ➢ Q5 blogs ranged from one post for the quarter to 0.4 posts per week, with a median of one post *per month*).
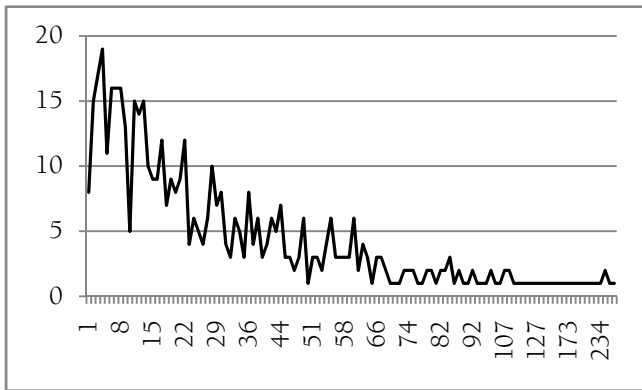

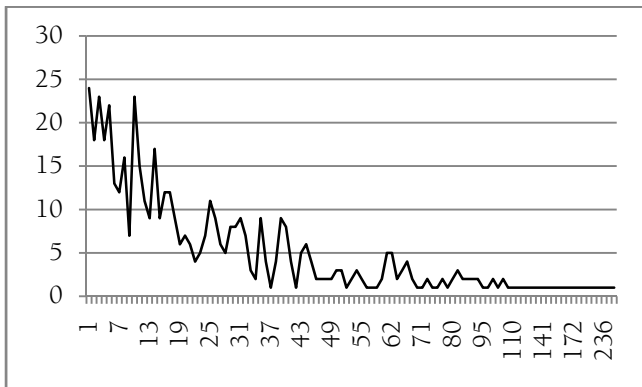*Figure 2.1: Blogs by number of posts, 2007*

Figure 2.2: Blogs by number of posts, 2008

Figures 2.1 and 2.2, presented together to allow comparison, omit blogs with more than 300 posts for the sake of clarity. The horizontal axis shows number of posts and only includes cases with at least one blog; the vertical axis shows number of blogs having that many posts.

Most Prolific Blogs in 2008

The blogs in Q1 for 2008 follow.

More than two posts per day (12)

Blog name	Count
ResourceShelf	919
Open Access News	791
beSpacific	770
Peter Scott's Library Blog	737
Library Stuff	343
Tame the Web	249
The Kept-Up Academic Librarian	241
Library Chronicles	236
Slaw	234
Collecting my Thoughts	224
LibrarianInBlack	216
Free Government Information (FGI)	195

One to two posts per day (24)

A Fuse #8 Production	176
bookshelves of doom	172
The Rabid Librarian's Ravings in the Wind	170
Beyond the Job	164
What I Learned Today...	163
Stephen's Lighthouse	159
DigitalKoans	157
Library Mistress	141
Library Boy	137
Travelin' Librarian	134
Text & Blog	121
Bad Librarianship Now!	118
Phil Bradley's weblog	111
Information Junk	110
UBC Academic Search - Google Scholar Blog	108
the.effing.librarian	104
iLibrarian	104
:: CultureLibre.ca ::	101
Digitization 101	98
Head Tale	98
PLA Blog	96
The Blue Skunk Blog	95
Bluestalking Reader	92
Library Link of the Day	92

Six posts per week and less than one per day (11)

affordance.info	87
j's scratchpad	87
Information Literacy Weblog	86
LibrariAnne	86
Killin' time being lazy	84
The Reader's Advisor Online Blog	84
Catalogablog	82
Jennie Law	82
OUuseful.Info, the blog	82
Kids Lit	80
Speak Quietly: Ramblings…	80

Five posts per week and less than six per week (15)

Hedgehog Librarian: Prickly, Nocturnal, InfoDiva	77
Family Man Librarian	75
Library Technology in Texas	75
WoW! Wouter over het Web	74
Lorcan Dempsey's weblog	73
Laurie the Librarian	72
zydeco fish	72
Libraries & Life	71
Atomic Librarian	69
BlogJunction	68
Walt at Random	68
Caveat Lector	65
Government Info Pro	65
Libraries in the NHS	65
Typo of the day for librarians	65

Four posts per week, less than five per week (28)

3.46 to 3.9 posts per week (18)

Changes from 2007 to 2008

Comparing the two sets of quintiles, you can see broad changes from 2007 to 2008—noting that the two universes are different:

➢ Overall, there are 9% fewer posts.

➢ The median for Q1 is 13% lower, and the most prolific and least prolific blogs each have 21% fewer posts.

➢ The median for Q2 and the most prolific blog are 21% lower; the least prolific of the quintile is 16% lower.

➢ The median for Q3 is 18% lower; the most prolific blog has 17% fewer posts and the least prolific has 18% fewer.

➢ The median for Q4 is 17% lower; the most prolific blog has 19% fewer posts and the least prolific has 25% fewer.

➢ You can't get lower than one and still be in the universe, so the least prolific blogs for 2008 have the same number of posts—but there are more of them. The median is 25% lower and the most prolific level is 29% lower.

So do liblogs post less frequently in 2008 than in 2007?

Yes—taken in large groups. But blogs are individual entities. Let's look at quintiles for the *difference* between 2007 and 2008.

Change	High	Low	Blogs	Median	Posts
All	2500%	-100%	523	-34%	17,881
Q1	2500%	24%	105	63%	5,626
Q2	23%	-22%	105	-1%	7,315
Q3	-23%	-48%	105	-35%	2,813
Q4	-49%	-75%	104	-62%	1,904
Q5	-76%	-100%	104	-100%	223

Table 2.3: Change in posting frequency from 2007 to 2008

The universe in this case is the same as for 2007 posts—all blogs with at least one post in March-May 2007, including those with no posts in March-May 2008 (which show as -100%).

Here we see just how varied the landscape really is—and I've included the total posts within each quintile to indicate that it's not entirely a question of posting frequency, although that does enter into it.

➢ Q1 represents a significant increase in posts from 2007 to 2008—at least 20% (actually a little more than that, but there are only three blogs in Q2 with 20% or more increase). Incidentally, none of the blogs with extremely high growth percentages had only single posts in 2007, although one did grow from two posts to 35 posts. (The blogs with a single post in 2007 had either one post in 2008 or none.)

➢ Q2 neatly represents "relatively unchanged" blogs—that is, those that neither increased nor decreased by more than 20% or so. As a group, these are more prolific blogs, with more than two-fifths of all posts in one-fifth of the blogs (as compared to one-third of all posts in Q1). If you're wondering, 18 blogs have exactly the

same number of posts in March-May 2008 as in March-May 2007, and in one case (Library Link of the Day, with exactly 92 posts in a 92-day quarter) that's clearly deliberate.

➢ Q3, the "average blog" quintile, shows a decrease in posts of anywhere from just under one-quarter to just under one-half from 2007 to 2008. This fifth of the universe had one-sixth of the posts.

➢ Q4, more decrease than average, decreased by roughly half to three-quarters, a quite substantial decrease. Not surprisingly, these blogs are down to one-tenth of all posts.

➢ The blogs in Q5—most decrease—represent two categories: Those with very substantial decreases but some posts in 2008, and those that had no posts at all in 2008. As the median shows, more than half are in the latter category, but not a lot more than half—56 of the 104. The other 48, still 9% of the universe, only provided 1% of the posts, although a few of them still had a fair number of posts (including one blog with 25 posts that had 107 in 2007).

The other sixtysix

What of the 66 blogs that had posts in March-May 2008 but not in March-May 2007?

There are two distinct subgroups within that group. The first is blogs that began in June through December 2007—they simply were not around for the 2007 metrics. That's 49 of the 66 or 74%, totaling 1,252 posts in 2008—interestingly, an average of two posts a week (26 over the quarter).

The other 17? Eight have archive problems—either the archive was lost in a domain change, the archive is unreachable, or it was simply unworkable for metrics. Two were on formal hiatus during the period. The other seven just didn't have any posts in that quarter. Notably *none* of the 17 was a "newbie"—only two were less than three years old and one is one of the oldest liblogs around. Those blogs account for 483 posts in 2008, one-third of them in one blog.

Growth Blogs

This list includes Q1 blogs with at least one post every two weeks in 2007—79 of the 105 blogs in Q1. (Note: Order in the list is in actual decreasing percentages, including fractional percentages.)

Blog   Growth from 2007 to 2008

Fewer Posts in Most Blogs

We can safely say most liblogs had significantly fewer posts in 2008 than in 2007—roughly three out of five had a drop of more than 20%.

But most isn't all. The list of blogs with real growth is a long and impressive one, as you can see.

Would it make much difference if we disregarded ceased blogs in considering changes? Not really. Only seven blogs with posts in 2007 explicitly ceased by March 2008. A dozen or so are *clearly* still active, with posts since June 2008. One blog (one of the most active in 2007) is a special case: The 2008 archives are unworkable. If you remove all blogs with no posts in March-May 2008 except those that have had posts since June 1, 2008, you arrive at the following:

Change	High	Low	Count	Median	Posts
All	2500%	-100%	479	-30%	17,881
Q1	2500%	29%	96	74%	4,226
Q2	28%	-18%	99	0%	6,911
Q3	-19%	-43%	95	-31%	4,145
Q4	-44%	-67%	98	-54%	2,050
Q5	-68%	-100%	90	-83%	523

Table 2.4: Change in posting frequency ignoring some blogs

This isn't a significant change—and I'd guess almost half of the blogs removed for this analysis are still active. Only three of the blogs in Q3 show a 19% decrease, so it's still fair to say nearly 60% of blogs had significant decreases—at least 20%.

How do the most prolific blogs fare in terms of frequency change? Of 108 Q1 blogs for 2008, nine were not part of the 2007 counts. Of the rest, 44 are in change Q1, with increases of 24% or more. 36 are in Q2, relatively little change overall. 15 have mod-

erate decreases in frequency (Q3). The other four are in Q4, with quite substantial decreases in frequency.

How Long?

Liblogs vary enormously in typical length of posts as well as number of posts. Some bloggers specialize in essays; some rarely blog more than a few sentences at a time.

The universe for this chapter is, once again, smaller—for unfortunate reasons. For 2007 length calculations, there were 48 blogs with posts for which I couldn't determine length. For 2008, there are 45. Most blogs have the same problems in both years. In most cases, the problem has to do with hidden posts—you don't see most or all of the post without expanding each one. A few cases are a little different: A WordPress template that makes it nearly impossible to do the work needed to determine length over three months.

We discuss overall length first. Much more interesting is the length per post. Even more than frequency, these are observations, not judgments: There's no "right length" for blog posts.

Where blogs have large amounts of link-network overhead for each post, I deducted those lengths. Typical post overhead remains.

Blog Length Quintiles for 2007

I was able to calculate total post lengths for 475 blogs for March-May 2007. Except for blogs and percentages, all figures are word counts.

	High	Low	Blogs	Median	Total
All	186,467	26	475	5,346	5,133,681
Q1	186,467	15,650	95	26,023	3,239,742
Q2	15,500	7,391	95	10,725	1,032,261
Q3	7,379	3,900	96	5,288	528,081
Q4	3,874	1,688	95	2,684	251,848
Q5	1,670	26	94	877	81,794

Table 3.1: Total word length quintiles for March-May 2007

This table should be self-explanatory—e.g., the shortest blog in the longest quintile (Q1) had 15,650 words while the longest had 186,467; the median was 26,023 words; 95 blogs fell into this category; they total 3.2 million words—63% of the total for all 475 blogs.

Breaking down Q1, no blog exceeded 200,000 words for the quarter but four exceeded 100,000 words. Six had 50,000 to 90,000 words. Forty had 25,000 to 48,000 words, with nine right around the median (25,000 to 26,999 words). That leaves 45 blogs with 15,650 to 24,800 words.

Blog Length Quintiles for 2008

I was able to calculate lengths for 489 blogs for March-May 2008.

	High	Low	Blogs	Median	Total
All	204,517	39	489	4,725	4,483,897
Q1	204,517	14,040	98	20,446	2,773,455
Q2	14,000	7,280	98	9,250	961,938
Q3	7,248	3,255	98	4,689	478,724
Q4	3,254	1,286	98	2,110	210,047
Q5	1,278	39	97	574	59,733

Table 3.2: Total word length quintiles for March-May 2008

The differences from 2007 are subtle.
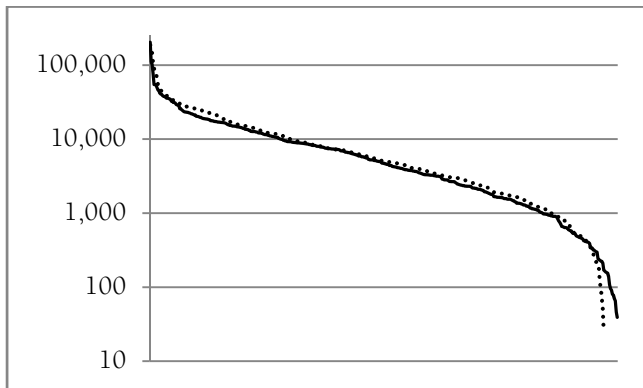
Figure 3.1: Word counts for liblogs, 2007 and 2008

Figure 3.1 shows word counts (vertical axis) for each blog in both years, plotted on a logarithmic scale. (The horizontal axis is just one increment for each blog.) The two lines don't differ much and it's a smooth curve for blogs with roughly 800 to 40,000 words.

Longest Blogs in 2008

These blogs had the most content (excluding some overhead) for March-May 2008, noting that some blogs couldn't be measured.

Open Access News 204,517
ResourceShelf 112,822
beSpacific 102,087
Collecting my Thoughts 77,287
Library clips 54,857
Free Government Information (FGI) 54,800
Slaw 53,800
Beyond the Job 50,177
The Blue Skunk Blog 46,495
Tame the Web 45,876
LibrarianInBlack 42,035
Peter Scott's Library Blog 40,699
The Rabid Librarian's Ravings in the Wind 40,277
The Reader's Advisor Online Blog 38,788
bookshelves of doom 37,980
Library Chronicles 37,727
Confessions of a Science Librarian 36,841

What I Learned Today... 36,009
Library Boy 35,714
Tales from the "Liberry" 35,380
the.effing.librarian 35,205
PLA Blog 34,636
DigitalKoans 32,612
Walt at Random 32,493
Stephen's Lighthouse 31,477
Bluestalking Reader 31,452
Caveat Lector 30,977
Digitization 101 29,593
The Kept-Up Academic Librarian 29,184
Museum 2.0 29,069
Academic Librarian 28,440
Off the Mark 25,820
UBC Academic Search - Google Scholar Blog 25,707
Librarian 24,929
The Gypsy Librarian 24,374
Misadventures of the Monster Librarian 23,748
The Misadventures of Super_Librarian 23,357
eFoundations 23,297
Head Tale 23,216
The Zenformation Professional 23,025
Free Range Librarian 22,930
Pop Culture Librarian 22,586
UK Web Focus 22,438
Nodalities 22,056
Travelin' Librarian 21,993
Attempting Elegance 21,580
Connie Crosby 21,340
Hedgehog Librarian: Prickly, Nocturnal, InfoDiva ... 21,308
Text & Blog 20,480
Pop Goes the Library 20,412
ricklibrarian 20,297
j's scratchpad 20,068
Informationoverlord 19,810
Rambling Librarian :: Incidental Thoughts of a Singapore Liblogarian 19,753
the log of a librarian 19,260
Silversprite 19,087
Library Stuff 19,000
Jennie Law 18,873
Librarilly Blonde 18,830
Library Mistress 18,752
Speak Quietly: Ramblings About Libraries 18,684
Killin' time being lazy 18,424
EBM and Clinical Support Librarians@UCHC 18,324
Not So Distant Future 17,803
The Information Literacy Land of Confusion 17,783
johnmiedema.ca 17,670
The Krafty Librarian 17,654
Swiss Army Librarian 17,374
Government Info Pro 17,336
Phil Bradley's weblog 17,234
Kids Lit 17,120
zydeco fish 17,053

Blog Length Changes 2007-2008

As with posts, length for each blog changes individually, not in lockstep. Blogs with zero length in 2008 are included only if there were also no posts in March-May 2008. That yields a universe of 469 blogs.

Change	High	Low	Blogs	Median	Words
All	2388%	-100%	469	-32%	3,942,801
Q1	2388%	38%	94	94%	920,314
Q2	37%	-16%	94	5%	1,654,845
Q3	-17%	-47%	92	-32%	954,195
Q4	-48%	-79%	94	-62%	380,499
Q5	-63%	-100%	95	-100%	32,948

Figure 3.3: Changes in blog length from 2007 to 2008

As with posts, the biggest gathering of text is in Q2—which in this case includes blogs that increased significantly as well as most of those that didn't change much.

Most blogs had significantly fewer words in 2008 than 2007, but quite a few—111 in all—had at least 20% *more* text in 2008.

Overall length is a secondary factor in blogs. Still, when an individual writes a lot more, it's interesting. Here are the Q1 blogs—those that gained the most in length—that meet two other criteria: They're not group blogs and they had more than 3,000 words in 2007:

If the sort order seems strange, note that these blogs are in descending order of increase by percentage (stored to several decimal places)—and when there is an absolute tie, longer blogs appear first. If "miscellaneous" is the first word that comes to mind when you scan this list, which certainly isn't entirely a list of the best-known liblogs, I wouldn't be surprised.

Blog length is generally less interesting than *post* length—the average words per post within a blog. (Ideally, you'd want to look at distribution of post lengths within each blog, but that's an implausible metric to calculate or report for a real-world study.)

Post Length Quintiles for 2007

The universe for words per post is the same as the universe for total length: 475 blogs for 2007.

Length	High	Low	Blogs	Median	Words
All	1,300	10	475	254	5,133,681
Q1	1,300	393	95	496	1,338,396
Q2	392	285	95	329	1,188,526
Q3	284	214	94	254	1,125,710
Q4	213	146	95	175	969,004
Q5	145	10	96	110	524,181

Table 3.4: Post length quintiles for 2007

The three central quintiles are small ranges—in each case, the low number is at least 68% of the high number. If you characterized blogs in those quintiles as having "around 330 words per post," "around 250 words per post," and "around 175 words per post" respectively, you wouldn't be far off. You could also state those as roughly two paragraphs, one long paragraph, and one paragraph.

The extremes may be worth breaking down. At the high end are a dozen blogs with posts averaging "column length"—700 words or more. Another nine have at least 600 words per post, and another 22 run 500 words or more.

Blogs with terse posts tend to be linkblogs or have a mix of links and other short posts—or possibly many very short posts and a few long ones. The shortest posts belong to a fairly special blog, *Library link of the day*—and an average of ten words per post makes sense when you're just providing a link.

Post Length Quintiles for 2008

The universe for post length for 2008 is 489 blogs.

Length	High	Low	Blogs	Median	Words
All	1,645	11	489	245	4,483,897
Q1	1,645	424	98	560	1,268,836
Q2	422	289	97	340	905,900
Q3	288	217	98	246	1,117,454
Q4	214	147	98	188	631,302
Q5	146	11	98	108	560,405

Table 3.5: Post length quintiles for 2008
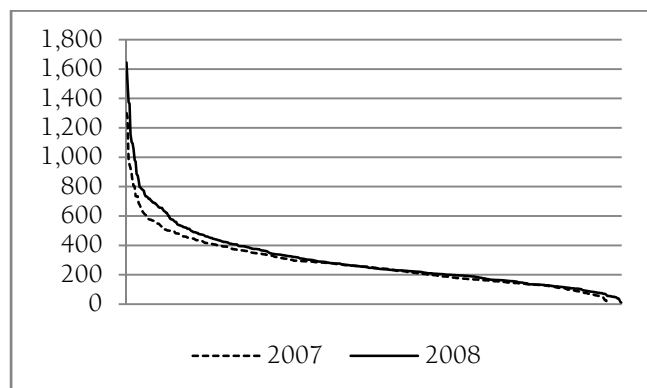
Figure 3.2: Words per post, 2007 and 2008

Unlike the graph for blog length, Figure 3.2 is linear. The pattern is interesting: In the upper midrange, posts are significantly longer in 2008—but in the lower ranges, there's no real difference.

In the case of post length, the top and bottom quintile are both interesting. The lists below include the top portion of the top quintile and the bottom portion of the bottom quintile—blogs averaging at least 500 words per post or no more than 100 words per post.

Blogs with Essay-Length Posts in 2008

The Zenformation Professional 1,645
Shelved in the W's 1,515
the log of a librarian 1,376
A LIBRARIAN AT THE KITCHEN TABLE 1,363
John Wilkin's Blog 1,157
REAL PUBLIC LIBRARIAN 1,103
Academic Librarian 1,094
Library clips 1,055
The PlanetEsme Plan 981
Museum 2.0 969
The Gay Librarian 887
Metalogger 874
The Gypsy Librarian 840
LibraryTechNZ 797
Librarian on the edge 793
Off the Mark 782
Pop Culture Librarian.. 779
Thoughts from Carl Grant 772
shimenawa 746
Obnoxious Librarian from Hades 734
always learning.. 732
Annoyed Librarian 727
Gather No Dust 717
Larocque and Roll 716
Open Libraries 707
Circ and Serve 700
T. Scott 693
nirak.net - Musings of an LIS Student 690
Loriene's Blog 688
MemberBlog.. 680
The Other Librarian 675
the pod bay door... 667
Rambling Librarian :: Incidental Thoughts of a Singapore Liblogarian… 658
Info Career Trends... 656
Everybody's Libraries... 656
Guardienne of the Tomes 655
Mary Ellen Bates - Info-Entrepreneur Tip of the Month. 643
Scholarly Electronic Publishing Weblog 634
The Singing Librarian Talks (or Writes…) 631
The Vital Library 622

Blogs with Terse Posts in 2008

More blogs have *very* long posts in 2008: eight average more than 1,000 words per post as compared to two in 2007 (the longest of which didn't have posts in 2008), and 13 average more than 800 words per post as compared to eight in 2007.

Post Length Changes 2007-2008

Here's how the average post length of various blogs changed from 2007 to 2008:

Change	High	Low	Blogs	Median	Words
All	657%	-100%	469	-2%	3,942,801
Q1	657%	38%	95	70%	714,427
Q2	37%	9%	94	17%	1,180,332
Q3	8%	-11%	90	-2%	1,172,045
Q4	-12%	-39%	95	-25%	676,108
Q5	-40%	-100%	95	-96%	199,889

Table 3.6: Changes in post length from 2007 to 2008

Did most blogs have longer posts in 2008 than in 2007? No—but it's close. 135 of the 469 blogs for which this could be measured—29%—have significantly longer posts and 217 (46%) have longer posts. If you exclude blogs with no posts in 2008, 32% of the remaining blogs have significantly longer posts and 51% have longer posts.

Blogs with Longer Posts

Here, again, we'll limit the list to blogs in Q1 (most increase in average length) that also had at least one post every two weeks (at least seven posts) during March-May 2007.

Conversations

Some blogs don't have comments—either because the blogger doesn't allow them or because the posts don't attract comments. And, of course, there are some blogs where I couldn't determine the number of comments—although there are also blogs where I couldn't track length but *could* count comments.

I don't distinguish between blogs that don't allow comments and blogs that just don't *have* any comments. Instead, quintiles include only blogs with one or more comments. As with length, in some ways raw counts are less interesting than conversational intensity—comments per post—but let's look at raw counts first.

Comment Quintiles for 2007

Comments showed up and were countable in 439 blogs for March-May 2007. Except for blogs and percentages, all figures in the table are comment counts (including totals).

Comments	High	Low	Blogs	Median	Total
All	1,689	1	439	19	27,009
Q1	1,689	69	88	162	20,462
Q2	68	28	90	42	3,933
Q3	27	11	93	16	1,638
Q4	10	5	85	7	603
Q5	4	1	83	2	193

Table 4.1: Comment quintiles for March-May 2007

In this case, we're very close to a 20:80 situation: The top 20% of blogs included 76% of all comments. If the sizes of quintiles vary more than you'd expect, that's because of the need to break quintiles at number breaks—specifically, Q3 had to be either somewhat larger than a pure 20% or quite a bit smaller.

Breaking down Q1 a little, one blog had more than 1,000 comments (*A Fuse #8 Production*) and also more than 1,500. Two more blogs had more than 900 comments, with another between 800 and 900 and another between 500 and 600—a total of five blogs with more than 500 comments each. Seven have 400 to 471 comments, seven more 301 to 358, and 18 more between 202 and 288.

Comment Quintiles for 2008

The universe of blogs with *countable* comments for 2008 is different and slightly smaller: 433 blogs. (In a couple of cases, comments could be counted for 2007 but couldn't be for 2008.)

Comments	High	Low	Blogs	Median	Total
All	1,219	1	433	18	22,600
Q1	1,219	65	87	122	16,441
Q2	63	27	85	42	3,622
Q3	26	13	90	18	1,634
Q4	12	5	91	8	716
Q5	4	1	80	2	187

Table 4.2: Comment quintiles for March-May 2008
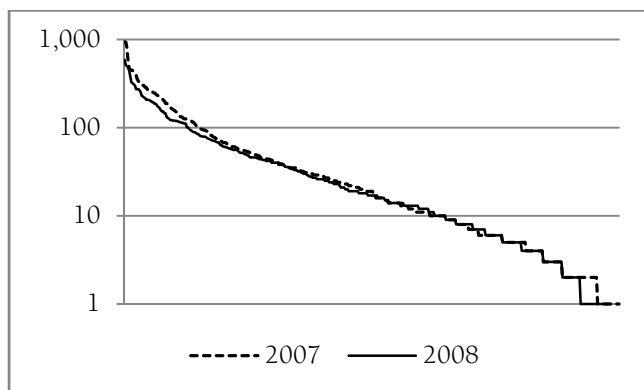
Figure 4.1: Comments per liblog, 2007 and 2008

Figure 4.1 shows the number of comments for each blog for both 2007 and 2008 on a logarithmic scale. The "steps" toward the right show multiple blogs for a given number of comments—the horizontal axis has one point for each blog. (The highest blog for 2007 and the two highest blogs for 2008 are off the chart.)

Blogs with Most Comments in 2008

The blog with the most comments in 2007—nearly 400 more than any blog in 2008—changed platforms, moving to the LJ/SLJ blog platform, which made it impractical to count comments for 2008.

Comment Changes 2007-2008

In order to calculate changes plausibly, I omitted not only all blogs with no comments in 2007, but also those that had neither comments nor posts in 2008 (or that had posts but no lengths, indicating a metrics problem). The resulting universe includes 385 blogs.

Change	High	Low	Blogs	Median	Total
All	4167%	-100%	385	-32%	20,290
Q1	4167%	75%	79	155%	6,326
Q2	71%	-11%	76	24%	6,015
Q3	-13%	-47%	77	-33%	4,813
Q4	-48%	-74%	72	-62%	2,723
Q5	-75%	-100%	81	-91%	413

Table 4.3. Changes in comments per blog from 2007 to 2008

Here are the liblogs in Q1 that had at least five comments in 2007—in descending order by percentage increase, and by descending order in number of 2007 comments when there are ties.

Conversational Intensity Quintiles for 2007

I define "conversational intensity" as number of comments per post, a more interesting number than the total number of comments. As you might expect, the universe for 2007 conversational intensity is the same as for total comments: 439 blogs.

CI	High	Low	Blogs	Median	Comments
All	28.86	0.03	439	0.90	27,009
Q1	28.86	2.17	88	3.70	16,887
Q2	2.13	1.05	87	1.41	5,541
Q3	1.04	0.64	88	0.86	2,550
Q4	0.63	0.31	89	0.43	1,400
Q5	0.30	0.03	87	0.15	631

Table 4.4: Conversational intensity (comments per post), 2007

Showing these figures to two decimal places is a little ridiculous—but otherwise, the ranges of identical figures are too broad to make useful quintiles possible.

Blogs with no countable comments were ignored—so it would be misleading to say that the average liblog had 0.9 comments per post in 2007. It may be worth listing the small number of liblogs that averaged more than ten comments per post in 2007 (a

very high conversational intensity for liblogs, although low compared to some other blogs), particularly since only four of these six achieved similar heights in 2008—and one nearly tripled its intensity in 2008.

Conversational Intensity Quintiles for 2008

The universe here is 433 blogs—and note that the medians for Q2 through Q5 are all significantly higher than for 2007 (the Q1 median is higher, but by less than 10%).

CI	High	Low	Blogs	Median	Comments
All	53.00	0.01	433	1.00	22,600
Q1	53.00	2.55	86	4.00	12,468
Q2	2.46	1.31	87	1.71	4,693
Q3	1.30	0.71	86	1.00	3,449
Q4	0.70	0.39	86	0.56	1,298
Q5	0.38	0.01	88	0.18	692

Table 4.5: Conversational intensity (comments per post), 2008
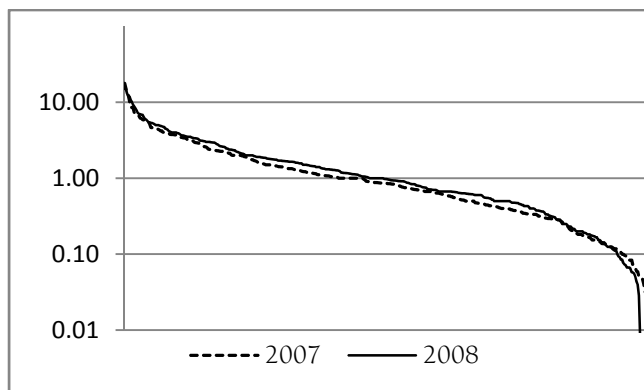
Figure 4.2: Conversational intensity (comments per post), 2007 and 2008

Figure 4.2 omits the single blog in each year with more than 20 comments per post and is again logarithmic to show detail. Note that the 2008 line is slightly but significantly above the 2007 line through most of the range—but, as usual, that doesn't show what happened with each blog.

Blogs with Most Conversational Intensity, 2008

When two or more blogs have the same conversational intensity (possibly including further decimal digits), they're listed in decreasing number of comments in 2008.

Changes in Conversational Intensity

As with changes in comment counts, this table omits blogs with no 2008 comments that also had either no counted posts or no calculated length, as in either case metrics problems could be involved.

Change	High	Low	Blogs	Median	Comments
All	1868%	-100%	385	-4%	20,290
Q1	1868%	102%	77	211%	6,164
Q2	100%	20%	77	54%	5,098
Q3	19%	-19%	75	-3%	4,462
Q4	-20%	-54%	78	-33%	3,533
Q5	-55%	-100%	78	-84%	1,033

Table 4.6: Changes in conversational intensity, 2007 to 2008

As a flat statement, "conversational intensity increased" is wrong—but it's right for 40% of the blogs. That is, Q2 begins at the lower limit of significantly increased conversational intensity, 20%. This table is also interesting because Q3, the average blogs, *exactly* encompasses what I consider to be "insignificant" change—anywhere from 19% to -19%. Just over half

the blogs had less conversational intensity (51%). The distribution of comments among the quintiles is also interesting—the changes in Q1 through Q4 are smaller than for most measures.

Looking at post length, we found that 60% of the blogs showed significant change. For conversational intensity, that goes up to 80%.

Liblogs Growing Most in Conversational Intensity

Here are liblogs that grew most (Q1) and that had at least five comments in 2007.

Remember: 102% is the *growth* percentage: It means the blog had slightly more than twice as many comments per post in 2008 as it did in 2007.

Getting the Picture

Some bloggers never use visuals as part of individual posts (ignoring such things as LiveJournal icons and social networking symbols). Many others rarely do. In a way, this chapter is a sideshow—we won't be looking at figures in later chapters—but it may be interesting.

"Figures" includes videos, drawings, charts, corporate logos that relate to an individual post, screenshots, book covers—what have you. In some cases, the counts are low for three reasons:

➢ I counted an autogenerated array of overlapping book jackets as a single picture—but a rectangular array as one picture per row.
➢ When there are multiple figures on a single row, I count them all as one.
➢ When figures join or overlap vertically in a way that makes it unclear whether they're part of one picture, I assume they are.

So, for example, a display of book jackets that shows five jackets per row and has 18 rows counts as 18 figures, not 90.

Figure Quintiles for 2007

I was able to count figures in 348 blogs in March-May 2007.

Figures	High	Low	Blogs	Median	Total	%
All	245	1	348	7	6,385	
Q1	245	27	70	46	4,458	70%
Q2	26	10	73	14	1,141	18%
Q3	9	6	53	7	381	6%
Q4	5	3	71	4	282	4%
Q5	2	1	81	2	123	2%

Table 4.7: Figures or illustrations in blogs for March-May 2007

Breaking down Q1, there were two blogs with more than 200 figures and nine more with 103 to 165 figures.

Figure Quintiles for 2008

Figures	High	Low	Blogs	Median	Total	%
All	172	1	361	7	6,092	
Q1	172	25	74	41	3,953	65%
Q2	24	12	73	17	1,272	21%
Q3	11	6	56	8	449	7%
Q4	5	3	80	4	307	5%
Q5	2	1	78	1	111	2%

Table 4.8: Figures in blogs for March-May 2008

Blogs with Most Figures, 2008

Changes in Figures 2007-2008

Table 4.9 shows changes from 2007 to 2008 for 312 blogs that had figures in 2007 and for which figure counts for 2008 are reliable.

Change	High	Low	Blogs	Median	Figures
All	1400%	-100%	312	-20%	5,194
Q1	1400%	83%	63	208%	1,165
Q2	82%	4%	56	31%	1,770
Q3	0%	-33%	63	-18%	1,396
Q4	-34%	-73%	62	-50%	783
Q5	-75%	-100%	68	-100%	80

Table 4.9: Changes in figures from 2007 to 2008

Here are liblogs with the most growth in figures from 2007 to 2008 that had at least five figures in 2007:

Figures per Post Quintiles, 2007

Figure 4.10 shows figures per post for the 348 posts for which figures could be measured in 2007.

F/Post	High	Low	Blogs	Median	Figures
All	20.67	0.01	348	0.32	6,385
Q1	20.67	0.69	71	1.00	2,781
Q2	0.67	0.40	75	0.52	2,207
Q3	0.39	0.20	74	0.29	776
Q4	0.19	0.09	68	0.15	425
Q5	0.08	0.05	61	0.05	196

Figure 4.10: Figures per post, March-May 2007

Only one blog had more than seven figures per post, and that blog only had three posts (including 62 figures). Six more had at least two figures per post, including one more with more than five.

Figures per Post Quintiles, 2008

F/Post	High	Low	Blogs	Me- dian	Figures
All	6.18	0.01	361	0.39	6,092
Q1	6.18	1.00	75	1.19	2,547
Q2	0.95	0.53	72	0.70	1,731
Q3	0.52	0.31	68	0.40	2,609
Q4	0.30	0.15	73	0.21	679
Q5	0.14	0.01	74	0.07	257

Figure 4.11: Figures per post, March-May 2008

Blogs with Most Figures per Post, 2008

The PlanetEsme Plan 6.18

Changes in Figures per Post

Change	High	Low	Blogs	Median	Figures
All	2441%	-100%	312	8%	5,194
Q1	2441%	114%	63	267%	1,007
Q2	112%	40%	63	63%	1,259
Q3	39%	-7%	65	8%	1,843
Q4	-9%	-59%	62	-29%	1,030
Q5	-60%	-100%	59	-100%	55

Figure 4.12: Changes in figures per post, 2007-2008

You can definitely say that most liblogs that had figures in 2007 had more figures per post in 2008: only 42% had fewer figures per post, and nearly 47% had at least 20% more figures per post. By and large, bloggers who use visuals at all are tending to use them more often.

Finally for this chapter, here are the liblogs with the greatest increase in figures per post that also had at least five figures in 2007.

omg tuna is kewl 1192%
"Self-plagiarism is style"...504%
lauren's library blog 367%
affordance.info..353%
CogSci Librarian 308%
Stephen's Lighthouse 283%
nirak.net - Musings of an LIS Student.........................281%
Connie Crosby...280%
Librarian 2..267%
Filipino Librarian..255%
Citegeist...250%
BookBitchBlog...245%
David's Random Stuff..192%
The Misadventures of Super_Librarian 188%
Grumpator...182%

Patterns of Change

So far, we've looked at individual facets of liblogs. Those are all interesting, but a blog is more than its individual facets.

This chapter looks at *patterns*—specifically, patterns of change, key to a lateral look at the landscape. Three elements make up the change pattern for a given blog:

➢ **Change in number of posts**: Were there more posts in 2008 than in 2007, fewer, or about the same number?

➢ **Change in post length**: Was the average post in a given blog longer in 2008 than in 2007, shorter, or about the same length?

➢ **Changes in comments per post**: Was the blog more conversational in 2008 than in 2007 (that is, did the average post have more comments), less conversational, or about the same?

I don't attempt to include figures (illustrations) in these change patterns partly for simplicity and partly because figures are, by and large, still secondary elements for most liblogs. Only about 60% of the blogs for which changes can be calculated have any figures at all—and only about 15% have even one figure per post.

If you found the tables in Chapters 1-5 formidable, take a deep breath before you turn the page. Table 6.1 shows the distribution of change patterns in liblogs, followed by a discussion of the table and what it shows. The short version: My naïve hypothesis is narrowly correct, in that the largest single cluster of liblogs does indeed have fewer posts, longer posts and is more conversational—but that oversimplifies the reality.

But Table 6.1 also oversimplifies the reality of change. It counts any upward change (including no change at all) as "longer" or "more" and any downward change, even 1%, as "shorter" or "fewer" or "less"—and that overstates the significance of small changes. More on that later in this chapter.

	Blogs	B%	Posts	P%
More Posts	**157**	**30%**	**8,919**	**50%**
Longer Posts	73	14%	4,250	24%
More Conversational	38	7%	1,507	8%
Less Conversational	29	6%	1,082	6%
No comment metrics	6	1%	1,661	9%
Shorter Posts	72	14%	4,105	23%
More Conversational	33	6%	1,832	10%
Less Conversational	31	6%	1,330	7%
No comment metrics	8	2%	943	5%
No length metrics	12	2%	564	3%
More Conversational	3	1%	174	1%
Less Conversational	2	0%	143	1%
No comment metrics	7	1%	247	1%
More Convers. subtotal	74	14%	3,513	20%
Less Convers. subtotal	62	12%	2,555	14%
Fewer Posts	**364**	**70%**	**8,962**	**50%**
Longer Posts	156	30%	4,784	27%
More Conversational	77	15%	2,356	13%
Less Conversational	65	12%	2,020	11%
No comment metrics	14	3%	408	2%
Shorter Posts	127	24%	3,462	19%
More Conversational	50	10%	1,782	10%
Less Conversational	68	13%	1,599	9%
No comment metrics	9	2%	81	0%
No length metrics	81	16%	716	4%
More Conversational	10	2%	260	1%
Less Conversational	5	1%	148	1%
No comment metrics	66	13%	308	2%
More Convers. subtotal	137	26%	4,398	25%
Less Convers. subtotal	138	26%	3,767	21%
Longer Posts total	229	44%	9,034	51%
Shorter Posts total	199	38%	7,567	42%
More Conversational total	211	40%	7,911	44%
Less Conversational total	200	38%	6,322	35%

Table 6.1: Change patterns in liblogs, 2007-2008

Understanding the Table

Table 6.1 shows patterns of change for each of 521 blogs (essentially all blogs with posts in 2007; two blogs fell out of pattern examination because of measurement problems).

Columns

The first column shows the particular pattern in nested form. So, for example, the third line (below headings) is for blogs with more posts, longer posts, and more comments per post ("more conversational").

The second column shows the number of blogs that fit that pattern. Third is the percentage of all blogs covered in this chapter.

The fourth column shows the number of posts in blogs that fit this pattern—included because 70 blogs each having one post over 13 weeks really aren't equivalent to 70 blogs averaging 26 posts each. The final column shows the fourth column as a percentage of all posts in blogs analyzed in this chapter.

Rows

The key point here is that **rows are nested**:

➢ "Longer Posts" is a subtotal of the three rows below it.
➢ "More Posts" is a subtotal of "Longer Posts," "Shorter Posts," and "No length metrics."
➢ The last two rows under "More Posts" and "Fewer Posts" offer a different slice, adding up all three "More conversational" or "Less conversational" lines respectively.
➢ The four rows at the bottom are also additional slices, showing the totals for longer and shorter posts and for more and less conversational blogs respectively.

Simplifications and modifications

In the handful of cases where there was no change from 2007 to 2008, I counted 0% as "more." That affects 18 blogs counted as "more posts" rather than "fewer posts" and one blog counted as "longer posts" rather than "shorter posts."

A couple of minor modifications from earlier chapters:

➢ Blogs that had no comments in 2007 but do have comments in 2008 are counted as "more conversational," where they were omitted from change quintiles (because it's not possible to calculate a percentage change from zero). Those that have no comments in either year are counted as "No comment metrics."
➢ Cases where there are no **posts** in 2008 are counted as "no length metrics" where they were included as -100% in change quintiles.

The net effect is that this chapter includes more changes in conversational intensity and fewer changes in post length. Are these the "right" choices? I can argue either way.

You can do a thought experiment to see why I didn't include figures, quite apart from their significance. It would increase the number of individual patterns from 18 to 54 and would considerably more than double the height of Table 6.1, along with making the labels extremely terse.

A Few Highlights

There are eighteen patterns for change in a blog, the eighteen rows with the most deeply indented labels. If patterns of change across the landscape were completely random, each row would have roughly 29 blogs and 993 posts and show 6% in each percentage column.

One pattern comes close to that "average": Blogs with more and longer posts but fewer comments per post—29 of them, with slightly more than 6% of all posts.

The outliers are more interesting:

➢ The single most common pattern is the "naïve hypothesis": Fewer posts, longer posts, more conversational. With 77 blogs and 2,356 posts, that pattern represents 15% of the blogs—but only 13% of the posts. That's hardly surprising: These blogs do have fewer posts in 2008 than in 2007.
➢ Oddly enough, the *next* most common pattern is the "discouraged" pattern: Fewer posts, shorter posts, less conversational. That pattern represents 68 blogs (13%) but only 9% of the posts.
➢ Right behind that comes one of two "not much information" patterns—fewer posts, no length metrics and no comment metrics (which most commonly means there were no comments in either year—it's frequently possible to count comments but not calculate length). That's an odd group of heavily diminished blogs: The 66 blogs average fewer than five posts each, so while they represent 13% of the blogs, they include only 2% of the posts.
➢ In fourth place, and the only other pattern with more than 50 blogs, is a variant on the first pattern: Fewer and longer posts but *less* conversational. In this case, the post counts are fairly typical—the pattern includes 12% of the blogs and 11% of the posts.
➢ There's one more pattern with substantially **more** blogs than you'd expect in a random scenario: Fewer and *shorter* posts, more conversational. That's 10% of the blogs and of the posts.
➢ There are enough blogs with incomplete metrics so that there are no real lightly-populated outliers with full metrics: The "average" pattern is also the smallest of the eight patterns with full metrics.

Looking at the totals

The six total lines are all interesting in different ways:

➢ **Post count**: While there are more than twice as many blogs with fewer posts than more posts—70% to 30%—the number of posts is essentially equal.
➢ **Length**: While the number of blogs for longer and shorter posts is close (44% to 38%), blogs with longer posts have a lot *more* posts (51% to 42%, the remainder being blogs with no length metrics).

➢ **Conversational intensity**: This is an even more striking pair—only 2% difference between increases and decreases (40% to 38%--22% of blogs had no comment metrics). Blogs with increased conversational intensity had a lot more posts (44% to 35%).

A Better Model

Unfortunately, splitting changes into straight more/fewer, more/less oversimplifies the reality of the situation. In practice, small changes aren't significant, particularly given that the sample only covers three months for each year.

A more realistic model has three categories for each metric:

➢ **Significant increase**: I've used a cutoff of 20%: the change metric must be at least 20% higher to represent significant increase. For a blog to grow from 100 to 110 posts in a year isn't all that significant—but for a blog to grow from 10 to 20 posts in a year is.
➢ **About the same**: The area of modest and possibly irrelevant change, from gains of 19% to losses of 19%.
➢ **Significant decrease**: At least 20% lower.

Tables 6.2 through 6.5 show the numbers for this model.

	Blogs	B%	Posts	P%
Significantly more posts	**108**	**21%**	**5,684**	**32%**
Longer posts	26	5%	637	4%
More conversational	12	2%	319	2%
About the same	4	1%	157	1%
Less conversational	9	2%	147	1%
No comment metrics	1	0%	14	0%
About the same length	46	9%	3,210	18%
More conversational	19	4%	1,182	7%
About the same	8	2%	429	2%
Less conversational	15	3%	669	4%
No comment metrics	4	1%	930	5%
Shorter posts	27	5%	1,397	8%
More conversational	7	1%	405	2%
About the same	5	1%	112	1%
Less conversational	14	3%	872	5%
No comment metrics	1	0%	8	0%
No length metrics	9	2%	440	2%
More conversational	1	0%	82	0%
Less conversational	2	0%	143	1%
No comment metrics	6	1%	215	1%
More conversational	39	7%	1,988	11%
About the same	17	3%	698	4%
Less conversational	40	8%	1,831	10%

Table 6.2: Change patterns for blogs with significantly more posts, 2007-2008

	Blogs	B%	Posts	P%
About the same post count	**91**	**17%**	**5,947**	**33%**
Longer posts	27	5%	1,168	7%
More conversational	12	2%	277	2%
About the same	3	1%	252	1%
Less conversational	9	2%	401	2%
No comment metrics	3	1%	238	1%
About the same length	38	7%	4,153	23%
More conversational	17	3%	1,144	6%
About the same	6	1%	588	3%
Less conversational	10	2%	735	4%
No comment metrics	5	1%	1,686	9%
Shorter posts	22	4%	490	3%
More conversational	6	1%	98	1%
About the same	3	1%	46	0%
Less conversational	9	2%	176	1%
No comment metrics	4	1%	78	0%
No length metrics	4	1%	136	1%
More conversational	1	0%	19	0%
About the same	1	0%	73	0%
No comment metrics	2	0%	44	0%
More conversational	36	7%	1,538	9%
About the same	13	2%	1,051	6%
Less conversational	28	5%	1,312	7%

Table 6.3: Change patterns for blogs, about same number of posts

	Blogs	B%	Posts	P%
Significantly fewer posts	**322**	**62%**	**6,250**	**35%**
Longer posts	85	16%	1,743	10%
More conversational	35	7%	709	4%
About the same	18	3%	497	3%
Less conversational	23	4%	433	2%
No comment metrics	9	2%	104	1%
About the same length	92	18%	2,235	12%
More conversational	40	8%	907	5%
About the same	19	4%	502	3%
Less conversational	30	6%	816	5%
No comment metrics	3	1%	10	0%
Shorter posts	65	12%	1,568	9%
More conversational	22	4%	1,094	6%
About the same	8	2%	88	0%
Less conversational	29	6%	362	2%
No comment metrics	6	1%	24	0%
No length metrics	80	15%	704	4%
More conversational	8	2%	224	1%
About the same	4	1%	68	0%
Less conversational	3	1%	116	1%
No comment metrics	65	12%	296	2%
More conversational	105	20%	2,934	16%
About the same	49	9%	1,155	6%
Less conversational	85	16%	1,727	10%

Table 6.4: Change patterns for blogs with fewer posts

Post length and conversational intensity for all blogs				
	Blogs	B%	Posts	P%
Longer posts	138	26%	3,548	20%
More conversational	59	11%	1,305	7%
About the same	25	5%	906	5%
Less conversational	41	8%	981	5%
No comment metrics	13	2%	356	2%
About the same length	176	34%	9,598	54%
More conversational	76	15%	3,233	18%
About the same	33	6%	1,519	8%
Less conversational	12	2%	2,626	15%
No comment metrics	55	11%	2,270	13%
Shorter posts	114	22%	3,455	19%
More conversational	35	7%	1,597	9%
About the same	16	3%	338	2%
Less conversational	52	10%	1,410	8%
No comment metrics	11	2%	110	1%
No length metrics	93	18%	1,280	7%
More conversational	10	2%	325	2%
About the same	5	1%	141	1%
Less conversational	5	1%	259	1%
No comment metrics	73	14%	555	3%
More conversational	180	35%	6,460	36%
About the same	79	15%	2,904	16%
Less conversational	153	29%	4,870	27%

Table 6.5: Post length and conversation changes for all blogs

The same notes apply as for Table 6.1. A reminder: "No length metrics" usually means that there were no posts in 2008, so length wasn't calculated—as opposed to the much smaller number of cases where there were posts but the blog couldn't be measured for length.

In cases where a given pattern has no examples, the line is omitted from the table.

Significant patterns

There are a lot more patterns using this model: 48 in all, at least potentially. That means a "typical" pattern in a random universe would include 11 blogs and 373 posts.

Interesting situations:

➢ Two patterns have more than 30 blogs each: Significantly fewer posts, about the same length and more conversational—and significantly fewer posts, longer posts and more conversational. These two patterns, both variants on the naïve hypothesis, combine to include 75 blogs (15%) but only about 9% of posts.

➢ More conversational blogs with longer or about the same length posts, where the number of posts stayed about the same, add another 29 blogs—and another 8% of posts. Combined, these four patterns total 104 blogs.

➢ Three other patterns (with full metrics) include at least 20 blogs and they're directly related patterns: All three are for blogs with significantly fewer posts and significantly less conversational intensity. Together, these three patterns total 82 blogs (16% of the universe for this chapter) but only about 9% of posts.

The problem with this model

The problem with this model, more realistic though it is, is evident from the tables: There are too many patterns to make quick sense of. Grouping similar patterns together and showing the blogs involved may help.

Blogs in Patterns

For most of you, the liblog landscape includes a few (or a few dozen) blogs you're very familiar with—and probably hundreds you've never heard of. Listing the blogs that fall into *some* of the change patterns may bring Tables 6.2 through 6.4 to life.

I've clustered together patterns that seem similar to me, but each pattern appears separately. Within a pattern, blogs are listed in decreasing number of 2008 posts. When an affiliation shows up more frequently than usual within a pattern (and there are at least three blogs), I note that.

The Naïve Hypothesis: Fewer but Richer

These blogs have significantly fewer posts in 2008 than in 2007, but the blogs are "richer" in some ways: Longer posts with more conversational intensity (and longer posts where there are no comment metrics), or posts about the same length with more conversational intensity—and blogs with more conversational intensity where I couldn't calculate length.

Those four patterns taken together total 92 blogs (18% of the universe for this chapter) and 1,944 posts (11% of the universe).

Fewer posts, longer posts, more conversational

This pattern is rich in blogs from academic libraries (40%), vendors (11%) and law (9%).

Fewer posts, longer posts, no comment metrics

Fewer posts, similar length, more conversational

Fewer posts, more conversational, no length metric

Growth Blogs

These blogs grew significantly in at least two measures (or in the only measure, if there was only one) and didn't have a significant reduction in any measure. That's a more diverse set of patterns than the "naïve hypothesis" set. It includes seven patterns that total only 55 blogs (11% of the universe) and 2,246 posts (13%).

Growth in all aspects

More posts, longer posts, no comment metric

More posts, more conversational, no length metric

More than half of these are from academics.

Relatively Stable Blogs

These blogs have no significantly lower factor and are otherwise either about the same or significantly higher for 2008. Ten patterns fall into this cluster, but only 50 blogs (10% of the universe)—but the blogs include 5,403 posts (30% of the total).

"Stable" doesn't mean unchanged—it means that measurable quantitative factors didn't increase or decrease by 20% or more. In some ways, it's surprising so few blogs fall into this pattern.

You could consider this group to be another piece of the "naïve hypothesis." It's rich in blogs from school librarians (18%).

Same length, number, no comment metric

This group primarily comes from academic librarians (60%), and it's probably worth noting that three of the most prolific blogs are in this overall cluster.

Same number of posts, no length metrics

More conversational:

About the same conversational intensity:

No comment metric:

Mixed Cases

These blogs have one (but only one) significantly lower metric, with others about the same or significantly higher. There are 15 individual patterns within this cluster, including 115 blogs and 3,919 posts—in both cases 22% of the universe.

More posts, longer posts, less conversational

More posts, same length, less conversational

More posts, shorter posts, more conversational

More posts, shorter posts same conversationally

More posts, shorter posts, no comment metric

More posts, less conversational, no length metric

About the same number, longer, less conversational

Same number and length, less conversational

Same number, shorter, more conversational

Same number and conversationally, shorter posts

NOTE: Throughout these lists, "same" means "about the same"—that is, the same plus or minus 19%.

Conclusion

I haven't listed all the blogs included in this part of the study. Some blogs had a "down year" in general—some disappearing, some simply reduced in general (or partly not measured). That includes ten patterns with quite a few blogs: 210 in all (40%) with 4,277 posts (24%). It includes 54 blogs with no posts in March-May 2008, some of them defunct, some on temporary hiatus. (If you're wondering, blogs that began in 2005 and 2006 were *slightly* more likely to have no posts in 2008, 11% in each case, than blogs that began in 2003 and 2004—7% and 9% respectively. I don't regard that as a big enough difference to be significant.)

Is a down year a bad thing? That depends on the blogger's intentions. Fourteen blogs had more posts—but the posts were significantly shorter and less conversational. That might suit the needs of the bloggers and represent success stories. A large group, 22 blogs, had fewer and shorter posts—but significantly more conversational intensity. Again, those may be success stories.

You can see the numbers for each pattern in Tables 6.2-6.4. You can see individual cases in the liblog profiles. I don't believe listing those blogs here would serve any purpose.

Correlations

Are there significant correlations between pairs of metrics? Any time you look at correlations, there are two meanings of "significant" to consider:

➢ **Statistically significant**: There's a good statistical "fit" between the two factors.
➢ **Meaningful**: There's some plausible reason to believe that the two factors *would* be correlated—that one has an effect on the other or that they somehow work in tandem.

When I began this project, I promised not to do a classic social sciences correlation free-for-all: Toss all the factors into SPSS and see what emerges. Of course, given lack of ready access to SPSS or equivalent, that's an easy thing to say—but as a humanist who's also numerate and somewhat statistically aware, I don't much care for random correlations. Show me that the change in average height of women in Boston during the 20th century correlates with the change in acreage devoted to vinifera grapes in Chile during the

period, and I'll say there's either some other factor influencing both of them or you're spouting statistical nonsense. (That's a bad example, since third factors are vaguely plausible in this case.)

I did test correlation on various pairs of metrics that *might* have meaningful relationships—and, in fact, on almost all possible pairs. The ones that emerged appear here. Most graphs don't have labels, to allow the most space for the graphs themselves.

Methodology

For each pair of metrics, I followed this practice:
➢ Removed blogs with no data or so little data as to be meaningless. In all cases, that means eliminating blogs with fewer than two posts in March-May 2008.
➢ Had Excel calculate Pearson's product-moment coefficient, the correlation between the two metrics. (The CORREL and PEARSON functions in Excel appear to do precisely the same calculations and return the same results. I used CORREL)
➢ If either correlation is medium or strong (absolute magnitude between 0.3 and 0.5 for medium, absolute magnitude over 0.5 for strong), I include the correlation itself and a scatterplot for the two metrics. I don't consider small correlation (with an absolute magnitude between 0.1 and 0.3) meaningful enough to bother mentioning, much less plotting.

Possible Age Correlations

Does the age of a blog correlate with its key metrics, or changes in those metrics? **No.** Every correlation came up between -0.15 and 0.22, in no case enough of a correlation to bother with a scatterplot.

A colleague had suggested that older blogs might be showing more of a decline in posts than younger blogs. I'd wondered whether older blogs might have more conversational intensity than younger blogs. In either case, it's possible—but not in a statistically meaningful sense across the appropriate universes, even eliminating trivial cases.

Count Correlations

There *should* be a strong correlation between the number of posts in 2007 and the number of posts in 2008, especially if we eliminate blogs with fewer than two posts in either year.

And so there is: For 443 blogs with at least two posts each year, the correlation is a very strong 0.927. You could say the number of posts during 2007 is a

strong predictor of the number of posts during 2008—does that surprise anybody?

Figure 7.1 shows 442 of the 443 blogs, eliminating the single blog with more than 1,000 posts in either year. This reduces the correlation slightly (to 0.907) but makes the scatterplot much more readable. Note that the scatterplot shows posts on logarithmic scales.
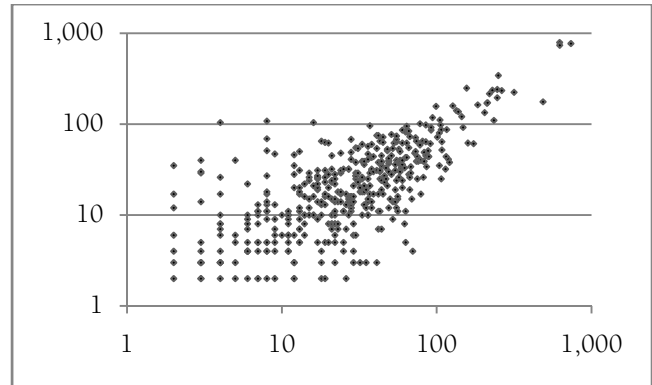
Figure 7.1: Correlation between 2007 and 2008 post counts

The vertical axis shows 2007 posts, the horizontal axis 2008. That's a fairly trivial correlation, although it makes an interesting figure.

Count and Length (2008)
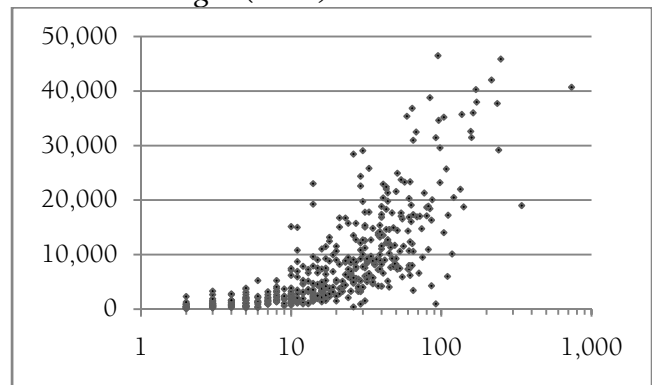
Figure 7.2: Correlation between post counts and total length (2008)

For Figure 7.2, the vertical axis is word count (linear), the horizontal axis number of posts (logarithmic). The correlation is 0.825, a strong correlation, although it may also be a trivial correlation (overall, blogs with lots of posts are likely to be long blogs). Removing the five longest blogs reduces the correlation to 0.662, considerably weaker (but still strong)—and makes the scatterplot much easier to portray, as in Figure 7.2. Once again, post count is shown on a logarithmic scale.

There's no significant correlation between number of posts and average words per post.

Posts and comments (2008)

Even eliminating blogs with no comments in 2008, there's no significant correlation between the number of posts in a blog and *either* the total number of comments (0.247) or comments per post (-0.06).

That *lack* of correlation may be significant: Blogs with more posts don't necessarily have more comments—at least not with a significant correlation.

Posts and changes

Are there correlations between the number of posts in 2008 and changes from year to year? For these measurements, I eliminated all blogs with fewer than two posts in either 2007 or 2008 (as well as eliminating zero situations for length or comments).

There's no real correlation between number of posts in 2008 and the percentage change in posting—the Pearson coefficient is 0.07. Neither are there correlations between number of posts and change in overall length (0.04) or number of posts and change in length per post (0.06). Finally, there's no real correlation between number of posts and change in comments (0.08) or change in comments per post (0.09).

Changes in posting and other metrics

Are there correlations between the *change* in post frequency from 2007 to 2008 and length-related or comment-related metrics?

For the 400 blogs with at least two posts and countable length in both 2007 and 2008, there's a strong correlation between the change in count and the change in total length (0.882). Note that this is a *stronger* correlation than between post count in 2008 and total length in 2008.

Removing 15 fringe cases—where the percentage increase in either measure was greater than 500%—reduces the correlation slightly (to 0.82) but allows for a more meaningful chart.
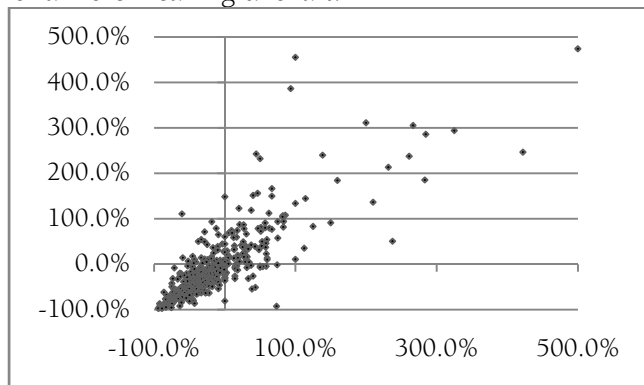
Figure 7.3: Change in posts vs. change in total length

The vertical axis shows change in overall length, the horizontal axis change in number of posts. There is

not a significant correlation between change in number of posts and change in length of posts.
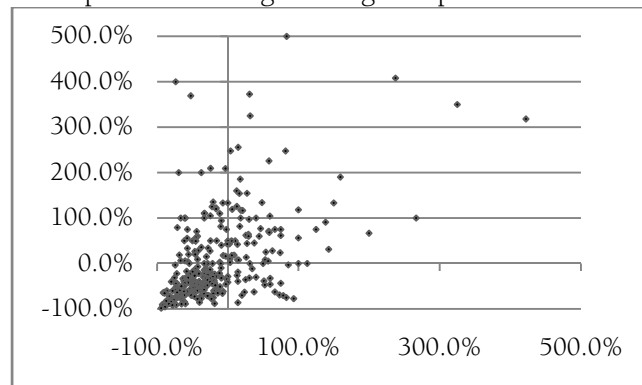
Figure 7.4: Changes in post count vs. changes in comment count

There's also a strong correlation between change in count and change in number of comments, although not between change in post count and any of the other comment-related metrics. Removing blogs with fewer than two comments in 2007 or 2008 (as well as those with fewer than two posts in either year), the remaining 322 blogs show a correlation between these two metrics of 0.557. Eliminating a dozen blogs where one or the other metric shows an increase of more than 500%, we have 310 blogs with a correlation of 0.49—technically a medium correlation, but on the border.

In Figure 7.4, the vertical axis is change in comments, the horizontal change in posts. As you can see, this is a much less well-behaved distribution than previous figures—and a lower correlation.

Length and Comment Correlations

Are there correlations between length of blogs (total or per post) and comments or changes in length or comments? In this case, blogs where length couldn't be calculated are omitted, along with blogs with fewer than two posts—and blogs with fewer than two comments, when looking at comment correlations.

There's a medium correlation (0.376) between total length of a blog and the number of comments in the blog—and again, that may not be entirely surprising. Removing four outlying cases—blogs with more than 55,000 words and blogs with more than 800 comments—yields a strong correlation (0.539) and a workable chart, Figure 7.5. In Figure 7.5, the vertical axis shows word count; the logarithmic horizontal axis shows number of comments.
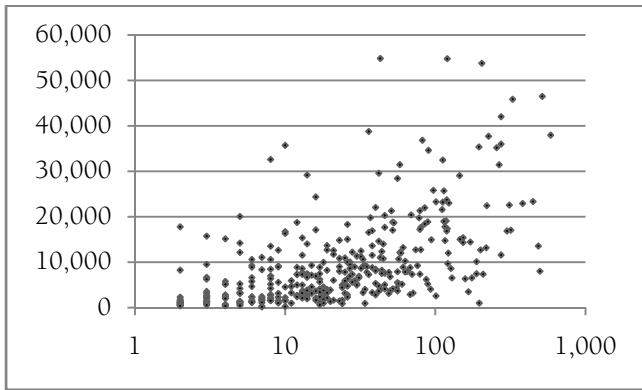
Figure 7.5: Number of posts vs. number of comments (2008)

There's also a medium correlation between change in overall length and change in overall comment counts—0.403 without removing extreme cases. Removing cases where comments grew by more than 550% and cases where length grew by more than 500% (eight blogs in all) increases the correlation to 0.475 (still a medium correlation) and results in Figure 7.6, with length change on the vertical axis and comment change on the horizontal axis.
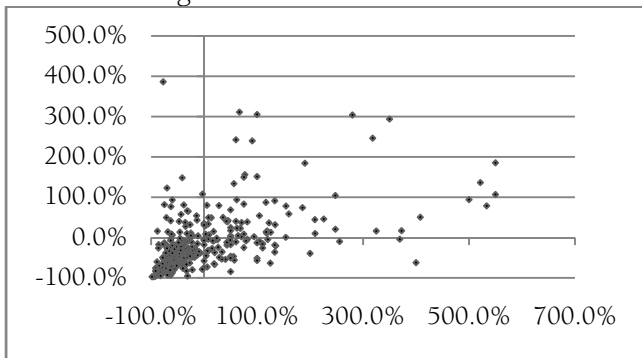
Figure 7.6: Change in overall length vs. change in number of comments
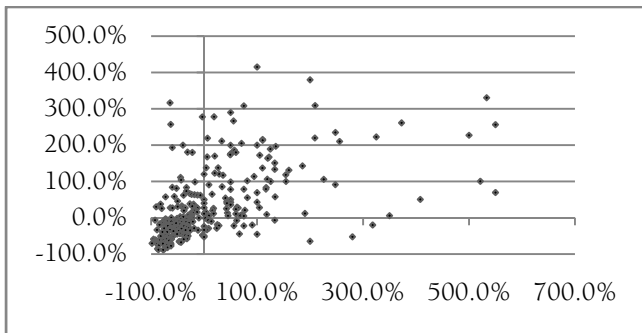
Figure 7.7: Change in comments per blog vs. change in comments per post

Finally, there's a medium correlation (0.393) between changes in comments per blog and changes in comments per post. Here again, removing extreme cases—blogs where comments increased by more than 550% and cases where comments per post increased by more than 500% (nine blogs in all) increased the correlation,

in this case to the lower edge of "medium correlation" (0.530) and yields Figure 7.7. The vertical axis is comments per post, horizontal total comments.

Conclusions

Of the seven correlations strong enough to be worth noting, six (Figures 7.1 through 7.6) seem intuitively likely. They're correlations that "make sense." The last one—changes in total comments and changes in comments per post—may be a bit more surprising, but it's not the kind of correlation that seems out of left field.

Consider all those pairs of metrics that did *not* show correlations, including the ones mentioned (all those where number of posts or change in number of posts is one metric) and those not mentioned (pairs involving length, comments and changes in either). In all, there are 45 such pairs, only seven of which showed medium or strong correlations.

What that means, if anything, is that you can't readily assume one aspect of a blog's behavior given knowledge of some other aspect. Yes, if the blog has lots of posts it's likely to have lots of text—and somewhat more likely to have lots of comments (but with no expectations for comments per post, one way or the other). Blogs that have more posts in one year are likely to have more posts the next year, with lots of exceptions. Changes in post count are likely to track changes in total length and comment count fairly well. And changes in number of comments have some similarity to changes in comments per post—but not a strong similarity. And that's about it.

The 2006-2008 Landscape

In the summer of 2006, I did a quantitative study of 213 liblogs I thought constituted "the great middle"—blogs that had some visibility but weren't the *most* visible liblogs. That study appeared in the August 2006 *Cites & Insights* (citesandinsights.info/civ6i10.pdf).

This chapter looks at some of the metrics for 2006-2008 for *some* of those blogs—namely, ones that have at least two posts in each of the three March-May study periods. That leaves 143 blogs—some of the others disappeared, some didn't have any 2007 posts, some had only one post or none in one of the quarters. We'll omit most charts and explanations but include some tables and relate those tables to the larger universe as appropriate.

Posts and Changes in Posting

In terms of overall posts, these blogs are reasonably typical.

- ➤ **2007**: These 143 blogs represent 27% of blogs with countable posts, and include 6,653 posts: 29% of the total.
- ➤ **2008**: These 143 blogs represent 27% of blogs with countable posts and include 4,829 posts: 25% of the total.
- ➤ **2006**: These 143 blogs include 8,145 posts.

Means and medians

- ➤ **Posts**: These blogs average 57 posts per blog in 2006, 47 in 2007 and 34 in 2008. That's just a little higher than the adjusted figure for 2007 for liblogs as a whole (44) and exactly the same for 2008. Median figures are 40 for 2006, 31 for 2007 and 19 for 2008—significantly higher than liblogs as a whole for 2007 (31 vs. 22) and slightly higher for 2008 (19 vs. 15).
- ➤ **Change in posting frequency**: These blogs averaged a 16% *increase* 2006-2007, a 4% increase 2007-2008, and a 23% decrease 2006-2008: The kind of figures you can only get by computing averages of averages! Median figures: -20% 2006-2007, -30% 2007-2008, -39% 2006-2008. (Compare -30% to -34% for the larger universe—a fairly trivial difference.)

Distribution (Quintiles)

2006	High	Low	Blogs	Median	Posts	%
Total	371	2	143	40	8,143	
Q1	371	85	29	117	4,302	53%
Q2	84	50	29	60	1,887	23%
Q3	49	30	28	40	1,103	14%
Q4	28	16	29	20	614	8%
Q5	15	2	28	8	239	3%

Table 8.1: Quintiles for post count, 2006 posts

Not only are counts generally higher than the overall counts for 2007, as you'd expect given other information, the distribution is differently shaped. That is, Q1 includes a little over half of all the posts—where for the larger universe in 2007 and 2008, Q1 includes nearly two-thirds of all the posts. Conversely, Q2, Q3, and Q4 all run three or four percentage points higher: Between them, they total 44% of the posts, where that 60% of blogs (in the larger universe) totals only 35% of the posts in 2007 and 34% in 2008.

How does this smaller group compare to the universe for 2007 and 2008? For 2007, Q1 includes 35 of these blogs (24%), Q2 includes 37 (26%), Q3 has 29 (20%), Q4 has 26 (18%) and Q5 has 16 (11%)—in other words, they skew high, in part because one-post blogs aren't included. For 2008, Q1 again in-cludes 35 of these blogs, but Q2 only includes 27 (19%), Q3 24 (17%)—and Q4 has 32 (22%), leaving 25 for Q5 (17%). In this case, more of these "middle blogs" are among those with the most posts, with the rest evenly distributed.

Change quintiles

I thought that 2007 was a peak year for the typical blog. I was apparently mistaken—at least for blogs in the great middle. While the 2006-2007 change quintiles are considerably more positive than the 2007-2008 change quintiles (for a larger universe and including blogs with one or no posts in 2008), more than 60% of these blogs still had fewer posts in 2007 than in 2006.

2006-07	High	Low	Blogs	Median	Posts	%
Total	1150%	-98%	143	-20%	6,653	
Q1	1150%	31%	28	112%	1,671	25%
Q2	29%	-6%	29	8%	1,506	23%
Q3	-7%	-27%	29	-20%	1,785	27%
Q4	-28%	-47%	28	-38%	1,188	18%
Q5	-48%	-98%	29	-70%	503	8%

Table 8.2: Changes in post count, 2006-2007

How about changes for a two-year period, from 2006-2008?

2006-08	High	Low	Blogs	Median	Posts	%
Total	400%	-96%	143	-39%	4,829	
Q1	400%	6%	28	48%	1,187	25%
Q2	5%	-28%	28	-17%	1,700	35%
Q3	-29%	-56%	30	-38%	1,070	22%
Q4	-57%	-74%	29	-64%	575	12%
Q5	-58%	-95%	28	-84%	297	6%

Table 8.3: Changes in post count, 2006-2008

Some 69% of blogs showed at least a 20% drop in posts from 2006 to 2008—and 45% had no more than half as many posts.

Changes for these blogs for 2007-2008 are typical of all liblogs.

Change patterns

It's implausible to offer a single table correlating changes across all three metrics (posts, length per post, comments per post) and all three years: There would be as many patterns as there are blogs.

I can provide year-to-year patterns for each of the three metrics, however, as there are no more than 18 possibilities for such patterns.

Here's what I found for changes in posting, using 20% limits for significant change.

2006-07	2007-08	2006-08	Blogs	%
More	More	More	5	3%
More	Same	More	4	3%
More	Fewer	More	7	5%
More	Fewer	Same	9	6%
More	Fewer	Fewer	12	8%
Same	More	More	2	1%
Same	More	Same	1	1%
Same	Same	More	1	1%
Same	Same	Same	5	3%
Same	Same	Fewer	4	3%
Same	Fewer	Fewer	21	15%
Fewer	More	More	5	3%
Fewer	More	Same	3	2%
Fewer	More	Fewer	10	7%
Fewer	Same	Same	3	2%
Fewer	Same	Fewer	11	8%
Fewer	Fewer	Fewer	40	28%

Table 8.4: Change patterns for number of posts

A few of those patterns may be worth listing, with growth from 2006 to 2008.

Continuous growth (more, more, more)

DigitalKoans 324%
:: CultureLibre.ca :: 237%
InfoMatters 117%
Vagabondages 116%
Vancouver Law Librarian Blog 62%

Mixed growth

More, same, more:
Incoherent Scribblings 400%
Bibliotekarens bibliotek 67%
Diary of a Subversive Librarian 45%
Ref Grunt 25%

Same, more, more:
The Imaginary Journal of Poetic Economics 294%
Library Technology in Texas 50%

Same, same, more:
The Misadventures of Super_Librarian 29%

Ups and downs—but up on the whole

More, fewer, more:
Citegeist 140%
Young Librarian 100%
Connecting Librarian 83%
The Gay Librarian 50%
Library Geek Woes 40%
LibrarianActivist.org 36%
shimenawa 25%

Fewer, more, more:
Biblioteksrelaterat 94%
Helenes hengekøye 33%
Family Man Librarian 32%
BlogJunction 28%
Love the Liberry 22%

Words

Taking away blogs where length couldn't be calculated in one year or another reduces this already-small universe from 143 blogs to 112 blogs. The quintiles for total length in 2006:

Words	High	Low	Blogs	Median
Overall	144,504	131	112	9,980
Q1	144,504	19,660	22	33,169
Q2	18,009	12,801	23	15,452
Q3	12,472	6,654	22	9,980
Q4	6,647	2,999	23	4,149
Q5	2,909	131	22	1,696

Table 8.5: Words per blog, 2006 quintiles

Changes in length

2006-07	High	Low	Blogs	Median
Overall	942%	-99%	112	-70%
Q1	942%	69%	22	178%
Q2	67%	7%	23	30%
Q3	5%	-15%	23	-7%
Q4	-16%	-54%	22	-34%
Q5	-55%	-99%	22	-74%

Table 8.6: Changes in overall length, 2006-2007

2006-08	High	Low	Blogs	Median
Overall	314%	-99%	112	-30%
Q1	314%	39%	22	96%
Q2	38%	-12%	22	3%
Q3	-14%	-41%	23	-30%
Q4	-42%	-78%	23	-56%
Q5	-79%	-99%	22	-87%

Table 8.7: Changes in overall length, 2006-2008

It's interesting that the quintiles here aren't much different from the overall change quintiles for 2007-2008—but the overall figures do include blogs with one and no posts in 2008, excluded here.

Words per post

As with the larger universe, these figures are inherently more interesting than overall word count.

2006	High	Low	Blogs	Median
Overall	847	25	112	240
Q1	847	392	22	529
Q2	385	283	23	332
Q3	282	196	22	240
Q4	194	144	23	167
Q5	136	25	22	94

Table 8.8: Words per post, quintiles, 2006

These are generally slightly lower numbers than for the universe as a whole in either 2007 or 2008. But let's look at the changes.

2006-07	High	Low	Blogs	Median
Overall	303%	-73%	112	18%
Q1	303%	59%	22	91%
Q2	58%	34%	23	45%
Q3	33%	7%	22	18%
Q4	6%	-18%	23	-5%
Q5	-22%	-73%	22	-39%

Table 8.9: Changes in words per post, 2006-2007

Table 8.9 is *quite* interesting: It suggests that, at least for length of posts, the "naïve hypothesis" may have happened a year earlier. Most of these blogs had fewer posts in 2007 than in 2006—but most had more words per post. In fact, 68% of the smaller universe had no fewer words per post—and 47% had *significantly* more words per post.

2006-08	High	Low	Blogs	Median
Overall	608%	-90%	112	13%
Q1	608%	65%	22	111%
Q2	62%	23%	22	42%
Q3	22%	5%	23	13%
Q4	3%	-28%	23	-11%
Q5	-31%	-90%	22	-51%

Table 8.10: Changes in words per post, 2006-2008

Change patterns

2006-07	2007-08	2006-08	Blogs	%
Longer	Longer	Longer	12	11%
Longer	Same	Longer	16	14%
Longer	Same	Same	3	3%
Longer	Shorter	Longer	4	4%
Longer	Shorter	Same	13	12%
Longer	Shorter	Shorter	5	4%
Same	Longer	Longer	13	12%
Same	Longer	Longer	3	3%
Same	Same	Same	12	11%
Same	Same	Shorter	2	2%
Same	Shorter	Shorter	7	6%
Shorter	Longer	Longer	3	3%
Shorter	Longer	Same	1	1%
Shorter	Longer	Shorter	3	3%
Shorter	Same	Same	3	3%
Shorter	Same	Shorter	6	5%
Shorter	Shorter	Shorter	6	5%

Table 8.11: Change patterns for average length of posts

Here also, the actual blogs with longer posts in 2008 than in 2006 may be worth listing by pattern.

Longer posts each year

Mixed but never shorter, longer overall
Longer, same, longer:
Same, longer, longer:
Same, same, longer:

Sometimes shorter, finally longer
Longer, shorter, longer:
Shorter, longer, longer:

The percentage shown above is the percentage change in average length per post from 2006 to 2008; the

triplets of words are for 2006-2007, 2007-2008, and 2006-2008 in that order.

Comments

I included blogs with at least one countable comment in each year—which still reduces the universe to a mere 94 blogs. Those 94 blogs account for 4,940 comments in 2006, 6,289 comments in 2007, and 6,020 comments in 2008.

2006	High	Low	Blogs	Median
Overall	597	1	94	23
Q1	597	88	19	125
Q2	78	33	18	45
Q3	31	18	18	24
Q4	17	9	20	13
Q5	8	1	18	5

Table 8.12: Comments per blog, 2006, quintiles

Changes in overall comments per blog

2006-07	High	Low	Blogs	Median
Overall	2291%	-99%	94	23%
Q1	2291%	169%	19	300%
Q2	160%	56%	19	107%
Q3	48%	0%	18	23%
Q4	-8%	-46%	19	-29%
Q5	-51%	-99%	19	72%

Table 8.13: Changes in comments per blog, 2006-2007

2006-08	High	Low	Blogs	Median
Overall	3485%	-93%	94	-25%
Q1	3485%	100%	19	350%
Q2	89%	-2%	19	56%
Q3	-3%	-45%	18	-25%
Q4	-48%	-68%	19	-55%
Q5	-70%	-93%	19	-86%

Table 8.14: Changes in comments per blog, 2006-2008

Conversational intensity: Comments per post

2006	High	Low	Blogs	Median
Overall	9.63	0.02	94	0.78
Q1	9.63	1.83	19	2.55
Q2	1.82	0.95	19	1.26
Q3	0.93	0.56	19	0.75
Q4	0.59	0.19	17	0.31
Q5	0.18	0.02	20	0.13

Table 8.15: Conversational intensity quintiles, 2006

There weren't many comments per post in 2006, by and large.

2006-07	High	Low	Blogs	Median
Overall	1615%	-94%	94	47%
Q1	1615%	226%	19	344%
Q2	215%	75%	19	117%
Q3	70%	25%	18	47%
Q4	24%	-12%	19	0%
Q5	-25%	-94%	19	-54%

Table 8.16: Changes in conversational intensity, 2006-2007

Most blogs in this subset had significantly higher conversational intensity in 2007 than in 2006. A full 64%, just shy of two-thirds, showed increases of 20% or more. Taken in conjunction with other tables in this chapter, that's strong support for the idea that "the change"—fewer posts, longer posts, more comments per post—happened more in 2007 than in 2008.

2006-08	High	Low	Blogs	Median
Overall	5044%	-92%	94	49%
Q1	5044%	198%	19	494%
Q2	197%	77%	19	115%
Q3	74%	33%	18	49%
Q4	27%	-12%	19	9%
Q5	-14%	-92%	19	-54%

Table 8.17: Changes in conversational intensity, 2006-2008

Blogs *within this group* maintained higher conversational intensity, although it didn't increase (except in the top group) from 2007 to 2008. Of course, the blogs in Q1 for 2006-2007 aren't necessarily the same blogs as in 2006-2008—in fact, of the 19 blogs with the greatest increase in conversational intensity over the two-year period, only 12 were also in Q1 for 2006-2007. (It may annoy some of you to know that the blog with the greatest percentage increase for 2006-2007 was also the blog with the greatest increase for 2006-2008.)

The patterns for conversational intensity change over the two years are unusual, largely because of the large increases in conversational intensity from 2006 to 2007. Given that and the small number of blogs in this part of the study, it may not be surprising that only 12 of 18 possible patterns actually occur.

Here are the patterns, followed by blogs by pattern for those blogs that wind up with significantly more conversational intensity between 2006. and 2008.

2006-07	2007-08	2006-08	Blogs	%
More	More	More	21	22%
More	Same	More	15	16%
More	Less	More	11	12%
More	Less	Same	7	7%
More	Less	Less	6	6%
Same	More	More	7	7%
Same	Same	Same	4	4%
Same	Less	Less	4	4%
Less	More	More	5	5%
Less	More	Same	6	6%

Less	More	Less	2	2%
Less	Less	Less	6	6%

Table 8.18: Change patterns for conversational intensity

Continually growing conversational intensity

Mixed growth, no decrease

More, same, more:

Same, more, more:

Some less, but more overall

More, less, more:

Same, more, more:

Less, more, more:

Conclusions

What does this subgroup tell us? If it's really typical of liblogs as a whole, the big change happened in 2007 and continues at a slower pace in 2008.

Unfortunately, it's not only a relatively small sample, it's not at all a random sample. Still, the numbers are persuasive.

Subgroups

Do subgroups of liblogs show major differences from liblogs as a whole? That's the question this chapter tries to answer—and I'm not sure there are definitive answers.

That's partly because most subgroups aren't very large, and the smaller the subgroup, the less meaning you can assign to deviations from the norm. Oh, you can certainly take note of the fact that two of three active marketing liblogs in 2007 had little or no activity in 2008—but that's a narrow observation.

I'll lay out the observable facts for 15 of the larger subgroups, although a few of those subgroups are small enough (fewer than 15 blogs) that I'd be hard-pressed to assign much importance to the results. For each subgroup, I provide a table showing the eight key metrics (posts, words per posts, comments per post and figures per post—and the change in each of those from 2007 to 2008) and the number of blogs in

each quintile for each metric. The five quintiles for each metric are followed by "Norm," that is, the total number of blogs with numbers for that metric, divided by five. The norm is the number you'd expect to see in every box if the blogs were "perfectly normal" in distribution.

Where I see differences I'd consider at least possibly meaningful or at least interesting, the line appears in boldface. That means at least one of the quintiles has enough more or fewer blogs than the norm to be interesting.

What's interesting? For a group with 25 blogs (and a norm of 5.0), I wouldn't consider 6 or 4 to be particularly interesting—but 7 or 3 might be, and 8 or 2 would *certainly* be. Since I provide the full table, you can draw your own conclusions.

As appropriate, I also add a few words about the differences.

Blogs without Known Authors

This subgroup includes anonymous blogs, pseudonymous blogs and blogs whose authors use only a first name.

Metric	Q1	Q2	Q3	Q4	Q5	Norm
Posts	19	17	19	24	20	19.8
Words per post	22	22	16	16	16	18.4
Comments per post	**28**	**17**	**15**	**11**	**12**	**16.6**
Figures per post	**13**	**16**	**9**	**16**	**12**	**13.2**
Change in posts	22	18	18	20	20	19.6
Change in length	**18**	**17**	**21**	**12**	**23**	**18.2**
Change in comments	14	19	19	13	21	17.2
Change in figures	15	13	10	16	17	14.2

Figure 9.1: Blogs without known authors

Striking deviations: The large number of blogs with most comments per post, the number of sharp declines in words per post and the number with significantly fewer posts.

Group Blogs

Metric	Q1	Q2	Q3	Q4	Q5	Norm
Posts	**17**	**22**	**14**	**10**	**17**	**16.0**
Words per post	**13**	**10**	**21**	**12**	**17**	**14.6**
Comments per post	**10**	**7**	**12**	**21**	**16**	**13.2**
Figures per post	**7**	**11**	**8**	**10**	**18**	**10.8**
Change in posts	17	18	18	13	16	16.4
Change in length	11	14	15	15	19	14.8
Change in comments	**9**	**13**	**12**	**9**	**25**	**13.6**
Change in figures	**13**	**6**	**15**	**9**	**17**	**12.0**

Figure 9.2: Group blogs

Striking deviations: Blogs with more posts (and few blogs with fewer posts), blogs with average word length (and few blogs with longer posts), blogs with fewer comments (and few blogs with more comments).

Signed Blogs

Metric	Q1	Q2	Q3	Q4	Q5	Norm
Posts	72	69	72	72	70	71.0
Words per post	63	65	61	70	66	65.0
Comments per post	**48**	**64**	**59**	**75**	**59**	**61.0**
Figures per post	56	43	51	48	44	48.4
Change in posts	67	71	67	72	67	68.8
Change in length	65	63	55	68	55	61.2
Change in comments	**53**	**45**	**45**	**56**	**86**	**57.0**
Change in figures	**47**	**46**	**42**	**37**	**60**	**46.4**

Figure 9.3: Signed Blogs

Striking deviations: Few blogs with most comments, many blogs with biggest drop in comments, many blogs with biggest drop in figures.

Australian Blogs

Metric	Q1	Q2	Q3	Q4	Q5	Norm
Posts	**2**	**6**	**3**	**3**	**5**	**3.8**
Words per post	**6**	**1**	**3**	**4**	**4**	**3.6**
Comments per post	**7**	**3**	**4**	**2**	**3**	**3.8**
Figures per post	**5**	**2**	**1**	**2**	**2**	**2.4**
Change in posts	4	2	5	3	4	3.6
Change in length	**3**	**3**	**2**	**5**	**2**	**3.0**
Change in comments	**4**	**1**	**5**	**0**	**5**	**3.0**
Change in figures	**1**	**4**	**3**	**1**	**4**	**2.6**

Figure 9.4: Australian Blogs

Striking deviations: Few blogs with most posts (and more with more), more blogs with longest posts (and few with longer), more blogs with most comments—and few blogs with average figures, either Q2 or Q4 for changes in comments, and Q1 or Q4 for changes in figures.

Canadian Blogs

Metric	Q1	Q2	Q3	Q4	Q5	Norm
Posts	**12**	**8**	**8**	**15**	**5**	**9.6**
Words per post	**7**	**7**	**14**	**10**	**6**	**8.8**
Comments per post	**4**	**5**	**9**	**9**	**10**	**7.4**
Figures per post	**5**	**5**	**7**	**5**	**11**	**6.6**
Change in posts	10	13	13	7	9	10.4
Change in length	**8**	**14**	**8**	**6**	**10**	**9.2**
Change in comments	**6**	**6**	**2**	**10**	**19**	**5.4**
Change in figures	**7**	**2**	**7**	**8**	**8**	**6.4**

Striking deviations: Few blogs with fewest posts, more with average-length posts, fewest comments and fewest figures, more with Q2 length change, biggest drop in comments; few with Q2 change in figures.

United Kingdom Blogs

I do include the two Scottish blogs in this group.

Metric	Q1	Q2	Q3	Q4	Q5	Norm
Posts	9	11	3	7	3	6.6
Words per post	7	5	5	7	7	6.2
Comments per post	3	9	6	6	5	5.8
Figures per post	4	2	7	3	9	5.0
Change in posts	5	8	8	7	2	6.0
Change in length	5	10	4	4	4	5.4
Change in comments	5	7	5	5	7	5.8
Change in figures	3	4	7	5	4	4.6

Figure 9.6: United Kingdom Blogs

Striking deviations: Fewer blogs with average number of posts or fewest posts, more with more posts; fewer blogs with most comments or more figures; fewer blogs with biggest decrease in posts and more with Q2 length change.

Academic Library Blogs

These are blogs from people self-identified as working in academic libraries. Relatively few are specifically devoted to academic librarianship. This is the largest "affiliation" group—and I'd guess a fair number of those without explicit affiliations are in academic libraries. Note that law, medicine and science took priority over academic as an affiliation (but there were too few science blogs to do a table).

Metric	Q1	Q2	Q3	Q4	Q5	Norm
Posts	20	33	36	41	22	30.4
Words per post	27	27	26	31	27	27.6
Comments per post	25	25	26	26	23	25.0
Figures per post	19	14	22	18	23	19.2
Change in posts	31	31	29	31	31	30.6
Change in length	30	27	25	24	28	26.8
Change in comments	24	25	21	17	40	25.4
Change in figures	21	18	11	13	27	18.0

Figure 9.7: Academic Blogs

Striking deviations: More blogs with fewer posts (and fewer with most posts), more blogs with biggest drop in comments (and fewer with second-biggest).

Blawgs

Blogs from people who identify themselves as being in law libraries.

Metric	Q1	Q2	Q3	Q4	Q5	Norm
Posts	5	6	6	2	3	4.4
Words per post	2	5	6	3	5	4.2
Comments per post	0	6	3	6	5	4.0
Figures per post	1	2	1	4	8	3.2
Change in posts	3	5	0	9	4	4.2
Change in length	3	5	3	5	3	3.8
Change in comments	2	6	2	5	3	3.6
Change in figures	1	3	2	4	3	2.6

Figure 9.8: Law Blogs ("Blawgs")

Here, the universe is small enough that it's hard to say which differences are striking—but the total absence of most-heavily-commented blogs and blogs with average change in posts is interesting.

Medical Library Blogs

Metric	Q1	Q2	Q3	Q4	Q5	Norm
Posts	7	3	2	2	2	3.2
Words per post	3	3	1	2	4	2.6
Comments per post	1	3	4	2	3	2.6
Figures per post	3	3	0	3	3	2.4
Change in posts	4	3	2	4	1	2.8
Change in length	2	4	2	1	2	2.2
Change in comments	2	2	1	4	3	2.4
Change in figures	3	1	3	1	1	1.8

Figure 9.9: Medical Library Blogs

Another relatively small group—but this time there's a clear and unsurprising deviation: Quite a few of these blogs have lots of posts.

Public Library Blogs

Metric	Q1	Q2	Q3	Q4	Q5	Norm
Posts	23	12	17	19	21	18.4
Words per post	20	18	13	20	14	17.0
Comments per post	22	15	21	11	14	16.6
Figures per post	21	15	12	8	8	12.8
Change in posts	16	15	18	18	19	17.0
Change in length	18	11	14	17	19	15.8
Change in comments	19	9	12	14	20	14.8
Change in figures	7	11	12	15	21	13.2

Figure 9.10: Public Library Blogs

Striking deviations: More blogs with most figures (and few with fewer or fewest)—but also more blogs with

biggest decrease in figures (and few with biggest in-crease).

School Library Blogs

Metric	Q1	Q2	Q3	Q4	Q5	Norm
Posts	4	3	2	3	3	3.0
Words per post	2	2	5	2	3	2.8
Comments per post	**4**	**5**	**2**	**1**	**1**	**2.6**
Figures per post	**6**	**2**	**2**	**1**	**1**	**2.4**
Change in posts	2	5	2	4	2	3.0
Change in length	2	2	5	3	3	3.0
Change in comments	**5**	**3**	**2**	**2**	**1**	**2.6**
Change in figures	**1**	**2**	**6**	**1**	**2**	**2.4**

Figure 9.11: School Library Blogs

The most striking deviation is no surprise: More blogs with most figures (and few with fewer or fewest pictures).

Vendor Blogs

Some of these are official vendor blogs. Others are from people who *currently* work for vendors—which makes this an odd subgroup, since people do change sectors.

Metric	Q1	Q2	Q3	Q4	Q5	Norm
Posts	**5**	**8**	**1**	**1**	**4**	**3.8**
Words per post	**5**	**5**	**4**	**2**	**0**	**3.2**
Comments per post	**5**	**3**	**4**	**2**	**1**	**3.0**
Figures per post	3	3	3	3	2	2.8
Change in posts	**5**	**5**	**7**	**0**	**1**	**3.6**
Change in length	**3**	**6**	**1**	**4**	**1**	**3.0**
Change in comments	**5**	**0**	**4**	**4**	**3**	**3.2**
Change in figures	3	2	5	3	3	3.2

Figure 9.12: Vendor Blogs

With two exceptions, most large deviations are *absences*—the lack of any blogs with terse posts, larger decrease in posts or Q2 change in comments. The exceptions: More blogs with more posts—and more with an average change in number of posts.

Visibility

If you're trying to sell ads on your blog or hoping for some sort of sponsorship, demonstrable visibility is crucially important. If you're writing a blog because it's a class assignment, visibility means nothing at all. Most of us are somewhere in between—and I suspect few of us have a solid notion of what visibility really means or where we fit in.

This is the third time I've studied some portion of liblogs (and the fifth time I've studied some group of library-related blogs). The first two times, visibility—or reach, an alternative term that means much the same thing—was a significant factor in the study.

This time, visibility served only as a crude filter to limit the universe being studied. I didn't want to spend time looking at blogs that probably wouldn't survive for more than a few months and I didn't want to publicize blogs when they may be deliberately aimed at a tiny group—what I call the "family and close friends" blogs.

This chapter is part history, part contemporary reality. We'll look at the tools I used in 2005 and 2006, why I changed to new tools when this study began—and why none of those tools seems particularly useful at this point. The last section will look at a single new tool that appears workable *only* as a crude filter—which, in future studies, is the only use I'd make of visibility measures.

2005: Reach Metrics

"Investigating the biblioblogosphere" appeared in the September 2005 *Cites & Insights* (citeandinsights.info/civ5i10.pdf). I hated the term even then (and say so in the second sentence of the article), but didn't abandon it until later. My study followed on work by Jon Garfunkel at *Civilities* (civilities.net), but isn't directly related to it.

I wanted to identify "*a* Top 50"—a set of blogs by library people (but not large-group blogs) that were among the most widely-read or had the greatest reach in 2005. After identifying them—and winding up with 60 blogs out of more than 230 initial candidates—I recorded various metrics for the blogs, listed the leaders for each metric and offered brief comments on each blog.

Here's what I had to say about the Reach metric, my basis for choosing 50 of the 60 blogs:

Starting point: Bloglines subscriptions
I reset Bloglines to show all listings, clicked on each feed, added up the numbers, jotted down the total on the printed list, then unsubbed all but one feed for each blog.

Narrowing the Field
The next step was to consider "reach" beyond apparent Bloglines readership, particularly because Bloglines numbers can be tricky.

The usual metric for blog reach is the Technorati result. I didn't focus on that, partly because I found it difficult to keep Technorati going long enough to do

many searches at once, partly because I think Technorati overvalues blog linking and, specifically, blogrolls. I did get Technorati linked-site numbers for the 60 "finalists" in the study, and only a few of them vary much in rank from the Reach score I prepared.

I did "link:" searches on three major open-Web search engines that support such searches: Google, MSN and AllTheWeb (one of several Overture-owned sites using the same search engine).

The results were startlingly varied, although blogs at or near the top of the list in one measure tended to stay there in others. But the numbers! For *The Shifted Librarian*, Google showed 9,430 links; MSN showed 76,676; and AllTheWeb showed *449,000*! Totaling all 231 sites, here's how extreme the differences were:

➢ Bloglines showed a total of 31,636 feeds (which I extrapolated to 126,544 readership).
➢ Google showed a total of 71,401 links.
➢ MSN showed a total of 297,345 links.
➢ AllTheWeb showed a total of 2,295,436 links.

While links tended to be proportional across search engines, that wasn't always the case. I could suggest at least half a dozen rankings based on that set of raw data. In the end, I calculated a "Reach" figure based half on apparent direct readership (Bloglines times 4), half on a weighted average of reported links, where weighting was proportional to the overall numbers:

Reach = (4 * Bloglines) + (0.67 * Google) + (0.16 * MSN) + (0.02 * AllTheWeb).

The spreadsheet for this analysis is available at http://waltcrawford.name/liblograw.xls. Feel free to manipulate the calculations to arrive at your own rankings and draw your own conclusions.

Results and Metrics

Any metric such as Reach, which runs from six to 47,602 for a group of 231 items, will have obvious breaks in the sequence—gaps that can be used to separate groups of entries. I was looking for 50 to 60 blogs for further analysis. One of several gaps was in that range. After eliminating a few blogs (for reasons noted below), there were 48 blogs in the "top 50."

I added blogs that, while not scoring in the top 50 on Reach, were either in the top 40 of Bloglines subscriptions or the top 30 in Google links, MSN links, or AllTheWeb links. Any of these—and, arguably, others as well—could belong in the "top 50."

2006: Tweaking the Algorithm

I was a little more ambitious in 2006—and a little more cautious about identifying "top" liblogs. It struck me that it was already pretty easy for people to identify the big-name liblogs, and that in some ways another, larger group was more interesting.

Here's some of what I said at the start of "Looking at liblogs: The great middle," *Cites & Insights* 6:10 (August 2006) (citesandinsights.info/civ6i10.pdf):

I knew I wanted to do something similar this year—but I wanted to do it differently. In the past few months, thinking about blogs has become more complex, especially as feeds and aggregators have become so easy and popular.

For example:

➢ Feeds eliminate the need to blog all the time in order to be visible. With aggregators, "blogging to be blogging" can be a danger: Your posts seem forced or repetitive and may encourage people to unsubscribe.
➢ For those who aren't out to be A-listers or politicians, size of audience has diminished in importance. The hope now is to find the *right* audience, which might be anywhere from half a dozen friends to a few thousand strangers.
➢ More than one commentator has suggested that the most interesting blogs are in the great middle—blogs with more than a handful of readers but not so popular as to carry the burden of popularity.

Redefining the scope

I fine-tuned criteria this year. There's no limit on group size, but I eliminated official blogs of all sorts. There had to be at least one posting during March through May 2006. I require at least one feed that Bloglines can recognize—but I dropped the English-only criterion.

The biggest change has to do with reach and readership. To the extent that an observer can gauge reach and readership, I wanted to avoid the "A list" in favor of a broader group of liblogs in the Great Middle. What's the Great Middle? It's the middle of the power-law curve: Blogs with more than a handful of readers, which garner some attention but aren't among the most popular in the field. There's no clear definition of that middle, just as there's no clear definition of reach or readership. In this case, it's a little less than half of the liblogs that meet other criteria (and that I could find), omitting roughly the most widely read sixth and least widely read third. "Roughly" is the right word in all cases.

Expanding the candidate pool

I began with LISWiki blogs not already in my Bloglines list, then went through Open Directory blogs that didn't show up elsewhere, and finally picked up

new items from PubSub. I added blogs with 16 to 689 total Bloglines subscriptions…

That left me with 368 candidates—far too many even for this expanded essay. I checked something like 650 liblogs in all, of which 554 are still active, aren't official or corporate, have an RSS feed, and have at least one subscription.

Draining the pool

I had to cut more. There are many ways to cut, and I didn't find "natural breaks." Using "half taken from the upper middle" as my target, I eliminated the most widely subscribed 90 and least widely subscribed 183 from the original 554, leaving 281 blogs with 19 to 196 Bloglines subscriptions. Note that only 25 of last year's candidates had more than 196 Bloglines subscriptions. Unquestionably, liblogs across a broad range have become more popular.

I wanted to cut that list a little more, but I needed more than Bloglines. I did the same set of "reach" measures as in 2005, with one minor tweak and one significant addition. Bloglines OPML output translates directly into a spreadsheet that made it easy to search for links: Highlight the URL cell, copy, paste into the "link:" search, and go. As with last year, I checked link: counts in Google and MSN Search—but this time I used Yahoo! instead of AllTheWeb. I then added one figure that I believe is more meaningful than any of these three: the *visible* result from Yahoo!

What's the visible result? The number of sites Yahoo! *shows* you with its "very similar" algorithm active. Anyone who's spent time looking at web search engines knows that any result count greater than 1,000 represents a claim, because the search engines won't show you more than 1,000 results. In practice, deduped results usually aren't anywhere near the 1,000-result limit. Yahoo! will show 100 results per page and give an accurate count of results displayed on the last page it shows; it also offers larger link: results than the other two engines. That made it an obvious choice. The deduped number is nice because it reduces the echo-chamber effect of blogrolls, where the presence of a blog on another blog's blogroll may result in hundreds of apparent links, only one of which is significant.

Consider the three raw link: results, noting that I had *already* removed 90 blogs likely to have very high link: results (and nearly 200 likely to have relatively low link: results):

➤ **Google**: The highest number was 5,370 (compared to 9,430 last year); several had no link: results at all.

➤ **MSN**: The highest number was 34,669, compared to 76,675 last year; again, several had no link: results.
➤ **Yahoo**!: Every candidate had at least five Yahoo! links; the high was 179,000 (compared to 449,000 last year).

These numbers don't mean much of anything, particularly given the skew of blogrolls. What can you do with ratios of 2,600:1 (Yahoo!) even after you've eliminated extremes? I'd concluded that last year's Reach numbers weren't very good. Using that same formula yielded a smaller range this year—from a high of 13,497 to a low of 84, a ratio of 161:1, considerably smaller than last year's 7,778:1 but still too broad given that this year's 161:1 omits the liblogs likely to have the highest figures. A slightly modified version of last year's formula, using adjusted deflators, yielded a range of 10,590 to 82, a ratio of 128:1.

The "visible Yahoo!" number had a good feel to it (and an upper limit of 1,000). I calculated a new Reach factor, adding the visible Yahoo! count to twice the Bloglines subscriber total. That yielded a high of 1,387 and a low of 51—a ratio of 27:1.

This time, there *were* obvious outliers. By dropping nine blogs with new Reach factors above 700 and 21 below 70, I had a ratio of only 10:1 for a candidate pool of 251 blogs.

Since 2006

Spelling out the 2006 Reach factor, it was (**R = Y! + 2B**), where Y! is the visible Yahoo! count, B is the Bloglines subscriber total and R is the new Reach factor—a much simpler number than in 2005.

But I also made one other change, fundamentally more important. I didn't identify the Reach for any given liblog—and I didn't use it to arrange liblogs or in any other manner. The *only* use was to define the "great middle"—the universe of liblogs studied in 2006.

That meant listing the blogs in alphabetical order, which isn't ideal, but seemed (and seems) preferable to a calculated number that had an increasingly arbitrary feel to it.

In 2007, I did two book-length studies of **library** blogs—that is, official blogs from academic and public libraries, respectively. I didn't use Reach as a filter—but I did look at Visibility, a very similar idea.

Things had changed so much that the whole concept behind Reach had become tenuous:

➤ AllTheWeb was gone or no longer interesting.

- ➢ Google pretty much admitted that their link: numbers represented some unstated portion of actual links, making them useless as a metric.
- ➢ Yahoo! started redirecting links: searches to its Site Explorer—and, at the time, I found the interface and numbers inscrutable. (It's improved somewhat since then, but it would still be a lot of effort to determine the visible number in Site Explorer.)
- ➢ With Google Reader growing, it was hard to say whether Bloglines subscriptions was a representative figure.

The Visibility algorithm

I needed something that didn't require too much effort—and something that "made sense." In the end, I used two numbers that popuri.us would report in a single step: Bloglines subscriptions (apparently including all feeds) and Technorati links.

I also recognized that visibility isn't linear: A blog with 1,200 Technorati links and 1,000 Bloglines subscriptions is not "ten times as visible" as one with 120 links and 100 subscriptions—it's perhaps twice as visible.

There's an easy way to handle cases where the significance of a change becomes relatively smaller as the numbers involved in the change get larger: Logarithms. The Visibility Factor, then, was the base-10 logarithm (Log10) of the sum of Technorati links and Bloglines subscribers, both as reported by popuri.us.

That algorithm provided a nicely narrow range, from 0.0 to 5.2 for library blogs—but with only one blog exceeding 3.9. It was easy to characterize the scale. Anything less than 1.0 was essentially invisible (the combination of subscribers and links from other blogs is less than 9). 1.0 to 1.9 was "slightly visible." 2.0 to 2.9, generally the most widely-populated range, was "visible." 3.0 to 3.9 was "widely visible"—and anything over 4.0 "highly visible."

I used those numbers and phrases in the two library blog books and had planned to use them for this study. But things changed.

- ➢ Bloglines frequently failed to report subscriber counts to popuri.us queries.
- ➢ Technorati would sometimes fail to report link counts.
- ➢ I felt the need for Google Reader counts.
- ➢ I became aware that Technorati's link count wasn't a great number—that the *authority* number (the number of other blogs linking to a blog within the last six months) was much less susceptible to gaming and a better indicator in general.

So I tried using the Technorati authority, Bloglines subscription count and Google Reader subscription count as a basis for a Visibility number. I'd already decided *not* to use Visibility within the study, using it only to avoid "invisible" blogs.

Things get worse

That worked for a while—but the Bloglines and Google Reader counts got harder and harder to get as time went along. In the final process of building the 2007-2008 universe, I dropped the Visibility algorithm altogether and used a simple yes/no test: A liblog I hadn't seen before needed a Technorati Authority of at least seven (which means seven other blogs have linked to it within the past six months) to be included. That limit was used in adding the last 35 liblogs; the tweaked Visibility algorithm was used for more than 90% of those included.

Something Better?

Where do things stand now and for the future?
- ➢ I wasn't happy with that seven cutoff.
- ➢ I'd already decided *not* to use any measure of reach or visibility in future studies (including this one) except as a yes/no filter.
- ➢ I needed something simple, easy to find, and good enough.

Google Page Rank

An answer was staring me right in the face, every time I used Firefox: Google Page Rank (imported from the Google toolbar for Firefox). It shows up *automatically* when I'm checking a site for other things (e.g., having at least one post in the past year), it's in a narrow range, and it seems plausible.

On October 16 and 17, 2008, I checked all 607 blogs in this universe for Google PageRank. Seven blogs are no longer reachable (the sites have disappeared). In a few other cases (maybe ten), Google doesn't return a number, which I had to interpret as zero.

In all other cases, the Google Page Rank showed up as a number from one to seven. (The number can go up to 10, but other than Google itself, I don't think I've ever seen a 10. Wikipedia and Yahoo! show up as nine; a few "A list" blogs show up as eight.)

I plan to use GPR if (when?) I do future studies—not to rank blogs but to eliminate the least visible. The question, then, is *what* GPR should be required? To help answer that question, I looked at the distribution of liblogs by Google Page Rank. I would want a cutoff that wouldn't substantially reduce the current

sample size, but would serve to eliminate blogs that seemed designed to stay hidden.

Here's what I found in October 2008:

Google Page Rank	Liblogs
0	20
1	2
2	8
3	67
4	188
5	212
6	90
7	13

Table 10.1: Distribution of liblogs by Google Page Rank
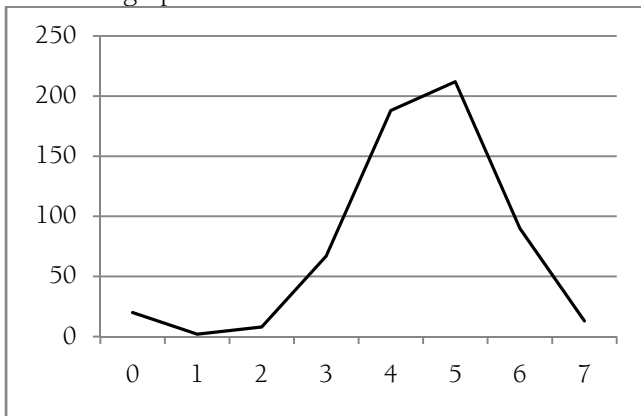
Here's the graphical version:

Figure 10.1: Distribution of liblogs by Google Page Rank

The vertical axis is number of liblogs, the horizontal axis Google Page Rank. It's a slightly lopsided bell-shaped curve, with most liblogs having GPRs of four or five.

Based on that distribution, a cutoff of two seems very inclusive—a cutoff of three less so. If I do another study, I'll probably use GPR 2 as a cutoff, combined with "recent activity" limits. A cutoff of two includes 578 of 600 liblogs, or roughly 96%. Someone wishing to be slightly less inclusive could certainly use GPR 3 (eliminating 1.3% of the liblogs), but I'd argue against any higher number as eliminating too many up-and-coming blogs.

Liblogs and the Larger Blogosphere

As I was preparing this study, Technorati released its *State of the Blogosphere / 2008* (technorati.com/blogging/state-of-the-blogosphere/). This year's report goes beyond earlier ones with lots of analysis based on a survey of 1,000 bloggers.

Most observers seem to focus on polls demonstrating how "mainstream" blogs have become (which I don't doubt) and the growth in blogging—and ignore history, even though Technorati provides a direct link to the 2007 report and earlier reports.

Here, then, a few facts about blogs and related facts about liblogs. I assume Technorati's actual numbers are factual; I see no reason to assume otherwise.

Mostly a Ghost Town

The "blogosphere" is much like Second Life: If you compare actual residents (active blogs) to counted residents (started blogs), it's mostly a ghost town.

What's "mostly"? 94% or more, depending on how you measure:

➢ The 2007 report said there were 70 million blogs as of April 2007, with 120,000 new ones emerging each day. If that 120,000 rate continued, there would be (or have been) about 120 million in June 2008, when the new study was done. The new study does not state the number of blogs or the number of new blogs each day–although it says "133 million blog records" since 2002, which presumably means 133 million blogs at some point. Technorati also quotes Universal McCann as saying that 184 million blogs have started as of March 2008. So let's say there are (or have been) somewhere between 133 and 180 million blogs.

➢ Meanwhile, Technorati says 7.4 million blogs had at least one new post within 120 days—a pretty modest measure of "active"—and just over 5 million posted in June. But if 120,000 new blogs were being created each day (each with at least one post), you could reduce that 5 million to a mere 1.4 million ongoing blogs. Of course, on that basis, 7.4 million is *smaller* than the number of new blogs during a 120-day period.

➢ Those figures make no sense, so let's be as charitable as possible and say between 5 and 7.4 million blogs are active, not just one-shot wonders. That's somewhere between 5.5% of 133 million and 3% of 180 million.

Liblog comparison

At the same time I checked for Google Page Rank (October 16-17, 2008), I checked for the most recent post. Nine blogs had either disappeared or no longer *had* posts.

Of the remaining 598 blogs, 86% would qualify as active using the 120-day cutoff: 514 in all. 71% of the 598 had at least one post within the preceding

month—423 in all, or 82% of the active blogs. But that excludes *all* new blogs—nothing that began in 2008 was included in the study.

Extrapolating, if the Technorati percentages hold for liblogs, there would be or have been somewhere between 9,300 and 14,000 English-language blogs begun since 2001. It's possible that there have been that many, but it seems unlikely.

40% Drop in Daily Posts Since 2007

That's the truly impressive Technorati figure–and it does appear to be a direct comparison:
- In April 2007, Technorati counted an average of 1.5 million posts per day.
- In June 2008, Technorati counted an average of 900,000 posts per day: 40% fewer.
- Looking back, Technorati reported 1.2 million posts per day in April 2006–and 900,000 in August 2005.
- In August 2005, Technorati reported 14.2 million blogs and said 55% of them–or 7.8 million–were active. If that's right, the active blogosphere is basically as active as four years ago, with a lot of churn in between.
- That also comes out to about one post every eight days for the active blogosphere (although of course the average doesn't exist–only 1.5 million blogs had posts in a seven-day period)

Compare that with liblogs:
- For March-May 2007, 523 blogs had countable posts, for a total of 22,969 posts.
- For March-May 2008, 533 blogs had countable posts, for a total of 19,616 posts. (That 533 doesn't include 54 blogs with posts in 2007 but not 2008–and includes 64 new blogs and blogs without posts in the 2007 quarter).
- That's a drop—but a drop of 8.5%, which is a whole lot better than 40%!
- Those 533 blogs averaged about 213 posts per day as a whole, or about one post every 2.5 days per blog.

Overall, liblog posting declined at a much slower rate than blogs as a whole–and active liblogs are about three times as active as blogs as a whole.

Most Recent Posts

Here's what I found scanning the blogs (including those with no posts in March-May 2008) on October 16 and 17, 2008:

Most recent post	Blogs	Percent	Cumulative %
1-2 days	159	27%	27%
3-4 days	95	16%	42%
5-7 days	39	7%	49%

8-14 days	70	12%	61%
15-30 days	60	10%	71%
31-61 days	51	9%	79%
31-92 days	24	4%	83%
93-120 days	16	3%	86%
121-182 days	30	5%	91%
183-365 days	40	7%	98%
More than a year	14	2%	100%

Table 11.1: Days from most recent post

In this table, "1" means "same day" and "2" means "day before" and so on. The cumulative percentage sums up all the intervals to that point.

Other Tidbits

- One-third of bloggers in general operate anonymously or with pseudonyms. That compares with 18% for liblogs.
- Roughly 85% of blogs as a whole have comment systems. Roughly 20% of liblogs didn't have any comments in 2008–but that includes blogs that don't allow them and blogs that just didn't have them.

What's the Message?

While liblogs have fewer posts now than a year ago—and for both liblogs and blogs in general, it appears that the peak was probably early 2007—liblogs are doing much better than blogs as a whole.

And when somebody blathers about hundreds of millions of blogs or says "everybody will blog in the future," feel free to ignore them. It's trivially easy to start a blog—but a lot of people (95%? 97%?) find, sooner or later, that they really don't have much to say that belongs in a blog. Why should that be a surprise?

If There is a Next Time

Will I do another study of the liblog landscape? That depends on a number of factors—sponsorship, book sales, phase of the moon, other good or bad uses of my spare time.

I do plan to trim the current spreadsheet for possible use as a basis for another study—and that trimming is based on changes I'd make next time around:
- As noted in Chapter 10, I'd use Google Page Rank of 2 (or higher) as a cutoff for visibility—unless a blog was already in the universe and met other criteria.
- The handful of non-English-language blogs retained from 2006 will go away next time. They deserve separate studies.

- ➢ I probably wouldn't bother to look at use of illustrations. They're too heterogeneous in nature and I don't think the numbers say much of anything.
- ➢ Finally, it does make sense to purge the study of truly inactive blogs. It's tempting to use the standard I used for library blogs (at least one post in two of the three months of the study), but that may cut out some interesting, if *very* infrequently updated, blogs. It also makes sense to include blogs that *were* active but have since gone quiet, but only for one year.
- ➢ For now, at least, I'll plan on a very generous measure of "active": A blog must have at least one post within the last 180 days. Blogs in the current study that fail to meet that measure *and* had no posts in March-May 2008 will be purged. That can't possibly remove more than 63 blogs (including nine that no longer have any visible posts at all), just over 10% of the total.

The remaining chapters provide brief factual profiles for each liblog in this landscape. I don't offer my opinions for several reasons (including, frankly, not really having much to say about roughly half of the blogs because of my own interests and limitations).

 Liblogs aren't being born as rapidly as they once were and aren't being used as frequently by bloggers—but they're being used well and too good effect. The shiny new toy has become a slightly dulled, enormously workable tool, surrounded by other tools that work better for some uses. If you write a liblog, these are good times to be doing it. If you're thinking about it and you have something to say—well, there's always room for one more.

The Rest of the Story

The Liblog Landscape 2007-2008: A Lateral Look is available as a $35 trade paperback from Lulu (www.lulu.com/content/4898086) or Amazon, and may still be on sale at Amazon for $27.76. The Amazon/CreateSpace edition has ISBN-13 978-1440473845.

 The book includes larger graphs and profiles of each of the 607 liblogs, as well as an index.

Other changes from the book

Most tables have larger type in the book. I changed most of them from 10-point type to 9-point type in order to fit more tabular columns into the narrower columns of *Cites & Insights*.

Quite a few tables in the book end with percentage columns that were omitted in this issue because, even with smaller type, they wouldn't fit.

A few textual changes have been made: "About the same" changed to "Same" in a number of subheadings, "from 2007 to 2008" changed to "2007-2008" in a few subheadings.

In lists of blogs standing out for one metric or another, the book uses blanks to separate subgroups of blogs within a list and for easier reading. Those blank lines were removed here—part of the process of bringing 55 C&I pages (121 pages in the book) down to the 47.5 pages you see here (not including added material, "The Rest of the Story").

In a few cases, references to chapter numbers may be a little mysterious—but since table and figure numbers are retained and begin with chapter numbers, you should be able to figure out what's what

Why This Issue?

The Liblog Landscape 2007-2008: A Lateral Look is by far the most complete review of changes in blogging behavior within the library field, and quite possibly in any field.

 For all that, the book has not sold enough copies either to justify further study (which might happen anyway) or, perhaps more significantly, to spread the word. So this free version of the first 11 chapters of the book serves two purposes:
- ➢ It makes the results of this survey available to the field as a whole.
- ➢ It's another test of Andersonomics, the idea that free digital content will yield income.

Masthead

 URL: citesandinsights.info/civ9i7.pdf