Considering the Sources:
Comparing Linking Patterns in Usenet and Blogs

Mary McGlohon[*]
Carnegie Mellon University
5000 Forbes Ave.
Pittsburgh, PA 15213
mmcgloho@cs.cmu.edu

Matthew Hurst
Microsoft Live Labs
1 Microsoft Way
Redmond, WA 98005
mhurst@microsoft.com

Abstract

Usenet is a decentralized discussion community predating blogs by decades. Just as there are a wide range of political blogs, many Usenet sub-communities focus on politics. However, we find these two communities are very different in terms of source content. In this work we compare linking patterns of the political Usenet with a well-known political blog and news aggregator, Memeorandum, with respect to *coverage* of news and blog sources, *attention* of stories, and *timeliness* of links.

Introduction

Usenet, arguably the first online decentralized discussion system, has remained an active community in the midst of other newer social media. To better understand how the community functions and which topics and sources are considered important, it is useful to make comparisons between the linking behavior occurring in Usenet compared with that of the blogosphere.

In this work, we obtain data from politically-related Usenet newsgroups and from a widely-used political blog and news aggregator Memeorandum. We study the rate of activity in Usenet over time, and compare which news and blog domains receive the most *coverage*– that is, which sources are most trusted by readers. We next consider the question of *attention* to these stories in both Usenet and Memeorandum. Finally, we consider *timeliness*: we analyze timestamps on the links and identify where in Usenet the most timely conversation occurs.

Preliminaries

Related Work

Microsoft's Netscan Project has conducted a very thorough study of Usenet discussion patterns, depicting hierarchy of newsgroups and their changes between 2000 and 2004 (Turner et al. 2005); studying the *social roles* of Usenet authors (Fisher, Smith, and Welser 2006); and visualizing them (Viegas and Smith 2004). Other related work has examined the relationship between blogs and mainstream media, setting up a system that identifies emotionally charged news articles for communities of different political orientation (Gamon et al. 2008). Other work has concentrated on political discussion networks and their influence on the election, by studying a large set of political blogs. Authors studied linking patterns in two distinct clusters of "conservative" and "liberal" blogs (Adamic and Glance 2005).

Data

Political Usenet: We collected data from nearly 200 newsgroups with posts between 2004 and 2008, using a subscription service. In the interests of capturing a representative subset of data relating to political discussions, we selected all newsgroups available with the substring "polit" in the name. Details may be found in (Mcglohon and Hurst 2009). In these newsgroups combined, there were 37 million articles (19.6 unique articles not including cross-posts). Unless mentioned otherwise, we multi-counted cross-posts; ignoring them did not usually affect aggregate rankings.

Memeorandum: We used Memeorandum as a sample of the blogosphere[1] Self-described, "Memeorandum offers you a window into this new world of news, focusing primarily on U.S. politics and current affairs. It auto-generates a news summary every 5 minutes, drawing on experts and pundits, insiders and outsiders, media professionals and amateur bloggers." A screenshot and description is shown in Figure 1. Essentially Memeorandum aggregates top news stories, and discussions of those stories both news and blogs. The Memeorandum data covered September 2005 through June 2008. (For comparative analysis, we only used Usenet data from that time period).

Link classifier

For comparison, we also classified links as news or blog content. This is done via a knowledge- or rule- based mechanism, centered on a large list of known news/blog sources. This list is augmented with specific rules that identify news articles per site, eliminating spurious links to other types of content. The data consisted of 35 million hyperlinks

---

[1]www.memeorandum.com. We use a third-party aggregator for the sake of easy reproduction of results: using an aggregator obtains a representative sample that is politically oriented.

Figure 1: A Memeorandum screenshot. The top "parent" story, *Poll shows McCain in a 5-point lead over Obama* appears, with a number of "discussion" links, typically blog posts linking to the parent.
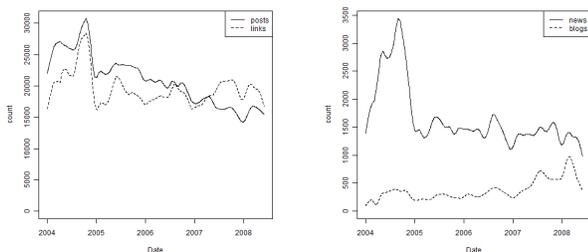

Figure 2: Top: Aggregate posting and linking rates in the political Usenet newsgroups. Bottom: News and blog linking rates over time. Both are smoothed using kernel smoothing.

among all articles, including cross-posts. 3 million were news sources, while around 700,000 were blog posts. We next discuss how these links were distributed over time.

Link and domain occurrence and coverage

Rate of posting and linking in Usenet

Overall posting activity in Usenet is declining (Turner et al. 2005); the same occurs in our subset of political newsgroups. We show aggregate posting and linking rates in Fig. 2. The steep drop after 2004 is a result of the "election bump" that the (mostly American) newsgroups experienced– groups such as `can.politics` experienced spikes in the time periods before their own elections. Some groups did not even experience an overall decline in posting; `it.politica` in fact increased in posting frequency between 2004 and 2008. It is also notable that the aggregate link activity appears steady– this could perhaps indicate that users are relying more on outside content or links within Usenet itself. With respect to the external media linked, we notice that blog linking increases, while mainstream news decreases. Potential reasons for this are numerous– not only are blogs prospering as a medium, but the mainstream media are beginning to adopt blogs as part of their content (for example, the New York Times blogs) where they formerly had "columnists".

Knowing where users are linking indicates what they con-

sider to be most important or credible. With this in mind, we next drill down into which of these news and blog domains have the most coverage, and how that compares to what we find in Memeorandum.

Coverage of news and blog domains

We analyzed the link rates to different domains. Note that in order to have an accurate correspondence between Memeorandum and Usenet, since Memeorandum is a single source, we needed to count each unique link once; repeated links in Usenet were not counted. The following is a comparison of the top news domains linked in Usenet and Memeorandum. For each domain, the rank in the other data set is shown in parentheses.

News Domains with Most Coverage		
Rank	Usenet domain	Memeorandum domain
1	news.bbc.co.uk (18)	nytimes.com (3)
2	washingtonpost.com (2)	washingtonpost.com (2)
3	nytimes.com (1)	news.yahoo.com (6)
4	guardian.co.uk (17)	themoderatevoice.com (699)
5	cnn.com (8)	corner.nationalreview.com (489)
6	news.yahoo.com (3)	newsbusters.org (191)
7	newsmax.com (72)	latimes.com (12)
8	epochtimes.com (NA)	cnn.com (5)
9	msnbc.msn.com (9)	msnbc.msn.com (9)
10	repubblica.it (NA)	weeklystandard.com (135)

Blog Domains with Most Coverage		
Rank	Usenet domain	Memeorandum domain
1	smirkingchimp.com (549)	thecarpetbaggerreport.com (35)
2	crooksandliars.com (11)	hotair.com (68)
3	freerepublic.com (706)	captainsquartersblog.com (71)
4	thinkprogress.org (4)	thinkprogress.org (4)
5	globalsecurity.org (NA)	bluecrabboulevard.com (1675)
6	dailykos.com (6)	dailykos.com (6)
7	democraticunderground.co (308)	talkleft.com (58)
8	lewrockwell.com (314)	firedoglake.com (64)
9	nraila.org (2225)	powerlineblog.com (40)
10	newshounds.us (165)	washingtonmonthly.com (33)

There are a few outliers: many non-U.S.-based news sources rank much lower in Memeorandum. As Memeorandum is largely concerned with American politics while our Usenet data contains a wide range of countries and languages, this is not surprising. However, even when we focus only on some newsgroups dedicated to US politics (such as `us.politics`), we find that the BBC site ranks in the top five. Differences between blog link rates are even more pronounced. We notice that most top-ranking Usenet blogs scarcely appear in Memeorandum at all. While news domains are somewhat correlated, blog domains show little correlation between Usenet and Memeorandum.

Coverage of political factions

One can also divide the blog links by political affiliation. Using the white list of the BLEWS system (Gamon et al. 2008), we created "A lists" for both Usenet and Memeorandum– that is, the most-covered bloggers from both the conservative and liberal factions. What is most striking is the differences in numbers of links to liberal versus conservative blogs. In total there were 10 times more links to liberal blogs

than there were to conservative blogs. We find no such bias in Memeorandum. "A Lists" are shown below:

Usenet "A List" Bloggers

Liberal	Links	Conservative	Links
commondreams.org	102597	jihadwatch.org	9003
truthout.org	75393	newsbusters.org	4805
alternet.org	51141	rightwingnation.com	4141
antiwar.com	39963	americanthinker.com	3734
crooksandliars.com	35719	danielpipes.org	2745
democraticunderground.com	33339	powerlineblog.com	2602
rawstory.com	29402	michellemalkin.com	2520
indybay.org	26731	coxandforkum.com	2189
huffingtonpost.com	24884	spectator.org	2085
thinkprogress.org	24774	scrappleface.com	1823

Memeorandum "A List" Bloggers

Liberal	Links	Conservative	Links
thecarpetbaggerreport.com	5575	hotair.com	5207
huffingtonpost.com	4816	bluecrabboulevard.com	4026
themoderatevoice.com	4513	corner.nationalreview.com	3846
thinkprogress.org	4196	powerlineblog.com	3360
dailykos.com	3668	outsidethebeltway.com	3213
talkleft.com	3667	newsbusters.org	2863
firedoglake.com	3437	gatewaypundit.blogspot.com	2731
washingtonmonthly.com	2979	michellemalkin.com	2465
crooksandliars.com	2897	redstate.com	2428
mydd.com	2853	riehlworldview.com	2335

Interestingly, in the Usenet sample, links to message boards outnumbered links to a traditional blog in both liberal and conservative spheres, which might suggest that, Usenet participants preferred content that is more Usenet-like.

Attention to links

Since Memeorandum is a single source while the Usenet data set has nearly $500,000$ authors, we chose to use unique counts for a story in the previous comparison of domain counts. However, there needs to be a measure of how much attention a certain story or source is given.

Attention in Memeorandum

For Memeorandum we can measure attention by how many "children" each parent story has. We find that *salon.com* was disproportionately sparking a great deal of discussion, compared with the coverage on the page in general. In contrast, *thecarpetbaggerreport.com* does not even show up in the top ten, while it was the most "covered" blog. Rankings are below:

Memeorandum Most Discussed News/Blog Sources

Rank	News Domain (Children)	Blog Domain (Children)
1	nytimes.com (83319)	michellemalkin.com (9162)
2	washingtonpost.com (56585)	thinkprogress.org (8942)
3	news.yahoo.com (19643)	hotair.com (7408)
4	cnn.com (10971)	captainsquartersblog.com (7116)
5	msnbc.msn.com (8910)	blogs.abcnews.com (5229)
6	latimes.com (7545)	firedoglake.com (3929)
7	abcnews.go.com (5351)	powerlineblog.com (3796)
8	salon.com (4320)	crooksandliars.com (3499)
9	foxnews.com (3892)	blog.washingtonpost.com (3426)
10	corner.nationalreview.com (3782)	talkingpointsmemo.com (3329)

In Memeorandum, the most followed individual links were scandal-related: one about presidential candidate Mc-Cain [2] with 219 related stories, one about New York Governor Eliot Spitzer [3] with 178 related stories, and one about Attorney General Alberto Gonzales' resignation [4] with 170.

Attention in Usenet

There are also two ways we identified to measure Usenet attention: a) how many times a link has appeared (since we have multiple sources), and b) how much the initial link is "followed" by other threads. Total domain counts are similar to those in the last section, and omitted for space. The "most followed" domains, however, present some interesting differences. Overall, the news domains were similar to what we found to be the most-covered news domains. The most followed blog domains were more varied. The rankings for each medium are below:

Usenet's Most Followed News/Blog Sources

Rank	News Domain (Replies)	Blog Domain (Replies)
1	news.bbc.co.uk (649508)	xoverboard.com (167101)
2	cnn.com (649090)	freerepublic.com (142881)
3	commondreams.org (623091)	whatreallyhappened.com (123414)
4	washingtonpost.com (619811)	thinkprogress.org (96429)
5	nytimes.com (508034)	takebackthemedia.com (84277)
6	guardian.co.uk (392954)	dailykos.com (75913)
7	msnbc.msn.com (357416)	globalsecurity.org (58821)
8	worldnetdaily.com (239113)	buzzflash.com (54165)
9	usatoday.com (221236)	balloon-juice.com (52601)
10	news.yahoo.com (215564)	lewrockwell.com (50726)

One thing to notice is that the lists do not have the same set of sources. For instance, `xoverboard.com`, a webcomic, is the most followed blog domain, but it ranks 13 on the total counts: with $9,524$ independent links, an "explosion" factor of around 17.5. What is perhaps more surprising is that almost all of these independent links were to a single post, a comic depicting an "Internet Argument"[5], which garnered nearly $167,000$ replies.

While the bloggers Memeorandum covers focus on scandals, it would appear Usenet users care most about conspiracy theories. The second most followed link was a post suggesting that President Bush had a Nazi affiliation [6], with $69,000$ posts in reply, followed by posts about World Trade Center conspiracies. The most followed news stories included of a column suggesting that the Bush family benefits from the Iraq war [7], and an inflammatory conservative commentary on the Mark Foley scandal [8]. Note that these are not necessarily a single discussion (the largest single thread we had contained 9,000 replies), but rather multiple threads and cross-posts, so these stories were given wide visibility.

Timeliness

The previous two sections demonstrated that the participants in the political Usenet and the political blogosphere have different reading/linking patterns. It is of interest to not only

---

[2] `nytimes.com/2008/02/21/us/politics/21mccain.html`

[3] `nytimes.com/2008/03/10/nyregion/10cnd-spitzer.html`

[4] `nytimes.com/2007/08/27/washington/27cnd-gonzales.html`

[5] `xoverboard.com/cartoons/2007/070416_argument.html`

[6] `takebackthemedia.com/bushnonazi.html`

[7] `www.commondreams.org/views.shtml?/views04/0223-08.htm`

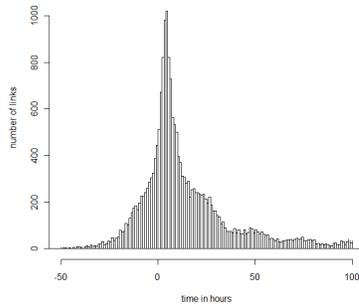[8] `worldnetdaily.com/news/article.asp?ARTICLE_ID=52244`

Figure 3: Distribution of timeliness in Usenet. The 0 point on the x-axis indicates the time at which the story appears on Memeorandum.

consider *what* is being linked to, but also *when* it is being linked. We use Memeorandum as a measure of when news stories appear on the blogosphere. We say that a Usenet author is "scooping" Memeorandum if they post a link to a news/blog before Memeorandum does. This is not an exact measure of where stories break first, but it is useful for determining whether Usenet users are breaking and discussing stories independently, or simply rehashing blogs.

Figure 3 shows a histogram (bucketed at hourly intervals) for how long before or after Memeorandum a link occurs in Usenet. We show counts from 50 hours before Memeorandum to 100 hours after. Of 20262 shared unique links in that range, and of those, 4606 occur before Memeorandum. Roughly, the mode of the distribution is at 4 hours after memeorandum. Around 22 percent of the time, when the domains have shared links, Usenet scoops Memeorandum.

When drilling down to newsgroups, we find some groups tend to break stories before Memeorandum more often than others do. Ranking is omitted for space. As expected, groups with highest linking volume generally scoop the most; however, there are some notable exceptions, `mn.politics` and `az.politics`. `mn.politics` is the 79th most-linking group and `az.politics` is 44th. However, when it comes to first-links, the two groups rank 7th and 10th, respectively. This suggests that there may be value in looking particularly at local political groups, which may provide a more concentrated view of a smaller sphere.

Conclusion

This is the first study that compares the political segment of Usenet to the political segment of the blogosphere over such an extended period of time. We have examined the political Usenet with respect to *coverage*, *attention*, and *timeliness*, and drawn comparisons with a blog/news aggregator. We report the following observations.

- While posting rates in the political Usenet are declining for most (but not all) newsgroups, linking rates remain steady. Linking rates to news are gradually declining, while links to blogs are increasing.

- Usenet has different linking patterns from that of the traditional A-list bloggers (when comparing a subset with a 2004 study) and blog aggregator Memeorandum, particularly in the choice of blog links: Usenet users often seem to prefer linking to non-Usenet discussion boards on the Web over traditional blogs. There is also a large liberal bias in Usenet which is not observed in Memeorandum.

- We compare the times at which links occur in Usenet to the times occurring in Memeorandum, and find that several times Usenet is "scooping" Memeorandum. While there is a general trend of higher-traffic newsgroups doing the most scooping, we find that some lower-traffic groups often have high timeliness.

Our findings raise a number of other interesting ethnographic questions and opportunities for application. Why does Usenet have such a heavy liberal bias? In the emergence of new media, why do some Usenet groups maintain a high level of activity while others decline? How do we mine Usenet to extract timely stories? The knowledge gained in this study can serve as a basis for exploring these questions.

Acknowledgments

References

Adamic, L. A., and Glance, N. 2005. The political blogosphere and the 2004 U.S. election: divided they blog. In *LinkKDD '05: Proceedings of the 3rd International Workshop on Link Discovery*, 36–43.

Fisher, D.; Smith, M.; and Welser, H. T. 2006. You are who you talk to: Detecting roles in usenet newsgroups. In *HICSS '06: Proceedings of the 39th Annual Hawaii International Conference on System Sciences*. Washington, DC, USA: IEEE Computer Society.

Gamon, M.; Basu, S.; Belenko, D.; Fisher, D.; Hurst, M.; and Konig, A. C. 2008. Blews: Using blogs to provide context for news articles. In *International Conference on Weblogs and Social Media*.

Mcglohon, M., and Hurst, M. 2009. Community structure and information flow in usenet: Improving analysis with a thread ownership model. In *International Conference on Weblogs and Social Media*.

Turner, T. C.; Smith, M. A.; Fisher, D.; and Welser, H. T. 2005. Picturing usenet: Mapping computer-mediated collective action. *Journal of Computer-Mediated Communication* 10(4).

Viegas, F. B., and Smith, M. 2004. Newsgroup crowds and authorlines: visualizing the activity of individuals in conversational cyberspaces. In *Proceedings of the 37th Annual Hawaii International Conference on System Sciences, 2004.*, 10 pp.+.