

Bloggers during the London attacks: Top information sources and topics

Mike Thelwall

School of Computing and IT
University of Wolverhampton
Wolverhampton WV1 1SB
+44 1902 321470

m.thelwall@wlv.ac.uk

ABSTRACT

Blogs are probably most associated with the high profile postings of a few highly popular bloggers who debate or comment on major news stories, but for each 'A-lister' there are numerous faceless bloggers who write about their own daily lives and/or interests. Hence it is interesting to investigate the extent to which an event with extensive media coverage, such as the London attacks, is reflected in blogspace as a whole. This paper reports a descriptive analysis of blog postings around the London attacks of July 7, 2005. The core of this study is the development of methods to identify and report on bloggers' activities in a way that is not dominated by prolific bloggers or repetitive blog postings. We report daily trends for the top links and topics for three sets of data: all bloggers' postings; the postings of bloggers who mentioned London at least once; and the blog postings mentioning London. Although only 5% of active bloggers ever mentioned London by name, the attacks appeared to be the most significant event in blogspace during the two weeks after the initial bombings. Bloggers who posted about London were found to be atypical, linking and posting much more frequently than general bloggers. The results suggest a dichotomy between externally-focused, news-aware approximately daily bloggers and internally-focused diary-like approximately weekly bloggers.

1. INTRODUCTION

Blogs are web sites that contain reverse-chronological order dated postings. They are used for a wide variety of purposes, from personal diaries to political campaigns, and a number of 'A-list' bloggers now have a significant online readership. The ease of creation and zero cost allows blogs to be adopted by a wider section of the population than any other major web publishing technology, albeit still a comparatively computer-literate and privileged subset of the world's population. Blogs have become an important mechanism through which information is communicated and openly discussed. As such blogspace provides a unique opportunity to study widespread information dissemination and discussion, including the kinds of topics that are most discussed and the main information sources used. Here, blogspace is defined as the collection of all blog postings.

There are various mechanisms that support blog discussions. First, many blogs allow readers to post comments in reply to individual postings. Second, bloggers commonly copy and quote others' postings, in order to respond to them, often creating a link to the original. Third, some blogs support 'trackback links', which are links to other bloggers' postings that discuss the original blogger's story. Finally, blogrolls are link lists posted in a navigation bar throughout a blog, often linking to a set of related bloggers and web sites. The most promising for identifying the most used information sources for bloggers, however, are

standard links in individual blog postings (i.e. not blogrolls, comments or trackback links). These have the advantage of being universally available in blogs (unlike trackback links).

This paper tackles the challenge of analyzing large-scale blog data to generate meaningful information about the behavior of blogs and bloggers in a given context (the London attacks of July 2005). The primary contributions are the introduction of analytical tools to define useful scopes for analysis (by slicing the data), and the introduction of appropriate processing methods to produce the most intuitively meaningful results to give findings that are not dominated by spam or by prolific blogs or bloggers. The focus of the case study analysis is to identify the top information sources used and topics discussed by bloggers during the London attacks.

2. LITERATURE REVIEW

2.1 Blogs and Bloggers

Blogs are usually put together using special purpose blog software and often hosted for little or no cost, for example at blogger.com or blog.com. The software makes it easy for individuals to post online content, typically using a simple interface allowing the blog owner to type their content into a box. Blogs can be used as a kind of online diary, and their ease of use supports their use for relatively informal publishing. As blog technology has evolved, so have the facilities available to bloggers. Currently (2006), a well appointed blog could expect:

- A home page containing the most recent postings, in reverse chronological order.
- An archive containing all previous months' postings.
- A unique URL (permanent URL) for each posting.
- A facility for visitors to record comments on postings.
- For each posting, a list of links to other bloggers' postings commenting on it and linking to it (via the track-back technology).
- On each blog page a list of links to favorite sites (the blogroll).
- Links and pictures within some postings.
- Pages of additional information, such as an author bibliography.

The variety of uses of blogs seems to be almost as diverse as human interests. Although blogs tend to be run by individuals and reflect their personal lives or interests, the community nature of many groups of blogs has attracted research to see how topics are organized or how online communities form and operate [1,2]. At the other extreme to personal blogs and small-scale groupings of bloggers, there is a small set of 'A-list' blogs that are widely-read enough to challenge traditional mass media in terms of influence. One example of this is Instapundit (instapundit.com), the blog of

Glenn Reynolds, a law professor at the University of Tennessee. Reynolds' blog has attracted over a million links, according to MSN search (Feb 3, 2006).

Academic interest in blogs seems to have spawned from three different directions. Computer scientists have developed methods to automatically extract public opinion from large numbers of blogs for market research [3]. Media theorists are interested in the extent to which blogs can challenge traditional mass media in the selection and reporting of news stories [4,5]. Sociologists have studied the dynamics of various blogger communities and blogger motives [6-9]. It is difficult to summarize the findings of the extremely diverse body of blog research, except by pointing to the wide variety of uses of blogs and the fact that blogs do sometimes create genuine online communities [1], and can serve to disseminate useful information and to be a valuable resource [10]. These findings should be interpreted in the context of the relatively affluent, US-centric nature of blogging [11], likely political biases amongst bloggers [12], the low quality, opinionated nature of the information in many posts [13].

2.2 Quantitative blog and RSS analysis

The largest scale studies of the blogosphere have been conducted by computer scientists, either using link analysis or word/term frequency analysis. The link structure of blogspace has rapidly expanded, displaying a burst nature, with rapid creation of links, and also a local community-like structure in terms of groups of bloggers tending to interlink [14,15]. The relatively sparse nature of linking prompted other researchers to analyze the text of blog postings tracking word frequencies as a way to find out about blog-based discussions [16]. For instance, counting the number of blog postings containing the word 'Microsoft' gives a measure of the profile of the company, and plotting this daily gives a time series showing the fluctuations of interest in the 'topic' Microsoft. Analyzing many topics through word frequencies in this way showed differing characteristics, including: spiky chatter, a constant level of discussion with occasional increase; or just spikes, sudden bursts of discussion around a topic, eventually dying away.

Two Really Simple Syndication (RSS)-based lines of research have focused on extracting significant events from blogs, using natural language processing techniques [17] or simple word frequency approaches [18]. These have shown that automated analysis of blog-like data has the potential for automatic extraction of relevant public-opinion-related information that can be targeted to specific applications, such as science policy debates.

2.3 Link Analysis

Away from blog analysis, information science methods have been developed to analyze the link structure of collections of web sites, mainly using universities as case studies [19-22]. A key finding is that all links are not equal. In particular, links between pages of the same web sites (i.e., site self-links [23]) are typically excluded because they are often used for navigational purposes. In addition, multiple links between a single pair of web sites are sometimes combined and treated as a single link for counting purposes [24] because some links are replicated throughout all pages of a web site, for example by inclusion in a standard links list in a site-wide navigation bar (like a blogroll). Finally, data cleaning can be needed to remove spam-like pages and mirror sites, and a classification of reasons why links have been created is needed to be able to effectively interpret the results of a link analysis exercise [25].

3. OVERVIEW AND DATA

This paper employs a variety of methods and each is described separately in conjunction with its results. The methods are described concretely and directly through the case study, rather than abstractly and separately from the specifics of the London data.

The data used in this paper is a collection of postings harvested by Intelliseek between 4 and 24 July, 2005 and made available to the research community [26]. For each post, the data recorded includes the text (HTML), its blog and post permanent URL, as well as its date and links. This set of data is described as the *full corpus* (1,417,711 blogs) and for the text analysis we also selected a random sample of 64,000 blogs to make processing manageable (64,000 blogs, 360,619 posts, 626,632 unique words; 5.63 posts per blogger). The subset consisting of all postings containing the word 'London' is the *London sub-corpus* (73,161 blogs, 151,331 posts, 443,752 unique words; 2.07 London posts per London blogger). The subset consisting of all postings made by bloggers that mentioned London at least once during the period is the *London posters sub-corpus* (73,161 blogs, 1,471,736 posts, 1,539,652 unique words; 20.12 London posts per London blogger). The statistics reported above are after spam removal (see below).

3.1 The main corpus and sub-corpora

The Intelliseek data is only a fraction of the whole blogosphere and we do not know how it was created, although presumably the creation process involved some link crawling i.e. automatic discovery of new blogs by following links from old blogs. Hence it is a convenience sample and its unknown generation method is undesirable for research. Nevertheless, its sheer size is some protection against bias although we suspect that link crawling and other technical problems specific to blog crawling make it likely that it is US-biased [27], even if unintentionally. The definition of the London sub-corpus is rather arbitrary, although it is a logical continuation of the topic concept used by previous researchers [16]. Some postings discussing the London attacks of July 2005 do not contain the city's name, and many unrelated posts do contain it. Nevertheless it is a convenient method for selecting a proportion of the corpus blogosphere that is likely to be related.

3.2 Link spam elimination and data cleaning

The data contains large numbers of links generated for advertising purposes, mainly for pornography and gambling. These links are undesirable from a data analysis perspective because they are relatively trivial in cause compared to other links. We imagine that at least half of the links were in fact created by a single individual spammer with the aid of a program. These are less important for analysis purposes than links that were created individually, particularly if the links are created to point to third-party information [25]. Hence we decided to remove from the data all blogs posting spam links. These were identified by (a) manual scanning of full link lists to identify anomalies (posts with many links or repetitive groups of links), and (b) using a program to list the top link targets, with manual investigations to check whether these top link targets were the result of spam. A program then filtered out all blogs that were directly identified as spam or ever hosted an identified spam (out)link. This process is imperfect and will miss advertising links that spread across a wide range of blogs, especially if those blogs also use links for non-advertising purposes. Nevertheless, this process appears to be effective in eliminating the most numerous sources of spam.

Links between pages of a single blog (blog self-links) were also eliminated. Following standard link analysis practice [25], we believe that these blog self-links are less useful for analysis purposes because links to an external blog or site are more definite indicators of the importance of the target; for example they are not normally for internal site navigation purposes, and presumably often represent an effort to seek out relevant information, rather than just recalling previous work (posts).

4. RESULTS

4.1 Basic descriptive statistics

Figures 1 and 2 give the overall activity levels of the three cleaned corpora. It is clear that there are large daily variations in the number of bloggers and blog postings in the data set. In particular the large falls in posts may indicate weekends (9-10, 16-17, 23-24 July) or data collection issues rather than natural changes in the blogosphere. Hence the analysis will focus on relative statistics rather than absolute statistics, where possible.

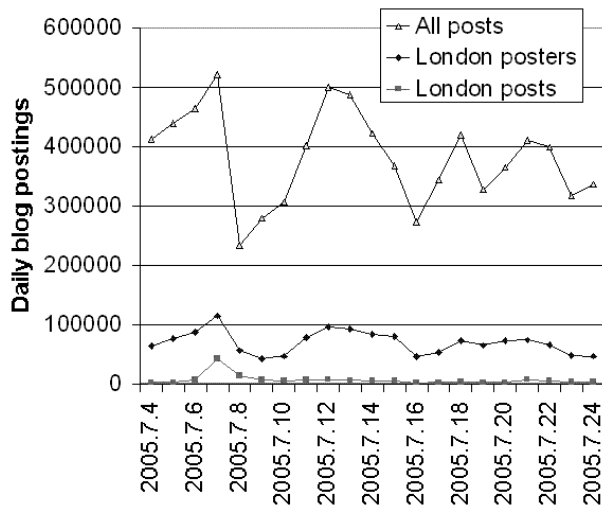


Figure 1. Daily blog postings for the three corpora.

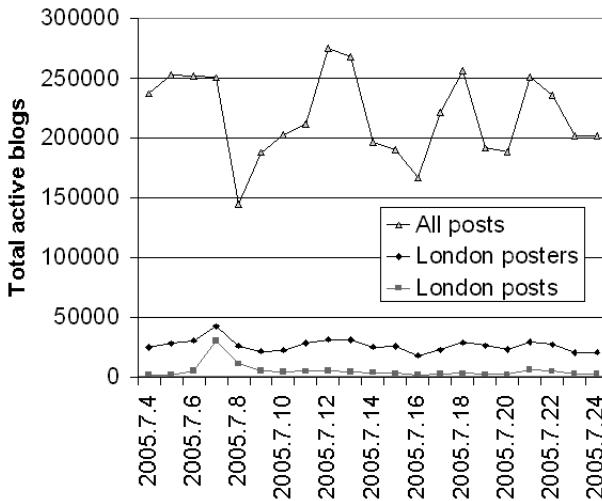


Figure 2. Daily active blogs in the three corpora.

Figure 3 gives the proportion of London-related postings in the corpus (lower line). This shows a clear peak at the time of the attacks (July, 7) and a second peak at the time of the second, failed attacks (July, 21). Figure 3 also shows the activity levels of

the London posters (bloggers who had made at least one post that included the word London). This level of activity does not follow that of London: it seems that these bloggers had significant topics to discuss that were unrelated to London. The same pattern is reflected in the proportions of active bloggers (Figure 4). The importance of the London attacks is much more marked in Figure 4: 12% of active bloggers mentioned London at the time of the attacks, although only 8% of posts mentioned it. The same comparison of graphs shows that the London posters are relatively active bloggers: they produce around 50% more postings than the proportion that they form of active bloggers.

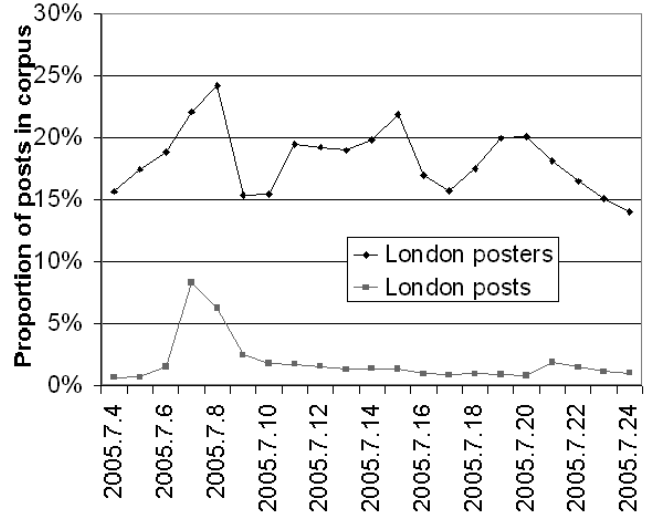


Figure 3. Proportion of postings in full corpus.

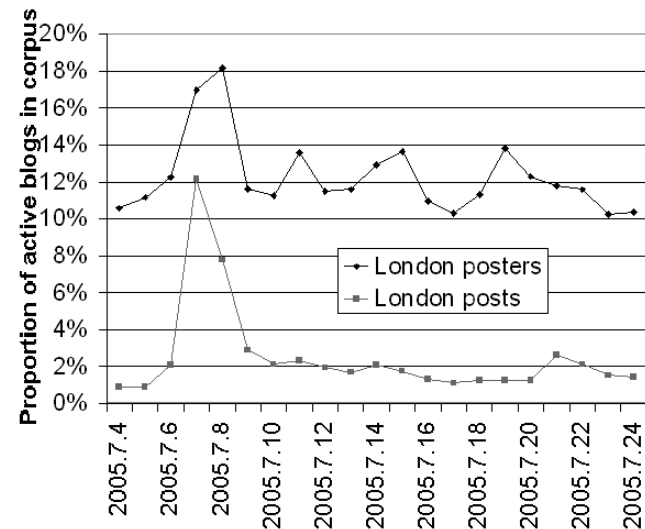


Figure 4. Proportion of active blogs in the full corpus.

4.2 Basic link statistics

Figures 5 and 6 give the total number of blogs with (out)links and the proportion of blogs containing links from each corpus. It is clear from Figure 6 that London posters are heavy link users, creating around four times as many as the average blogger. Nevertheless, it is also clear that the link creation is not a function of the London story, but is more in the nature of the bloggers: bloggers that mention London are likely to be high link users before, during, and after the events. Three of the five dips in the graphs coincide with weekends. It seems that many bloggers take

the weekend off and perhaps the average weekday blogger is different from the average weekend blogger; for example tending to link less (Figure 6).

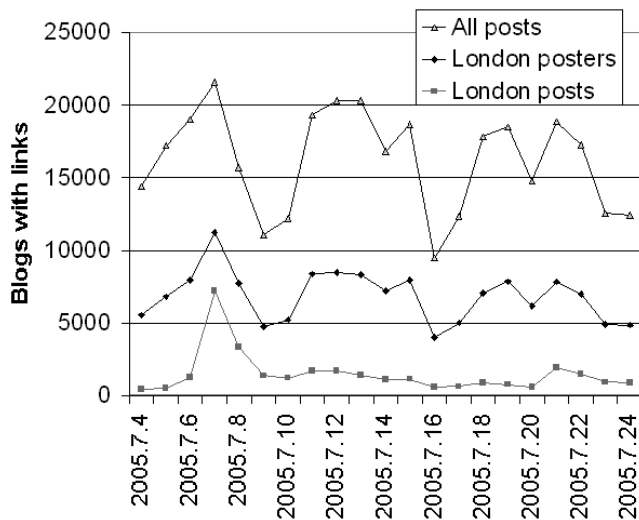


Figure 5. Daily counts of blogs containing links.

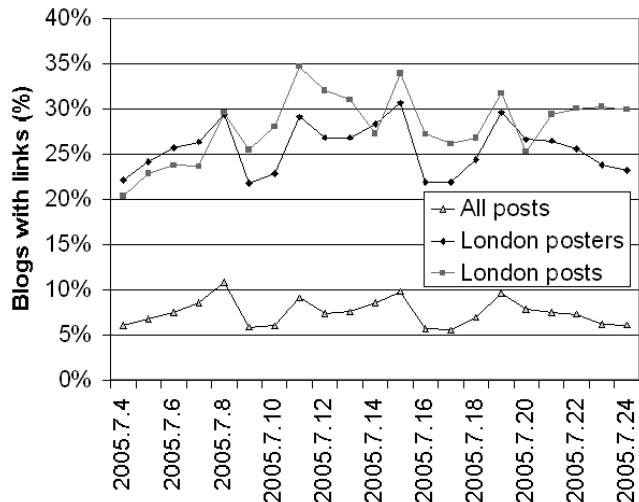


Figure 6. Percentage of blogs containing links.

4.3 Top information sources

The distribution of link target domains (not shown) follows on every day the hooked power law typical of links and many other internet phenomena [28,29], with a few domains attracting a significant proportion of the links and many domains only being linked to once. We wanted to know the main sources of information used by bloggers, and how these changed during the London attacks. For this, we truncated all link target URLs to their domain name and listed all the domains linked to by each blog (i.e. ignoring multiple links to a single domain, even if to different pages, or if the links were from different postings in the same blog). This process eliminates the effect of multiple, repeated postings of similar stories from individual blogs, and multiple postings of similar links from single blogs [24].

The BBC news web site, the New York Times and a UK newspaper, the Guardian, were top sources of London-related information for bloggers (Figure 7). Nevertheless, there was a huge fall in the hegemony of the traditional sources of London information on the day of the attacks. In fact there was a huge

increase in links to these sources when the attacks happened, but a very wide range of sources of information were used, not just the main ones. Hence, although the BBC in particular was a natural source of information about the attacks, the issue was perhaps so big that it could not dominate to the same extent as usual. From Figure 8, the top sources of information for London bloggers, the BBC is in fact the major news source and maintained its position at a constant level during the attacks. Other major news sources lost ground for London bloggers (in percentage terms) in the immediate aftermath of the attacks – presumably because of the increase in the range of sources linked to – but the major news sources quickly regained their position.

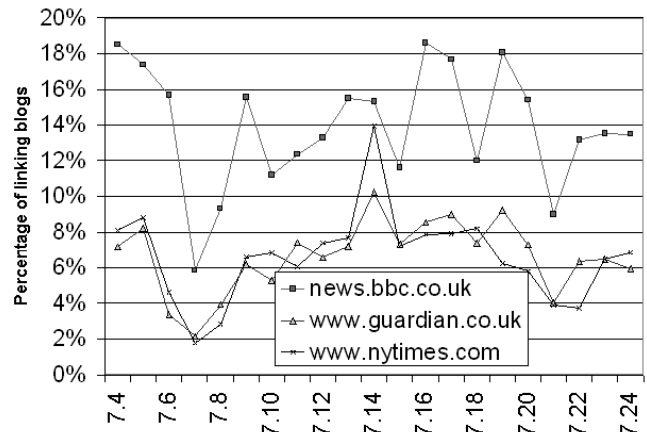


Figure 7. Top linked domains from postings containing the word London.

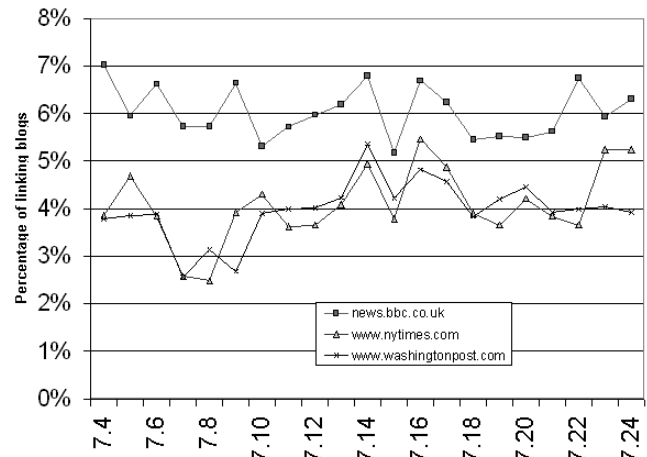


Figure 8. Top linked domains from linking London bloggers' postings.

Two more graphs give contrasting perspectives, serving to confirm the above findings. Figure 9 is identical to Figure 7, but reports the percentage of links per active blog, rather than per linking blog: the shapes of the graphs are very similar. Figure 10 shows all bloggers' linking to the four sites of Figures 7 and 8: again the graph is very similar in shape to Figure 8. The dominance of the BBC could either reflect leadership for web news or could be specific to blogspace: the BBC web site contained many interactive features that made it more blogger-friendly than many other news sites, at least in 2005.

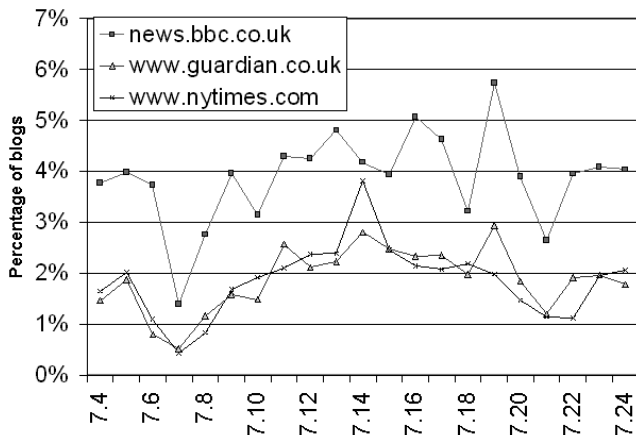


Figure 9. Top linked domains from London postings.

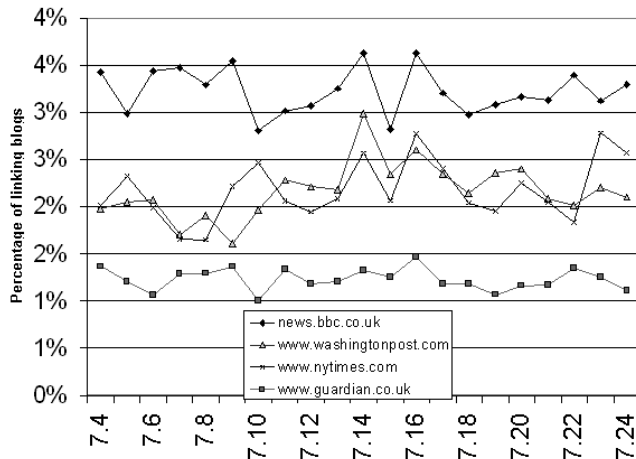


Figure 10. Top linked domains from all postings.

4.4 Top discussion topics

We analyzed the words that experienced the greatest increase in frequency during the lifetime of the corpus, assuming that the major discussion topics would be reflected by an increase in the usage of one or more relevant words [16].

For each word in each corpus we counted the percentage of blogs in which it occurred at least once on each day and then for each word we measured its (largest single day) spike, defined to be the maximum percentage increase in *blogs* (rather than posts) containing the word compared to the average percentage of all previous days. We then listed the words in decreasing order of spike size, creating a separate list for each corpus. Figures 11 and 12 present selected results from the top 10 spike words of each corpus. The results of counting per blog are much better than for counting per item or per word because of spam. For example all of the top 10 words in all three corpora were interesting with our method, but when counting on a per-posting basis and using otherwise the same technique, one of the top words was “eric”, the name of a blogger who suddenly started a bout of intensive posting. We did not want the statistics to be influenced unduly by individuals in this way.

Below are some of the top words, clustered by co-occurrence in at least 30% of blog postings at the time of the peak occurrence of the highest word. The first word is the highest spike word and the words in brackets are co-occurring words.

- **London posts corpus:** Olympic (bid), attack (terrorist, bombing), yesterday, police (shot), bomb, rule.
- **London posters corpus:** London (attack), terrorist, bombing, bomb, harry (potter), explosion (bus), Olympic.
- **Full corpus:** London (attack), harry (potter), terrorist, bombing, chocolate (factory), bottom, bomb.

Figure 11 shows the importance of the Olympics just before the London attacks. We presume that most of the London postings after the 7th were related to the attacks, but only a proportion would have mentioned the word attack (or ‘attacks’, we stemmed for plurals). The word police is interesting – the police force became increasingly central to the discussion of the events as the focus shifted to stopping further attacks or catching helpers of the bombers. Note also that the police themselves became the main story after the shooting of Jean Charles de Menezes in error. Our word list did not just include nouns and one of the top words was ‘yesterday’, which spiked very strongly on July 8 (not shown).

Figure 12 unsurprisingly shows that London was the most important term across the whole corpus, but it sits in uneasy combination with the release of a Harry Potter book and the film Charlie and the Chocolate Factory. It is not surprising that terrorist/terrorists was less popular than London: only a proportion of postings related to the attack would have discussed the perpetrators, and some may have used alternative words such as bombers or bombing. Note also the double spike for London, with one at the time of the failed second attacks.

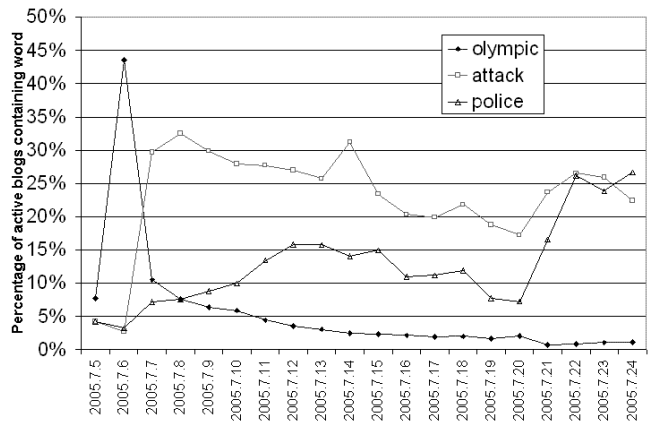


Figure 11. Top spike words from London postings (selection).

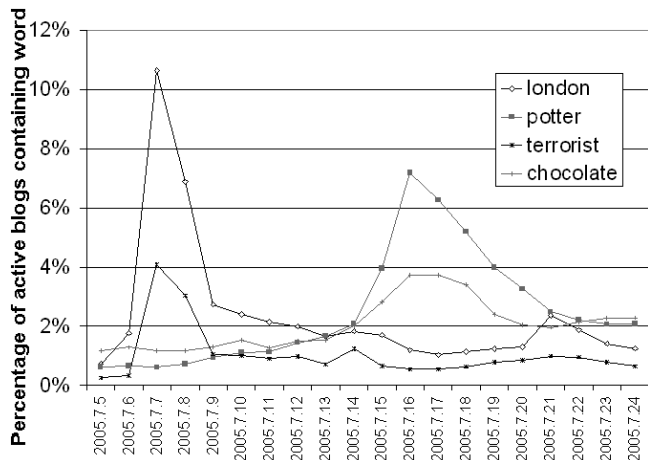


Figure 12. Top spike words from all postings (selection).

4.4.1 Comparison of top words to media timelines

We produced a list of the 1,600 words sharing the largest usage increase within the London postings. In order to compare these with media coverage of the events, we chose timelines as representatives of how the media summarize an event after the fact. We selected timelines from the BBC (<http://news.bbc.co.uk/1/hi/uk/4694069.stm>, accessed December 21, 2005), the London Telegraph (http://www.telegraph.co.uk/news/main.jhtml?xml=/news/exclusions/ufflondon_terror.xml, accessed December 21, 2005) and Wikinews (http://en.wikipedia.org/wiki/Timeline_of_the_2005_London_bombings, accessed February 12, 2006). All were different in format but all sought to give a retrospective summary of the attacks. This comparison has the potential to give insights into differences between the topics that were considered important by bloggers during the London attacks at the time, and those that were considered important by media sources and Wikinews in retrospect.

Figure 13 shows that the highest spike words in blogs tended to also be frequently mentioned in the timelines. Figure 14 confirms this but shows that some blog words were missing, including many related to the Olympics (Olympic, agree, host, game), although 'Paris' and 'bid' were included. The near absence of the Olympics is unsurprising as the decision to award the Olympics to London was quickly dismissed as a potential factor in the London bombings.

Other words in the top 50 spikes for blogs that were not used in any of the media timelines and are unrelated to the Olympics include: *suspend*; *motion*; *condemning*; *innocent*; *subway*; *Egypt*; *Iraq*; *prayer*. Some of these reflect a change in language: 'innocent' seems to be assumed in retrospect for both the bombing victims and the man shot by the police in London. Others seem to concern reactions to the events: prayers and motions condemning the attacks. The contemporary news also prominently featured an Egypt connection which turned out to be a dead end.

Perhaps most surprisingly, none of the media timelines explicitly mentioned a connection with Iraq, which was clearly important in the blogs. This may be a political point (at least for the BBC and The Telegraph) because the UK prime minister denied a connection between the invasion of Iraq and the bombings, although this was hotly disputed. In media theory terminology, the 'frame' [30] of the reporting appears to have shifted to avoid Iraq. Put simply, framing is the selection of facts about an event in order to promote a particular interpretation [31]. It is a standard device used by the media to create easily understandable news stories but the result can often be criticized for oversimplifying or ignoring complexity [32]. Framing is closely related to the semiotic concept of myth: text often invokes pre-existing sets of ideas and beliefs (called 'myths', whether true or not) which can add implicit meaning through intertextuality [33]. For example the mere use of the word Iraq in the current climate is likely to suggest war and perhaps also political judgments. Both framing and the invoking of myths are quite subtle ways in which the reporting of an event can be simultaneously factual and evoke a particular worldview or political perspective.

Except perhaps for the word Iraq, the word differences identified between blogs and timelines do not really provide a strong indication that blog discussions differ significantly from media reporting because some differences are statistically inevitable in

any word frequency analysis. This is perhaps surprising given that media reporting is known to be highly self-managed and politically sensitive for violent events, needing to carefully evoke emotions to attract an audience without alienating it [32]. The fault is probably with the level of analysis used here. A qualitative approach might have been more informative for this issue, for example a content analysis of a selection of London postings may have revealed a tendency for more partisan discussion of the events, as seemed to have been the case for high profile blogs discussing the Iraq war [34]. Alternatively, a quantitative comparison with a large collection of contemporary news stories may have given scope for statistical conclusions to be drawn about the difference between blogs and contemporary media coverage.

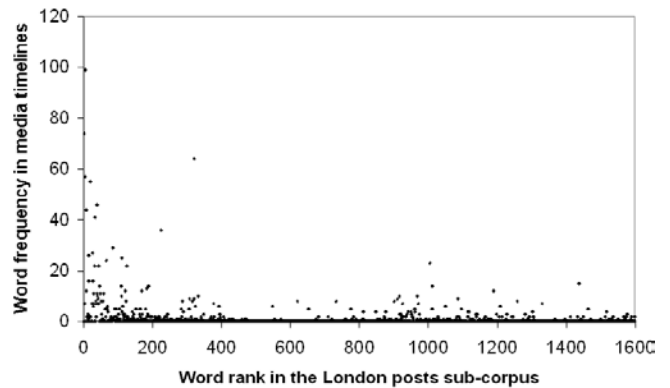


Figure 13. The top 1,600 spike words from London postings: frequency in three media timelines.

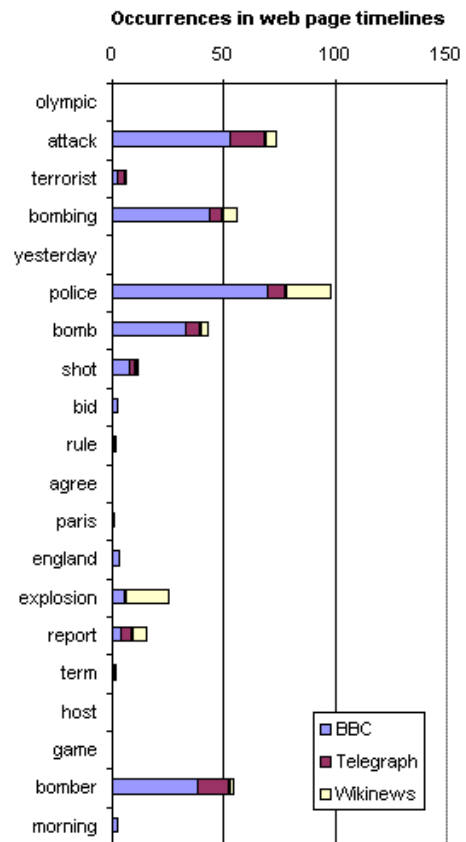


Figure 14. The top 20 spike words from London postings: frequency in three media timelines.

5. SUMMARY

We have deployed a range of methods for time series analysis of word frequencies and linking in blogspace, with key techniques being: the manual elimination of spam; *the aggregation of all postings by a single blogger on a single day for link counting and topic identification*; and the separation of the corpus by blog content and by posting content. The results without aggregation and spam elimination (not shown, but recall “eric” in Section 4.4) reflected marketing campaigns and individual bloggers rather than trends in blogspace and so we believe that our techniques are justified. Nevertheless, even after processing the data, all of our link graphs are quite jagged and hence it is not really possible to draw strong conclusions about the impact of the London attacks on linking or the top link targets in blogspace. The jagged nature of the graphs may be due to a combination of weekly periodic variations, technical issues with gathering and hosting blog data, and natural fluctuations in blogging unrelated to the London attacks.

At the start of this research, we had hoped to show that bloggers increased their external awareness as a result of the attacks, and that this was reflected in an increased tendency to create links. This is a possible interpretation of Figure 6, but the fluctuations in the graph render such a judgment unreliable. Nevertheless, we were able to reach some more reliable conclusions based upon splitting the corpus by blogger and posting type.

In terms of the web sites targeted by links from blogs, the appearance of major media sources is not a surprise, although the prominent position of a non-US source, the BBC, may be unexpected. One of the most interesting finding concerning the top link targets was the fall in the percentage share of the top linked-to sites at the time of the attacks (Figure 8), perhaps reflecting a widening of the search for information rather than a sudden reliance upon a few authoritative sources.

Despite a surprisingly small percentage of bloggers mentioning London ($73,161/1,417,711 = 5\%$), the attacks were the most significant event in the blogosphere, as shown by both the link and text analysis. Some high profile news and cultural events were also evident from this analysis, but were apparently less discussed than the attacks. Bloggers who posted about London were, on average, different to other bloggers because they posted more and linked more by a factor of about 4. This would be broadly consistent with a dichotomy between externally-focused, news-aware approximately daily bloggers and more internally-focused diary-like approximately weekly bloggers, although a wide spectrum of behavior is likely for both categories of blogger.

Finally, in comparison to major media representations of the London attacks, blog postings give insights into contemporary discussions and can highlight issues, such as the alleged Iraq connection, which may be forgotten or deliberately ignored in hindsight. This analysis has really only scratched the surface of the potential for blog analysis to give new insights into public debates and media coverage of major events.

6. ACKNOWLEDGMENTS

The work was supported by a European Union grant for activity code NEST-2003-Path-1. It is part of the CREEN project (Critical Events in Evolving Networks, contract 012684).

7. REFERENCES

- [1] Herring, S.C., et al., Conversations in the blogosphere: A social network analysis "from the bottom up", in Proceedings of the Thirty-eighth Hawaii International Conference on System Sciences (HICSS-38). 2005, Los Alamitos: IEEE Press.
- [2] Su, N.M., et al., A bosom buddy afar brings a distant land near: Are bloggers a global community? in Communities And Technologies 2005. 2005. p. 171-190.
- [3] Gance, N.S., M. Hurst, and T. Tomokiyo. BlogPulse: Automated trend discovery for weblogs. In WWW 2004 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics. 2004.
- [4] Gill, K.E. Blogging, RSS and the information landscape: A look at online news. In WWW 2005 2nd Annual Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics. 2005. Chiba, Japan.
- [5] Matheson, D., Weblogs and the epistemology of the news: Some trends in online journalism. *New Media & Society*, 2004. 6(4): p. 443-468.
- [6] Huffaker, D.A. and S.L. Calvert, Gender, identity, and language use in teenage blogs. *Journal of Computer-Mediated Communication*, 2005. 10(2): p. <http://jcmc.indiana.edu/vol10/issue2/huffaker.html>.
- [7] Nardi, B.A., et al., Why we blog. *Communications of the ACM*, 2004. 47(12): p. 41-46.
- [8] Gorgura, H. The war on the terror consensus: Anti-war blogs as an online sphere of dissensus. In *AoIR 5.0*. 2004. Brighton, UK.
- [9] Kim, J.H. Blog as an oppositional medium? A semantic network analysis on the Iraq war blogs. In *Internet Research 6.0: Internet Generations*. 2005. Chicago.
- [10] Bar-Ilan, J., Information hub blogs. *Journal of Information Science*, 2005. 31(4): p. 297-307.
- [11] Lin, J. and A. Halavais. Mapping the blogosphere in America. In *WWW 2004 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics*. 2004. New York.
- [12] Ackland, R., Mapping the U.S. political blogosphere: Are conservative bloggers more prominent? 2005, The Australian National University.
- [13] Cronin, B., Vox populi: Civility in the blogosphere. *International Journal of Information Management*, 2005. 25(6): p. 483-586.
- [14] Kumar, R., et al. On the bursty evolution of blogspace. in *WWW2003*. 2003. Budapest, Hungary, <http://www2003.org/cdrom/papers/refereed/p477/p477-kumar/p477-kumar.htm>.
- [15] Kumar, R., et al., Structure and evolution of blogspace. *Communications of the ACM*, 2004. 47(12): p. 35-39.
- [16] Gruhl, D., et al. Information diffusion through Blogspace. In *WWW2004*. 2004. New York, <http://www.www2004.org/proceedings/docs/1p491.pdf>.

- [17] Prabowo, R. and M. Thelwall, A comparison of feature selection methods for an evolving RSS feed corpus. *Information Processing & Management*, 2006, to appear.
- [18] Thelwall, M., R. Prabowo, and R. Fairclough, Are raw RSS feeds suitable for broad issue scanning? A science concern case study. *Journal of the American Society for Information Science and Technology*, 2006, to appear.
- [19] Polanco, X., et al., Clustering and mapping Web sites for displaying implicit associations and visualising networks. 2001, University of Patras.
- [20] Ingwersen, P., The calculation of Web Impact Factors. *Journal of Documentation*, 1998. 54(2): p. 236-243.
- [21] Björneborn, L. Small-world linkage and co-linkage. in *Proceedings of the 12th ACM Conference on Hypertext and Hypermedia*. 2001. Århus, Denmark: ACM Press.
- [22] Park, H.W., Hyperlink network analysis: A new method for the study of social structure on the web. *Connections*, 2003. 25(1): p. 49-61.
- [23] Björneborn, L. and P. Ingwersen, Toward a basic framework for webometrics. *Journal of the American Society for Information Science and Technology*, 2004. 55(14): p. 1216-1227.
- [24] Thelwall, M., Conceptualizing documentation on the Web: An evaluation of different heuristic-based models for counting links between university web sites. *Journal of American Society for Information Science and Technology*, 2002. 53(12): p. 995-1005.
- [25] Thelwall, M., *Link analysis: An information science approach*. 2004, San Diego: Academic Press.
- [26] Intelliseek, 3rd Annual Workshop on the Weblogging Ecosystem. 2005.
- [27] Vaughan, L. and M. Thelwall, Search engine coverage bias: evidence and possible causes. *Information Processing & Management*, 2004. 40(4): p. 693-707.
- [28] Pennock, D., et al., Winners don't take all: Characterizing the competition for links on the web. *Proceedings of the National Academy of Sciences*, 2002. 99: p. 5207-5211.
- [29] Adamic, L.A. and B.A. Huberman, Power-law distribution of the World Wide Web. 2000.
- [30] Entman, R. M., Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 1993. 43(4): p. 51-58.
- [31] Gamson, W.A. & Modigliani, A. Media discourse and public opinion on nuclear power: A constructionist approach. *American Journal of Sociology*, 1989. 95(1), 1-37.
- [32] Seaton, J., *Carnage and the media: The making and breaking of news about violence*, 2005, London: Allen Lane.
- [33] Fiske, J. *Introduction to communication studies (2ed.)*. 1991, London: Routledge.
- [34] Thompson, G. Weblogs, warblogs, the public sphere, and bubbles. *Transactions*, 2003. 7(2).
http://transformations.cqu.edu.au/journal/issue_07/article_02.shtml