

Australia's Political Blogosphere in the Aftermath of 2007 Federal Election

Dr Axel Bruns, Dr Jason Wilson, Barry Saunders
Creative Industries Faculty, Queensland University of Technology, Brisbane, Australia
a.bruns@qut.edu.au, j5.wilson@qut.edu.au, b.saunders@qut.edu.au

Lars Kirchhoff, Thomas Nicolai
Institute for Media and Communication Management, University of St. Gallen, Switzerland
lars.kirchhoff@unisg.ch, thomas.nicolai@unisg.ch

Australian political bloggers and citizen journalists appear to have played an important role in the campaign leading up to the 24 November 2007 federal election. They provided a highly critical corrective to mainstream journalism, seemed to influence public opinion on key election themes, and offered a coverage of political events which diverted from the customary focus on political leaders and bellwether locations only. Bloggers were wooed by political parties (such as the Australian Labor Party with its *Labor First* blog site), mainstream media (such as the online arm of public broadcaster ABC, which ran several blogs of its own), and journalism researchers (through projects such as *Youdecide2007.org*, which provided a space for a hyperlocal citizen journalism coverage of the campaign in participants' individual electorates¹).

But what remains unclear to date is exactly how information travels within the distributed network of the blogosphere itself, and from here to other (online) spaces of citizen and industrial journalism. To trace such movements may underline (or undermine) news and political bloggers' claims of influence and importance; it would highlight the extent to which blogging operates merely as an echo chamber for the political cognoscenti, or has impact in the wider population. It would provide insight into the extent to which news bloggers and mainstream journalists feed off and respond to one another's work, and outline possible avenues for mutually beneficial collaborations.

The main problem with tracking interactions in the blogosphere in a quantitative fashion is two-fold: on the one hand, scholars have yet to agree what metrics may be appropriate for measuring bloggers' actions and their influence on the wider political sphere (and indeed, what appropriate definitions for terms such 'influence' may be). On the other, even if an appropriate working model of what those metrics may be is agreed upon, the question of how to document them in a reliable, testable fashion across a potentially large population of sites also remains non-trivial.

This paper examines the use of automated tools to address such questions in the context of tracing patterns of interlinkage and clusters of interconnection in the Australian political blogosphere. While this approach is not without its own limitations (as noted in Bruns, 2007), early results have nonetheless proven instructive – network crawls on a number of current issues, for example, indicate the existence of a relatively stable structure in the Australian political blogosphere which involves a number of fairly clearly defined hubs (and leading sites within these hubs), most likely formed largely on the basis of shared ideology.

But there are significant methodological problems with such research, too – most of all, current tools remain fairly blunt in their analysis specifically of blog-based content. Link crawling and network mapping tools are generally unable to distinguish between the different forms of content that may be found on a blog: they simply crawl the entire page, or the entire Website, which means that navigational and other functional links on a page are included in a common category with links in blogrolls, posts, and comments (and even, in the case of some sites, ads) – and this takes place possibly for the site overall, combining recent and archival links. When these data are mapped, then, such maps represent a blurred, composite picture

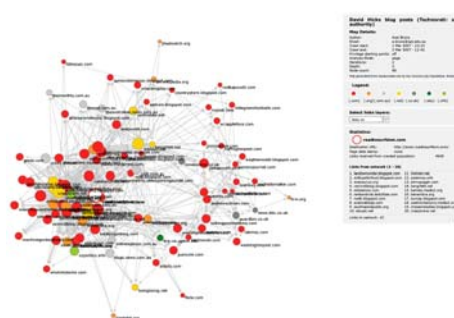


Fig. 1: Network of Bloggers in the Australian Political Blogosphere (from Bruns, 2007)

¹ Axel Bruns, Jason Wilson, and Barry Saunders are part of the research team for this project, alongside Terry Flew and Stuart Cunningham (all QUT).

only. Similar limitations apply for textual analysis-based automated data gathering and evaluation approaches, of course – by default, an automated textual analysis of a Website using a tool such as Leximancer will take in actual blog posts themselves as well as static pages, functional information, and other material.

Research Framework

While qualitative evidence for the networked patterns of discussion, debate, and deliberation in the blogosphere is readily available, it is more difficult to establish a solid quantitative picture of blog-based topical discussion networks and their cluster patterns, then. Large numbers of blogs (and individual blog posts, links, and comments) are likely to be involved in a quantitative study of blog-based discussion patterns. Hence, automated data collection and analysis is necessary. Any tools used for this purpose need to be able to distinguish between the different units of analysis: in terms of content, the blog posts themselves, blog comments, blogrolls, and ancillary (static) content; in terms of links, topical links in blog posts, commenter-provided links, blogroll links, and generic links elsewhere on the site.

Distinctions between these different categories build on the following assumptions:

1. Content:

The core underlying assumption is that *the vast majority of bloggers write about topics which interest them* (rather than claiming an interest they do not have). This should not be understood to claim that bloggers cover *all* the topics they are interested in, however – the topics covered on any one blog constitute merely that subset of all interests which a blogger has deemed it acceptable to reveal publicly to a general readership. On this basis, we assume that:

- a. The complete collection of all blog posts for a given blog provides a reliable indication of the interests of the individual blogger (as expressed publicly); the development of these interests may be further traced by tracking changes in topical coverage over time.
- b. A comparison of bloggers' interests (in total, for specific periods of time, and/or in relation to broad topical domains) across multiple blogs indicates the distribution of topical interest across the blogosphere (at least for the subset of the entire blogosphere included in the analysis).
- c. A comparison between the blogger's postings on specific topics, and the collection of reader comments to these postings, indicates the level of agreement or disagreement between blogger and commenters (at least for blogs with substantial commenting activity).

2. Links:

The core underlying assumption is that *links to other Websites indicate a recognition of the linked content as 'interesting'* (for a variety of possible reasons, and potentially indicating approval or disapproval). By extension, *this also confers a certain amount of reputation and attention on the creator of the linked content* (again, this accrued reputation can be either positive or negative).

We also assume that linking patterns predict traffic and influence. The more incoming links any piece of content has, the more likely visitors are to see it, and this increases its potential to influence readers. Further, the outgoing links of sites which themselves receive many incoming links are more powerful in directing traffic and conferring influence than the outgoing links of little-known sites. Google's PageRank and Technorati's authority ranking operate on similar assumptions.

On this basis, we assume that *patterns of interlinkage indicate the existence of a network of attention. These patterns are indicators of visibility and influence. In these patterns, the balance of incoming and outgoing links for any one site or page warrants special attention.* Specifically,

- a. Patterns of interlinkage between contemporaneous blogrolls indicate the existence of a long-term network of recognition between peers. Sites with many incoming *and* outgoing links may be understood as *hubs* for communication in this network; sites with many incoming, but limited outgoing

links may be understood as central *sources* for information; sites with many outgoing but few incoming links may be understood as (not necessarily central) *distributors* of attention to other members of the network. The gradual evolution of such networks can be traced over time.

- b. Patterns of interlinkage between contemporaneous blog posts (and other post-level content) indicate the existence of a network of debate on specific topics. Such networks of debate can be seen to persist over greater or lesser periods of time. Posts with many incoming links may be understood as making an important (possibly controversial) *original* contribution to the debate; posts with many incoming and outgoing links may be understood as making an important *discursive* contribution to the debate; posts with many outgoing links may be understood as *introductions to* or *summaries of* ongoing debate.
- c. Aggregated from the level of the blog post to that of the blog, these patterns of interlinkage also indicate the role of the overall blogs in topical debate networks. Blogs with many incoming *and* outgoing links may be understood as central *hubs* for communication on this topic; blogs with many incoming, but limited outgoing links may be understood as central *sources* for information on the topic; blogs with many outgoing but few incoming links may be understood as (not necessarily central) *distributors* of attention to other members of the network. A comparison of these short-term debate networks over time and across topics indicates the fluctuation of centrality; sites whose centrality remains high over time can be seen as having significant authority overall, while sites whose centrality is high only for specific topics can be seen as having significant authority only for those topics.
- d. Patterns of interlinkage between blog posts and comments indicate that posts or comments have an ongoing relevance to particular networked debates. If a comment is linked to in a further post (either on the blog on which the comment was posted, or elsewhere), it indicates that the comment has itself provoked further discussion and commentary, and that the conversation constitutes a dialogue between blogosphere authors and commenters. If blog posts are referred to in comments threads, especially if these are on other blogs, it indicates that the initial post has relevance and influence in an ongoing, networked debate.
- e. Patterns of linkage between current and archived posts on the same blog indicate the blog author's continuing interest in and coverage of relevant topics.

In order to conduct a quantitative analysis of blog-based discussion networks at a content and link level on the basis of this research framework, a number of tools must be used. Each introduces a number of necessary limitations to the breadth and depth of study possible. The three key elements of the research process are data gathering and processing, content analysis, and link network analysis (however, this does not imply that content analysis necessarily precedes network analysis, or vice versa). Subsequently, it is also possible to extract and identify common patterns and interrelations between content and network analyses. Additional qualitative and/or quantitative work beyond these initial stages could extend into social network analysis, to identify social networks within the blogosphere.

The present project addressed the key stages in this process in the following way:

- development of a master list of some 300-400 Australian blogs with a (broadly) political theme, based on blogs identified in preliminary issue crawling exercises such as those documented in Bruns (2007), augmented and extended on a continual basis;
- subscription to the RSS feeds for these blogs in order to identify new posts on these blogs as they become available;
- scraping and storing of blog content for each new blog post identified from RSS feeds;
- processing of scraped content, extracting:
 - o blog entry text only (discarding static content and comments) and

- o discursive links (discarding static, blogroll, and internal navigational links);
- automated analysis of blog entry contents (for specific blogs, across blogs, and for specific timeframes), using the textual analysis software Leximancer;
- automated analysis of discursive links, using a combination of tools including VOSON, Pajek, and UCINET.

Later stages of the project will extend this analysis by examining blogroll and comments contents in comparison to blog post contents.

Analysis

This paper presents initial results especially from the textual analysis component of the overall research project, focussing on the performance of three high-profile Australian political blogs during November 2007, January 2008, and February 2008. In doing so, it both presents these results themselves as well as serving as a proof-of-concept for the overall research project. Responses to these early results will be incorporated into later stages of the project.

The textual analysis conducted here builds on the Leximancer software. Leximancer begins by simply trawling through the body of text and counting how often each word occurs, and in doing so drops a number of meaningless 'kill words' (the, and, or, etc.); beyond this, it allows for further manual editing as well. In this second stage, we combined terms which have the same meaning (in our sample, for example ALP, Labor; Liberals, Libs, Liberal Party; Howard, John Howard) and removed further terms which are irrelevant for the intended analysis (such as finally, apparently, significantly, think, show, say, etc.). Choices made during this process are necessarily skewing the eventual result, of course. Perhaps the most contentious omission we made in the analysis discussed below is to drop the "Australia, Australian, Australians" cluster of terms from the analysis – our reasoning here was a) that in discussing Australian politics these terms will occur so often and in so many different contexts (Australian government, Australian elections, Australian media, Australia's policy, etc.) to be essentially meaningless, and b) that it is impossible for Leximancer to distinguish between mentions of the Australian election/media/... and *The Australian*, the national daily newspaper which would itself have featured prominently in the data (also because 'the' is itself a 'kill word', of course). Clearly this is not without its problems, however – at a later stage, we may conduct a number of comparative analyses to see whether leaving Australia(n/ns) in the sample makes a significant difference to the results. (Combining terms may also obscure differences in tone and register – any analysis focussing on such aspects may need to avoid term combinations, and possibly refrain from excluding 'kill words' altogether.)

In addition to this cluster of terms, a number of the other terms originally identified by Leximancer in its first pass over the data were also dropped. The full list of such dropped terms is different for each of the blogs analysed here, but usually includes terms such as the names of months and days, link words such as 'finally, first', etc., and terms occurring in frequent phrases ('instance' from 'for instance', 'writes', 'says', etc. from '[journalist/blogger] writes', and similar terms). We hope to fine-tune this list, and perhaps develop a list of standard 'kill words' that applies more generally for blogging analysis in the Australian political context.

The blogs examined for the purposes of this analysis included Andrew Landeryou's *The Other Cheek*, *Larvatus Prodeo*, and *Club Troppo*. We begin by providing a frequency analysis for the most common terms in each of these blogs (listing the top 20 most commonly used terms for each blog), where some notable differences in the key terms used in each of these blogs are clearly apparent. *The Other Cheek*, for example, contains many of what may be called the generic technical terms of politics – the names of major parties, 'politics/al', 'election', 'government', 'campaign', 'seat', candidate', etc. Owing to Landeryou's writing style, 'OC' itself also pops up as a very frequent term, as does 'patriot' as a term of endearment for those on either side of Australian politics with whose political views he agrees. (Further, 'Game' at #19 is reflective of another stylistic quirk: Landeryou's frequent use of the phrase 'Game on'.) On the other side of the political ledger, 'Greens' is also a relatively frequent term, but beyond this there is a notable absence of terms related to specific policies and initiatives – except perhaps for 'work' (which could relate to workplace relations), but then 'work' is also commonly used in non-issue-specific phrases ('this policy won't work', 'our agenda of work', etc.), so that its ranking among the top 20 terms here might be misleading. It is somewhat

surprising that outgoing Prime Minister John Howard (who was defeated in the November election) is not mentioned particularly often (he appeared only at #36 in the list, with 31 mentions) – but more than two thirds of the corpus used for this analysis stem from after the election; what this may mean, therefore, is that Landeryou is strongly focussed on *current* events and scandals, not so much on longer-term analysis.

The Other Cheek			Larvatus Prodeo			Club Troppo		
Concept	Absolute Count	Relative Count	Concept	Absolute Count	Relative Count	Concept	Absolute Count	Relative Count
Labor	145	100%	Labor	127	100%	national	263	100%
Liberal	145	100%	Howard	102	80.3%	government	241	91.6%
OC	143	98.6%	government	102	80.3%	years	228	86.6%
political	104	71.7%	election	91	71.6%	economic	204	77.5%
election	87	60%	party	80	62.9%	policy	202	76.8%
government	83	57.2%	Rudd	79	62.2%	people	194	73.7%
campaign	76	52.4%	year	78	61.4%	Howard	154	58.5%
people	65	44.8%	political	76	59.8%	election	133	50.5%
patriot	62	42.7%	campaign	66	51.9%	Labor	124	47.1%
seat	62	42.7%	policy	54	42.5%	world	122	46.3%
Rudd	56	38.6%	issue	53	41.7%	countries	114	43.3%
candidate	56	38.6%	Coalition	48	37.7%	politics	109	41.4%
state	50	34.4%	work	42	33%	problem	105	39.9%
Greens	49	33.7%	uranium	41	32.2%	public	91	34.6%
public	48	33.1%	public	41	32.2%	issue	91	34.6%
federal	42	28.9%	change	39	30.7%	work	90	34.2%
left	42	28.9%	seats	39	30.7%	Rudd	86	32.6%
work	38	26.2%	blog	38	29.9%	change	84	31.9%
Game	38	26.2%	candidates	37	29.1%	tax	84	31.9%
Age	38	26.2%	person	37	29.1%	price	78	29.6%

Table 1: Concept count for the three blogs

This is one clear point of difference with *Larvatus Prodeo*; here, over the same circum-election period, Howard ranked at #2 and was mentioned 102 times. Otherwise, the broad trend is perhaps similar to *OC*; generic technical terms such as ‘government’, ‘election’, ‘party’, ‘political’, ‘campaign’ are similarly strong, but ‘policy’ and ‘issue’ also get a showing. One notable term here is ‘uranium’ at #14 (41 mentions), which clearly points to the presence of more specific topical debates in addition to the broader coverage of electioneering and political processes. Note also ‘blog’ as a key term – in part, this may reflect the fact that *LP* has been a keen follower of the ‘polling wars’ conducted between mainstream commentators and political and psephological bloggers over how to correctly interpret pre- and post-election opinion polling (see Bruns, Wilson, and Saunders 2008), and that questions related to the role of citizen journalism in Australia are not uncommon on this blog. However, like ‘work’, its ranking might be inflated by more generic uses (‘as X writes in their blog’). By comparison, ‘blog’ appears only somewhat further down the order of terms for Landeryou, with 21 mentions.

The picture is considerably different for *Club Troppo*, where terms related to specific policy fields are notably more common. Here, even terms such as ‘Howard’, ‘Labor’, and (incoming Prime Minister) ‘Rudd’ are outranked by ‘economy/ic’ and ‘policy’, and ‘world’, ‘countries’ (often in the context of ‘developed/ing countries’), ‘problem’, ‘issue’, and ‘tax’ also make the top twenty list. Many more of these – ‘rates’, ‘recession’, ‘market’, ‘child/ren’, ‘international’, ‘community’ – occur on the next twenty places in the list, unlike *The Other Cheek* and, less strongly so, also unlike *Larvatus Prodeo*. To highlight just one further example for these differences: the term ‘intervention’ (used mainly in relation to the Howard government’s intervention in Northern Territory indigenous communities which aimed to safeguard indigenous children’s health and protect them from sexual and drug abuse) ranks at #57 for *Troppo*, at #63 for *LP*, and does not appear in *OC*’s list of key terms at all. Related to this, ‘child/ren’ ranks at #32 for *Club Troppo*, but does not appear in the list of key concepts for either of the other blogs. One reading of this suggests that *Troppo* focusses much more strongly on policy analysis over political wonkery and insider gossip; for *OC*, the balance is reversed, while *LP* sits somewhere in the middle between these extremes.

A second step in the analysis maps these key terms in relation to one another – terms which frequently co-occur in close proximity to one another in the text are located closer to one another on this map than terms which do not, in other words. The resulting maps provide further support to the observation that the blogs have different points of focus in their day-to-day coverage of politics – and by plotting all frequently-used terms on the map, the exact nature of these topical clusters becomes more easily discernable.

Figure 1: Topic map for *The Other Cheek*

For *The Other Cheek*, a number of interesting clusters emerge – most closely connected is probably the cluster of terms in the bottom left quadrant of fig. 1, which contains terms such as ‘bizarre’, ‘corrupt’, ‘courageous’, ‘friends’, ‘power’, ‘serious’, ‘failed’, ‘claims’, etc. This can be said to represent the ‘political gossip’ element in OC. A broader, looser cluster exists in the bottom centre – terms such as ‘love’, ‘leftard’, ‘electorate’, ‘government’, ‘office’, ‘policy’, ‘media’, etc. can be found here, and can be understood as pertaining to politics more generally. Above and to the right of this is a more close-knit cluster of terms around ‘election’ ‘minister’, ‘campaign’, ‘excellent’, ‘supporters’, ‘successful’, and ‘leader’, and overall top terms ‘Labor’ and ‘Liberal’ are also found here – clearly this cluster is related directly to the election and its immediate aftermath.

Further above this are a few smaller clusters of terms: 'Radio National', 'climate change', 'Philip Adams', 'Greens', 'ABC', 'radio', and the more distant outlier 'Climate Change Coalition' indicate Landeryou's continuing line of commentary not necessarily on the science of climate change itself, but on reporting (especially by the national public broadcaster ABC) on climate change issues; to the left of this, a small cluster around 'Stephen Mayne', 'Nameless' (a Landeryou code term for independent political commentary site *Crikey*, founded

may have discussed New Zealand economic policy in some detail). In the top left, a focus on global politics becomes visible: 'developed', 'developing', 'India', 'China', 'countries', and further right also 'free', 'poor', 'global', 'human', 'development', and 'US' all seem related – as, presumably, is (Friedrich von) Hayek, the Austro-British economist.

In the bottom right, finally, is a looser smattering of terms which seem more directly related to the coverage of the electoral race: 'Labor', 'Liberals', 'leader', 'Howard', 'Coalition', 'campaign', 'election', and also 'Senate' and 'Andrew Bartlett'; additionally, *LP* founder Mark Bahnisch also makes an appearance here, most likely as a result of *Tropo*'s repeated linking to Bahnisch's posts about the electoral and post-electoral race in *Larvatus Prodeo* and elsewhere. It appears evident, though, that campaign and post-campaign politicking does not constitute the core of *Tropo*'s interests, and that its focus instead lies with the policies themselves.

Conclusion

This first exploration of our data indicates the validity of our research approach. We chose the three blogs examined here for the relatively clear thematic and stylistic differences between them, and in both the frequency data described in the first part of analysis, and in the maps presented in the second part, some clear differences between the three blogs become evident.

This form of automated textual analysis enables researchers, therefore, to come to grips on a quantitative basis with the orientation of Australian political blogs – it becomes possible to measure and distinguish their thematic direction overall, and to identify how this might change over time as the political cycle proceeds: a similar analysis several months from now would likely produce a very different set of specific key terms, even though the general orientation of the blogs (in the present cases for example towards political gossip, wonkish play-by-play analysis of political tactics, or policy analysis) may remain consistent. This enables us move away from otherwise almost inevitable shortcut descriptions of specific blogs as 'left', 'centre', and 'right', and to develop a more sophisticated typology of blogs and blogging styles.

An obvious further step is to repeat this concept analysis process for the clusters of highly interlinked blogs that emerge from a hyperlink network analysis. Here, we may find that that blogs which link frequently to one another display a compatible thematic focus (gossip blogs link to gossip blogs, policy blogs link to policy blogs, etc.), but at the same time, the presence, for example, of *LP* founder Mark Bahnisch's name as a commonly used term in *Tropo* may also indicate that blogs specialising in one approach (say, policy analysis), link frequently to blogs which operate in a different, but complementary field (say, electoral coverage). Additionally, of course, a closer, qualitative look at some of the thematic clusters we have identified here will also produce further insight. What exactly is covered, for example, by the 'fiscal' or 'global politics' clusters on *Tropo*; what is the relationship between Stephen Mayne and Family First on *The Other Cheek*?

Finally, over the longer term it will be possible to trace the ebb and flow of topics, blogs, and blog clusters in the Australian political blogosphere, and to correlate this with the changing fortunes of political parties and the attendant shifts in political and policy focus over time. This longitudinal study is well beyond the scope of the present paper, but we believe that our current research contributes to the groundwork for this further research.

References

- Bruns, Axel. "Methodologies for Mapping the Political Blogosphere: An Exploration Using the IssueCrawler Research Tool." *First Monday* 12.5 (May 2007). 7 Jan. 2008 <<http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1834/1718>>.
- , Jason Wilson, and Barry Saunders. "When Audiences Attack: Lessons from the Australian Poll Wars." Leeds: Centre for Digital Citizenship, forthcoming in 2008. 19 July 2008 <[http://snurb.info/files/When%20Audiences%20Attack%20\(preprint\).pdf](http://snurb.info/files/When%20Audiences%20Attack%20(preprint).pdf)>.
- Club Tropo*. 2008. 19 July 2008 <<http://clubtropo.com.au/>>.
- Larvatus Prodeo*. 2008. 19 July 2008 <<http://larvatusprodeo.net/>>.
- Leximancer. 2008. 29 Apr. 2008 <<http://www.leximancer.com/>>.
- The Other Cheek*. 2008. 19 July 2008 <<http://andrewlanderyou.blogspot.com/>>.