Mapping the Australian Networked Public Sphere
Axel Bruns, Jean Burgess, Tim Highfield, Lars Kirchhoff and Thomas Nicolai

Social Science Computer Review published online 21 September 2010
DOI: 10.1177/0894439310382507

The online version of this article can be found at:
http://ssc.sagepub.com/content/early/2010/09/16/0894439310382507

Published by:
http://www.sagepublications.com

Additional services and information for Social Science Computer Review can be found at:

Email Alerts: http://ssc.sagepub.com/cgi/alerts

Subscriptions: http://ssc.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav
Mapping the Australian Networked Public Sphere

Axel Bruns\(^1\), Jean Burgess\(^1\), Tim Highfield\(^1\), Lars Kirchhoff\(^2\), and Thomas Nicolai\(^2\)

Abstract
This article reports on a research program that has developed new methodologies for mapping the Australian blogosphere and tracking how information is disseminated across it. The authors improve on conventional web crawling methodologies in a number of significant ways: First, the authors track blogging activity as it occurs, by scraping new blog posts when such posts are announced through Really Simple Syndication (RSS) feeds. Second, the authors use custom-made tools that distinguish between the different types of content and thus allow us to analyze only the salient discursive content provided by bloggers. Finally, the authors are able to examine these better quality data using both link network mapping and textual analysis tools, to produce both cumulative longer term maps of interlinkages and themes, and specific shorter term snapshots of current activity that indicate current clusters of heavy interlinkage and highlight their key themes. In this article, the authors discuss findings from a yearlong observation of the Australian political blogosphere, suggesting that Australian political bloggers consistently address current affairs, but interpret them differently from mainstream news outlets. The article also discusses the next stage of the project, which extends this approach to an examination of other social networks used by Australians, including Twitter, YouTube, and Flickr. This adaptation of our methodology moves away from narrow models of political communication, and toward an investigation of everyday and popular communication, providing a more inclusive and detailed picture of the Australian networked public sphere.

Keywords
network, mapping, public sphere, Australia, blogs, social media, methodology

Introduction
The rise of social media during the past decade, and of the Web 2.0 technologies that enable them, has been accompanied by a growing interest in examining the patterns of interaction and content creation in which social media users engage. Initial interest focused especially on the first major social media form, blogging, while more recent work has also examined activities in content sharing

\(^1\) Queensland University of Technology, Brisbane, Australia
\(^2\) Sociomantic Labs, Berlin, Germany

Corresponding Author:
Axel Bruns, Z1-515 Creative Industries Precinct, Queensland University of Technology, Kelvin Grove, QLD 4059, Australia
Email: A.Bruns@qut.edu.au
and collaboration sites such as Flickr and YouTube as well as social networking in Facebook, MySpace, and—most recently—Twitter.

We have also seen the development of increasingly sophisticated tools for crawling, tracking, and capturing social media content and interactions across and within specific sites, and for exploring, analyzing, and visualizing the resultant data. At the same time, there remain significant limitations to the capacity of these tools to pinpoint specific subsets of information and trends within the data, and a detailed discussion about the assumptions underlying some of the analyses and visualizations of these data, as well as about the ethical considerations that should guide data capture, storage, and analysis is yet to be had.

This article addresses key methodological questions from the perspective of a research program that has developed new approaches to mapping the Australian political blogosphere (Bruns, Highfield, Kirchhoff, & Nicolai, 2009; Bruns, Wilson, Saunders, Highfield et al., 2008; Bruns, Wilson, Saunders, Kirchhoff, & Nicolai, 2008; Kirchhoff, Nicolai, Bruns, & Highfield, 2009). We address a number of the shortcomings in conventional web crawling methodologies, outline the research approaches we have taken, and discuss findings from a yearlong observation of the Australian political blogosphere.

The article then looks ahead to the next stage of the research program, a $400,000 ARC Discovery project running from 2010 to 2012, which extends this approach to an examination of other social networks used by Australians, including Twitter, YouTube, and Flickr. This adaptation of our methodology moves away from narrow models of political communication, and toward an investigation of everyday and popular communication, providing a more inclusive and detailed picture of the Australian networked public sphere. In doing so, we seek to develop new alliances between traditional humanities approaches and large-scale computer-assisted analysis to investigate the social and cultural impacts of social media beyond the political blogosphere. In discussing this broader context for our work, in the latter sections of this article, we draw on theoretical developments in cultural and media studies to explore the relationship of everyday life with media change; and in particular, post-Habermasian theories of mediated publics.

Existing Work on Blog Mapping: Advances and Limitations

A number of themes are common in studies of blogospheres through large-scale data gathering methods. Much of the research that has been undertaken in this field over the past 5 or 10 years has focused on developing more or less comprehensive network maps of specific blogospheres—from the Adamic and Glance (2005) study of the partisan U.S. political blogosphere during the 2004 presidential elections to the map of political, social, and cultural clusters in the Iranian blogosphere by Kelly and Etling (2008), from the network of interlinkages between Australian political bloggers discussing the fate of Guantanamo detainee David Hicks (Bruns, 2007) to the map showing the interweaving of political blogospheres in various European countries around the appointment of the first President of the European Council in November 2009 (linkfluence, 2009).

Such mapping projects have produced much valuable information and have generated many impressive visualizations of interconnections and clustering patterns that appear to exist in the various blogospheres that they have examined. However, it is also important to query some of the assumptions upon which such studies has been based and to test the quality of the data sets with which they have worked. On closer inspection, a number of key limitations immediately become visible.

First, many network mapping studies rely on using network crawling tools for their data gathering. Such tools operate by starting with a seed sample of known blogs selected by the researchers and follow the hyperlinks present on these sites to find what sites the seed sample has linked to; they repeat this action a specified number of times to identify the network of interlinkages between all
the sites found in this process. However, this approach ignores the fact that not all of the links found on a given site are equivalent: some such links may merely assist with internal navigation on the site; some may link to advertising, institutional imprints, information about the content management system used, content licenses, or other background information; some links constitute the blogroll, a static feature used by bloggers to express mutual support and admiration, for example.

Only a subset of all links found on any given blog page actually form part of the blog post itself—and it is these links only which should be of immediate interest to most network mapping projects, as they alone form part of a discursive network of blog posts. It may be interesting to contrast this discursive network with the network constituted by blogroll links—to examine, for example, whether those blogs identified through the blogroll, as fellow travelers are also those which the blogger links to most frequently in the course of discussing issues of interest, but given the fact that blogrolls are usually relatively static and infrequently updated, and are absent altogether from many blogs, the study of the networks constituted by blogroll links is usually secondary to the study of discursive link networks. Worse, projects that do not distinguish between blogroll and discursive links (or which even simply include all links in their network maps) run the risk of finding false positives: they might find patterns of frequent interlinkage between two blogs that happen to include one another on their respective blogrolls but never actually respond to one another’s posts.

A second related problem for many network mapping studies is that most link crawlers do little more than following the links, wherever they may lead: few tools provide mechanisms for imposing temporal or geographic restrictions on the data set that is generated through crawling, for example. Even after only a few iterations, a basic crawl is likely to discover sites that have only very little to do with the initial focus of the seed sample of sites. The limited quality of the data set thus generated also limits its utility for network analysis: it would be valuable, for example, to compare the network of blog posts made on a specific issue during 1 month with that from the previous month, so as to examine the impact of new information coming to hand during that time and to track the evolution of blog-based discussion on the issue.

Finally, any attempt at an explanation of the interlinkage patterns identified in the process must necessarily also examine the content into which these links were embedded. Purely crawler-based research approaches falter at this hurdle—and many of the existing studies, for example, of political blog networks were forced to resort to manual coding of the political orientation of each site to reach their aims. But even where crawlers capture the blog pages they visit, researchers frequently encounter problems similar to those outlined above for link crawling: usually, it is not the entire blog page that is relevant for content analysis but only the blog post itself. Additionally, of course, the potentially vast corpus of textual data that may be gathered by large-scale crawls itself poses a significant problem, too.

Our Approach

To address such issues, our project, which focuses on tracking and mapping activity in the Australian blogosphere, uses crawlers only in a first, exploratory phase: we crawl the network of links from a seed sample of known Australian blogs to generate a more comprehensive list of blogs; this list is further refined through manual coding to ensure that comprises only Australian-based blogs (or, in a pilot study conducted during 2007–2009, Australian political blogs—see Bruns et al., 2009, 2008a/b; Kirchhoff et al., 2009). Any previously unknown Australian blogs identified in further stages of our research will also be added to this master list as they are discovered. Although this approach does not guarantee that literally all extant and active Australian blogs are included in the list, those blogs still left undiscovered are likely to be the least active, least linked to, and thus also least significant for the purposes of our study.
The most significant innovation in our research, however, is that subsequent to this early exploratory phase we no longer employ web crawlers for our data gathering, but instead use automated content harvesting processes that capture any new posts made by any of the blogs on our list. We identify and subscribe to the Really Simple Syndication (RSS) feed of each of the blogs we track and use a web scraper to harvest the entire blog page whenever an RSS feed alerts us to a new post. In a further step, we process the content of the captured page, separating the content of the blog post itself from comments, blogroll, headers, footers, sidebars, and other ancillary materials that may also be present on the page. We also extract any links that are present in the blog post itself; only these are used at a later stage to map the blog network.

The resultant data set thus contains the blog posts themselves (stripped of any extraneous content surrounding those posts on the blog page), alongside any links that were present in the posts; each post is also timestamped, enabling us to slice our data set to focus only on blog posts made during specific time frames. As a result, importantly, it is possible for us to create and compare maps of the discursive interlinkage patterns between Australian blogs during specific days, weeks, or months—and thus to track how clusters of interlinkages in the Australian blogosphere may form or dissolve over time, to identify how persistent or ad hoc such clustering patterns are. Such evolving patterns of interconnection may also be correlated with broader current events and underlying societal trends, of course.

Furthermore, by capturing entire posts (stripped of extraneous material), we are able to engage in large-scale textual analysis, for example, using automated concept mapping tools to extract the most frequently used keywords from the corpus of post data and map their patterns of co-occurrence. Our timestamping of blog posts allows us to examine and compare such dominant themes for specific time periods, for example, to correlate themes in the blogosphere with topics in the mainstream news or to measure and track the emergence and decline of trending topics from day to day and week to week.

Finally, textual analysis is combined with link network mapping: for example, to examine the key themes of discussion within the various clusters of frequently interlinked blogs (and thus perhaps to explain these clusters as communities of shared interest) or to investigate whether blogs that are characterized by similar topical interests also link to a comparable set of online resources. Obviously, such quantitative explorations can—and should—also be extended through further qualitative examination of the blogs concerned.

**Our Findings So Far**

A preliminary phase of our study, conducted between 2007 and 2009, focused on political blogs as a specific subset of the Australian blogosphere. We identified several hundreds of Australian blogs with a relatively strong focus on politics and tracked their posting and linking activities since November 2007—roughly 3 weeks before the Australian federal election that delivered a change of government. For comparison, our sample also included a number of mainstream news websites and professional columnists in a separate category. After further finetuning, we restarted our content capture processes at the start of 2009.

In the first place, our results from this preliminary work show a typical long tail-style distribution of activity within the Australian political blogosphere (which at this point, we also take to include professional columnist/bloggers at mainstream news sites and a small number of independent professionally operated opinion sites such as *Crikey* and *Online Opinion*):

As Figure 1 shows, there is a very small circle of hyperactive bloggers and columnists, dominated by the professionals for whom the publication of blog-style columns is part of their paid work, but also including a handful of highly active independent, unpaid bloggers (see Kirchhoff et al., 2009, for a detailed discussion). There is also a broader mid-field of substantially active authors in which
independent bloggers are dominant. An analysis of the links received by these sites (from *within* the total population of sites which we tracked—that is, not counting links from nonpolitical blogs or other websites inside or outside Australia) shows a somewhat more contracted distribution of attention, however. Although, as Figure 1 indicates, there is regular publishing activity across a long tail of minor blogs, inlinks are concentrated on a much narrower range of blogs, many of which are among the most active political blogs that dominate Figure 1. In other words, while many of the blogs in the long tail contribute to public debate and link to the leaders, it appears that those leaders only rarely or never reciprocate—the minor blogs act only in a framing capacity, commenting on and encouraging their own readers to explore the leaders’ opinions directly by following the hyperlinks to the major blogs. If such processes are found to be a common feature of political communication in the blogosphere, this could be described as a kind of “reverse two-step flow,” a mirror image to the two-step communication flow postulated in the heyday of the early broadcast era (Katz & Lazarsfeld, 1955): where in that model, information transmitted in a first step through the “push” medium of television would in a second step be further disseminated socially, in this reverse model—appropriate to the web as an information “pull” medium—there is a flow of *readers* (rather than information) upward from minor to major blogs, which in turn provide commentary and criticism, as well as further hyperlinks, directed at the content of mainstream news media (Figure 2).

**Figure 1.** Publishing activity by amateur and professional bloggers and opinion sites (November 2007–January 2008).
Our examination of political blogging during 2009 also points to a variable distribution of attention across major issues (cf. Bruns et al., 2009; Highfield, Kirchhoff, & Nicolai, 2009); unsurprisingly, different blogs have divergent overall interests. Our identification of bloggers’ attitudes toward specific political topics builds substantially on large-scale textual analysis and automated keyword extraction to document the topical preoccupations of specific blogs and blog clusters. Figure 3, for example, shows a significant cluster of keywords related to social policy in posts made by the leading blog Club Troppo during and after the 2007 federal election. By contrast, other leading blogs paid considerably more attention to the day-to-day conflict of election campaigning or to the political process of governmental transition (Bruns et al., 2008b). Club Troppo, in other words, was focused more on debating the merits of social policy initiatives than the cut and thrust of day-to-day electioneering. This approach, then, provides a clearer picture of a blog’s orientation than generic labels such as “left-wing” or “right-wing.”

Similarly, we are able to observe the overall political blogosphere’s attention and focus, both for longer time frames (such as election campaigns) and during specific short-term political events or crises. It is especially interesting to compare the coverage of such events by bloggers and by the mainstream media. At the height of the 2009 “Utegate” affair over alleged (and later disproven) political interference by the Prime Minister in a government loans scheme, for example, our examination shows that even in the face of extensive mainstream media coverage of these events, blogs continued to debate the major long-term challenges facing the nation, rather than engaging to a
significant extent with the short-term political drama. Figure 4 shows keywords relating to Utegate (chiefly the names of major protagonists Godwin Grech and the then opposition leader Malcolm Turnbull, as well as the Ozcar government scheme) in a loose cluster of themes to the side of a much denser cluster of terms that relate to the two major challenges facing the government: climate change and the global financial crisis, and which centrally features the name of the then Prime Minister Kevin Rudd. The blogs quite literally positioned Utegate as a sideshow to the main business of national politics, in other words. Analysis of mainstream news content would likely show a very different distribution of central themes. What this observation points to is that political blogging in Australia does not simply echo the key themes positioned as important by leading news outlets but maintains its own perspectives on what is important to discuss.

Political blogging in a narrow sense only accounts for a relatively small subset of all blogging activities by Australians, of course. Indeed, a key limitation of much contemporary blog mapping research is the persistent but rarely examined assumption that the “political blogosphere” represents the most significant area of online public communication in any given nation-state. Our larger project takes a much broader, more diverse and inclusive view of what topics and practices can be considered “public communication.” In the following section, this model of public communication is described both as a response to recent and significant shifts in the dynamics of social media itself and on the basis of theoretical developments in media and cultural studies—the disciplines upon which this study is based.

**Public Communication beyond the “Political Blogosphere”**

Although “social media” like Twitter, YouTube, or Facebook are frequently characterized—or trivialized—as tools for individualistic self-expression or social networking, they are also at the same time playing an increasingly significant role in global public communication; in some cases at a scale that far exceeds the original intentions of the founders of the platforms concerned. This is an observable shift—neatly illustrated by the emergence, evolution, and widespread popularity of the short message social network site Twitter.com. Twitter recently changed its tagline (the short slogan that acts both as invitation to users and as an evocation of the platform’s overall purpose) from the “me-centred” “Twitter is for staying in touch and keeping up with friends no matter where
you are or what you’re doing”¹ to something more akin to a global mission statement built around real-time events: “Share and discover what’s happening right now, anywhere in the world.”

This demonstrates a significant change in how everyday communication and personal media figure in public communication, associated with the widespread (but by no means universal) uptake and usage of social media platforms. Via social network sites, the everyday lives of individuals are being remediated into new contexts of social visibility and connection—through Facebook and Twitter status updates, videos uploaded to YouTube, and photos contributed to Flickr. For media and communication studies, there are new theoretical questions to be asked about the connections between everyday life, media representation, and questions of publics; and these questions have additional methodological implications.

However, public debates about the meaning and uses of social media platforms still tend to get stuck in between celebration and condemnation (for examples of contrasting but equally hyperbolic views, see Grossman, 2006 and Keen, 2007) of the scale at which “ordinary” people are directly participating in the production and circulation of media messages, particularly via online social networks. Both the utopian hype and the dystopian counter-hype that structure these debates are based in technologically determinist positions—assuming that digital technologies have direct “impacts” on society: in these debates, social media represent either the “democratization” (Grossman) or the decay (Keen) of public communication. Even for more thoughtful scholarly critics, the empirically observable cultures of mass participation in online culture fail to measure up to the more utopian ideals of their proponents (see, e.g., Turner, 2010, pp. 123–157). The focus on journalism and political debate in researching blogs, for example, may reflect the preexisting concerns of critics more than the emerging patterns of participation in user-created content and social media. A 2006 survey

Figure 4. Key themes in Australian political blogs at the conclusion of the Utegate affair (August 4–5, 2009).
of American bloggers found that people create and contribute to blogs for a wide range of reasons and that they primarily blogged about their everyday lives and personal interests, with politics occupying a minority position (Lenhart & Fox, 2006).

In the ways these debates are framed, both the promise and the disappointments of participatory culture and social media implicitly or explicitly tend to rely on normative, Habermasian models of public communication, shaping the lens through which the actual uses of social media are perceived. It is necessary to move beyond normative assumptions about the role of media in the maintenance of the public sphere—and beyond assumptions of the role of very specific modes of media participation (the production and consumption of political news or rational debates about political issues) in public discourse. It is also necessary to develop a more pluralistic and flexible concept of publics to comprehend the role that everyday creativity and communication may play within it, in the context of online social networks.

The literature on post-Habermasian theories of how “publics” emerge, and under what conditions of mediation, is too voluminous to deal with adequately here. However, it is necessary to stake out some ground in this debate to position the analysis of how social media spaces, including not only blogs but also YouTube, Twitter, and Flickr, might represent platforms for the emergence of publics.

Famously, the dominant appropriations of the Habermasian model of the public sphere (Habermas, 1989) imagine a universally accessible space where informed citizens engage in the political process through rational–critical debate. However, even without considering the impact of social media, from an identity-aware perspective the protocols of rational–critical debate are far from inclusive. As Warner (2005, pp. 51–54) argues, these protocols “include the idea that one needs to bracket one’s private self in order to engage in public discussion”—that is, the privilege to “rise above” one’s social or cultural identity, personal feelings, and lived experience, to engage in legitimate public debate. For Warner and others, while the public (singular) is an abstraction, “a kind of social totality” that tends to be imagined at the level of the nation-state (Warner, 2005, p. 65), there are also multiple, concrete publics (plural), constituted via communication, and structured by affect as much as by rational–critical debate. Such engagement can occur in and through popular culture (Hermes, 2006; McGuigan, 2005) and everyday communication, particularly around events of shared concern (Hartley & Green, 2006). By decentering more formalized spaces of rational debate (such as political journalism or even political blogging) and taking account of emergent, ephemeral communication, this literature provides us with a way of aligning contemporary theories of public communication with the observable dynamics of social media.

These theoretical moves—from “the public sphere” to affective and emergent publics—have obvious applications in researching the ephemeral and multiple discursive interactions that take place online via social media platforms such as Twitter, Flickr, and YouTube; where there are a large number of relatively contained interest-based communities and where the same networks that support mundane and interpersonal communication are at times activated around matters of shared public concern—from celebrity deaths to natural disasters and national elections. We argue that social media constitute platforms for the emergence of publics, in this post-Habermasian sense, and it is this proposition that drives our development of new empirical mapping methods for social media beyond the political blogosphere.

Looking Ahead

New empirical research is required that can track these identity or interest networks, to quantify and qualify the extent to which the remediation of everyday life via social media might be changing the face of public communication in these ways. For this, more large-scale, computer-assisted methods building on existing social media mapping techniques are required. However, this shift has major
methodological implications for cultural and media studies. Work based on the framework above has,
to our knowledge, never been undertaken in conjunction with large-scale computer-assisted tech-
niques. Instead, it has used ethnography to understand the motivations and meaning-making practices
of small groups of users (Lange 2007) or has elaborated new theoretical perspectives on the basis of
existing literature combined with small selections of illustrative texts (Crawford, 2009; Jenkins 2006).

Accordingly, coming stages of our research (during 2011 and 2012) will extend and adapt our
established blog mapping approach to other social media spaces such as Twitter, Flickr, and You-
tube, without restricting the sample to data sources that are narrowly defined as “political” in focus.
As we move deeper into tracking user activities in these various social networks, we will also
increasingly confront a range of complex ethical considerations, even if prima facie all of the mate-
rial which we capture is publicly available and thus—as published material—does not require us to
seek permission from authors to analyze it. Although any of the tweets we will examine are public
and discoverable through the standard Twitter interface, for example (we will not include any con-
tent that has been protected from public viewing), not all of them will have been intended for wider
public dissemination. As social networking tools are contributing substantially to a blurring of pub-
lic and private communication, then, so must our research ethics and research methodology work to
find appropriate answers to these challenges to enable us to continue our work while protecting the
privacy of those whose material it builds on.

Declaration of Conflicting Interests

The author(s) declared no conflicts of interest with respect to the authorship and/or publication of
this article.

Funding

Parts of this paper refer to work undertaken with the support of an Australian Research Council
Discovery Grant.

Note

twitter.com/.

References


blogosphere. Paper presented at Internet Research 10.0—Internet: Critical, Milwaukee, WI.


blogosphere in the aftermath of the 2007 federal election. Paper presented at Association of Internet

Hjorth (Eds.), Mobile Technologies: From Telecommunications to Media (pp. 252-256). New York: Routledge.

magazine/article/0,9171,1569514,00.html


**Bios**

**Axel Bruns** (a.bruns@qut.edu.au) is an associate professor in the ARC Centre of Excellence for Creative Industries and Innovation (CCI). See http://snurb.info/.

**Jean Burgess** (je.burgess@qut.edu.au) is a senior research fellow in the ARC Centre of Excellence for Creative Industries & Innovation, Queensland University of Technology.

**Tim Highfield** (t.highfield@qut.edu.au) is a PhD student at Queensland University of Technology, Brisbane. His thesis looks at political blogging in Australia and France.

**Lars Kirchhoff** (lars.kirchhoff@sociomantic.com) is VP Product Management of sociomantic labs. He researches Information Systems, Social Media, and Communication Management.

**Thomas Nicolai** (thomas.nicolai@sociomantic.com) is the managing director of sociomantic labs. He has contributed to Internet Research, Information Systems, and Design Science.