



Queensland University of Technology
Brisbane Australia

This is the author's version of a work that was submitted/accepted for publication in the following source:

[Highfield, Tim](#), Kirchhoff, Lars, & Nicolai, Thomas (2011) Challenges of tracking topical discussion networks online. *Social Science Computer Review*.

This file was downloaded from: <http://eprints.qut.edu.au/41656/>

© Copyright 2010 SAGE Publications

Notice: *Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source:*

<http://dx.doi.org/10.1177/0894439310382514>

Challenges of tracking topical discussion networks online

Tim Highfield (Queensland University of Technology)
Lars Kirchhoff, Thomas Nicolai (Sociomantic Labs)

Abstract

Attempts to map online networks, representing relationships between people and sites, have covered sites including Facebook, Twitter, and blogs. However, the predominant approach of static network visualisation, treating months of data as a single case rather than depicting changes over time or between topics, remains a flawed process. As different events and themes provoke varying interactions and conversations, it is proposed that case-by-case analysis would aid studies of online social networks by further examining the dynamics of links and information flows.

This study uses hyperlink analysis of a population of French political blogs to compare connections between sites from January to August 2009. Themes discussed in this period were identified for subsequent analysis of topic-oriented networks. By comparing static blogrolls with topical citations within posts, this research addresses challenges and methods in mapping online networks, providing new information on temporal aspects of linking behaviours and information flows within these systems.

Keywords

blogs, France, methodology, network mapping, social media, visualisation

Introduction

The network has become a widely-used concept for describing online friendships and connections between websites, appearing in studies of sites from Twitter (Huberman, Romero, & Wu, 2009) and blogs (Park & Thelwall, 2008) to profiles on Facebook (Gaines & Mondak, 2009). As automated data collection and visualisation processes have developed, the scope of these projects has become greater. There is now a wealth of network-oriented research within internet studies and associated fields. Often using hyperlinks as a principal data source, the gradual improvement of methods for studying online networks have enabled long-term observation of large systems such as blogospheres (for example, Etling, Kelly, Faris, & Palfrey, 2009).

While these long-term projects provide valuable overviews of how the chosen sample of sites are connected and form groups amongst themselves, their findings may be limited by the aggregation of

data in their analyses. Reports such as that by Etling et al. (2009) on the Arabic blogosphere may cover several months of blog activity, but use a composite view of links and sites in providing a summary of the connections made over the entire period. While this approach presents an overview of groups and links created during the study, it does not explain when or why links were made. As such, the next step for research into networks of sites such as blogs may be the analysis of smaller datasets and topics within an examination of longer periods of online activity. This paper outlines a preliminary attempt at tracking topical discussion networks – the collection of sites commenting on a particular event or issue, and the links between them – within a long-term study of French political blogging. In developing this approach, the research also notes some of the challenges facing this process, and suggests further methodological and conceptual aspects that may be improved by additional work in this field.

The research outlined here forms part of a larger, ongoing project comparing activity within the French and Australian political blogospheres between January and August 2009. Earlier work using this dataset has looked at the text content of blog posts around particular events, such as the inauguration of Barack Obama (Highfield 2009a/b), and bushfires and political scandals (Bruns, Highfield, Kirchhoff, & Nicolai 2009). The method tested in this study will be developed further during the wider project, using these events as additional case studies.

Online (social) networks

The development and popularity of social network sites such as Facebook and MySpace have led to various studies and methods investigating how these sites are used. Offering users the opportunity to create their own profile and connect to other individuals' profiles through a declaration, and acceptance, of friendship, these sites may be open to all, or built around a particular group of

people, interest, or location (boyd & Ellison, 2008). The ties between profiles can be used to determine an individual's (platform-specific) social network. Similarly, connections amongst sites that might not be considered as 'social network sites', such as blogs and Twitter, can be found through hyperlinks or 'following' relationships.

Concepts from network studies in sociology and mathematics, such as weak ties (Granovetter, 2001), small worlds (in Watts, 2004), and hubs and nodes (in Barabási, 2003) have been adapted for research into online networks. Through the application of social network analysis to computer-mediated communication, methods such as hyperlink analysis have been developed, forming part of the webometrics approach to measuring and analysing internet-based activity (Thelwall, 2009).

Attempts to study and map networked aspects of social network sites and blogs have looked at various phenomena, from interests and connections between university students on Facebook (Lewis, Kaufman, Gonzalez, Wimmer, & Christakis, 2008) to how news spreads along networks of Twitter and Digg users (Lerman & Ghosh, 2010). Amongst research into blog networks, an early attempt at tracking topical discussions in the blogosphere can be seen in Bruns' (2007) use of IssueCrawler to monitor mentions of David Hicks, an Australian detained at Guantanamo Bay.

Political blogging in France

Although research into topical networks online may focus on features such as Facebook status updates or trending topics on Twitter, this study looks at discussions posted and links created on French political blogs. Political blogging is one of many subgenres of blogging, alongside blogs dedicated to subjects as diverse as music, cooking, or knitting. However, political blogs are an ideal subject for research into discussion of news events, and interactions with the mainstream media,

through their commentator roles and referencing behaviours.

Defining a blog as political is not a straightforward process. However, related research has highlighted common characteristics of sites featured in political blogging studies, such as the sharing of political opinions, commentary on political events, policies, campaigns, and the work of the mainstream media, and campaigning for candidates or political issues, with blogs often forming highly polarised groups around ideology or party affiliation (see, for example, Drezner & Farrell, 2008; Tremayne, 2007; Cardon & Delauney-Teterel, 2006). Political blogs may also employ various approaches to their discussions. Vanbremeersch (2009) describes how the formation of blog networks leads to bloggers taking on different, complementary roles and practices. These roles include commentators, activists, and link-bloggers, whose posts are mainly lists of links, as bloggers attempt to find their place within the network and their own, individual voice.

The French blogosphere has been an active system for several years. By 2006, French bloggers had become the majority population amongst European bloggers (Forrester Research, 2006). Political blogging in France includes representation from national and local politicians and their advisors, militant supporters of parties, and journalists and analysts as well as politically engaged citizens. As with its U.S. counterpart, research into the French political blogosphere has investigated activity both during elections (see Vaccari, 2008; RTGI, 2007), and major political events, including campaigns surrounding the proposed European Union (E.U.) constitution (Fouetillou 2006b; Franck 2005), and social unrest in the banlieues of Paris and other major French cities (Echchaibi, 2009; Allard & Blondeau, 2006), both in 2005.

The 2007 French Presidential election also stimulated political blogging. As well as providing a platform for commentary, opinion, and partisan support, blogs and videoblogs were used as

campaign tools, particularly by the Parti Socialiste (PS) candidate Ségolène Royal. The online strategy of the PS included the creation of a large support network of blogs under the banner of the Royal campaign, forming a 'Ségosphère' or 'Ségoland' (Vaccari, 2008), contributing to the PS web presence during the election campaign (see RTGI, 2007; Fouetillou, 2006a). Large-scale studies of the political blogosphere, such as the election-oriented *Blogopole* (RTGI, 2007), have shown that French political blogs can be seen to form several clusters organised around political parties, and the presence of these clusters and self-defined groups of blogs will be further examined in this research.

Methodology

This paper forms part of a larger, comparative study of French and Australian political blogging, using data collected by Sociomantic Labs between 12 January and 10 August 2009. The data collection draws upon seed lists created using methods and reasoning developed in earlier studies of the Australian political blogosphere (see Kirchhoff, Nicolai, Bruns, & Highfield, 2009). The French seed list drew upon studies of the French political blogosphere by RTGI (2007) and ranking sites such as *Wikio.fr* to locate potential blogs for the study. In-post citations and blogroll links from these sites were then followed in a manual snowballing approach, identifying additional sites to include in the sample. Major national and regional French newspapers, as well as television, radio, and alternative media sites, were all included in the initial seed list, with a view to further comparative work expected to look at coverage of different topics in blog posts and news articles.

The seed list was created using an inclusive process. Blogs with some political content were included, rather than restricting the sample to sites entirely focussed on politics. If a new site appeared several times in blogrolls from sites already in the seed list, it would also be added, even if

its posts were not always about political issues or characters. As has been mentioned earlier, the term 'political blog' does not have a set definition, and it was decided that the seed list would include sites discussing several topics besides politics, in addition to those dedicated to specific political characters, parties, or issues. This inclusive process aids this paper, providing additional sites and data for evaluating from where bloggers get their information, and to which sites they may link. In particular, comparing different types of blogs, from party-specific to the more general commentators, may provide further information on the shape of topical networks across sections of the French political blogosphere.

An initial list of 450 French language political blogs and related sites, and the addresses for their respective feeds, was prepared. Sociomantic Labs carried out the crawling, scraping, and data preparation stages of the project between January and August 2009. In total, 22,939 posts were gathered from 148 active blogs over 221 days, from an overall return of 167,330 posts and articles from 217 active feeds. The data successfully collected represents activity from 48 percent of the feeds in the original seed list.

This paper acts as a preliminary test of method and sample. Later stages of this project will look at the hyperlinks and post content of other topical networks as well as the wider period covered by the collected data. For this study, a single case has been isolated from the seven months of data, acting as an example of a topic discussed within the French political blogosphere. The period chosen coincides with the inauguration of Barack Obama as U.S. President on 20 January 2009, and is based on previous research that used the event as a case study for an analysis of keyword content of French blog posts (Highfield, 2009b).

Findings

For this study, the networks featured were located using links originating from the 217 sites from which data was successfully collected. Two types of links feature: blogrolls and in-post citations. Blogrolls are relatively static page elements on a blog, often found in a sidebar, and comprise a list of links to sites that the blogger finds interesting, are run by friends or associates, or with which the blogger is, or would like to be, associated. In-post citations are links found within the text of blog posts, and are usually related to the specific content of the post.

Visualising blogroll networks

Before examining topical networks, such as that discussing the Obama inauguration, it is important to have a sense of the connections already present between the featured blogs at the stated structural level of the political blogosphere. This may be seen through identifying and visualising the network formed by blogrolls. Although blogroll links are not an accurate means for determining the interests of bloggers at a specific moment, they do provide a sense of the connections between sites, the blogs with which they feel affiliated or from which they may draw information.

For the purposes of this study, a blogroll was treated as a semi-permanent section of a blog, be it a list in a sidebar on the main blog page or on a separate page providing a list of links, clearly linked from the main page. The inclusion of lists on separate pages allowed for a more comprehensive view of the semi-permanent connections between the sites in the sample, especially as some blogs highlight a selection of links from the overall blogroll on the main page, but only show the full list on a separate page. Some blogs in the sample did not feature blogrolls in either format, while a few used automated widgets, operated from a third-party, which highlighted only some of their blogroll links. The links in these third-party widgets were not successfully extracted for this study, and so

were not included in the list of edges used to create Figure 1. In total, 193 of the 217 sample sites contributed at least one out-link from a form of blogroll. Following each link made on a blogroll on these sites leads to a directed network containing 5580 individual blogs, mainstream and alternative media sources, and other websites.

The number of links listed varied between sites. 52 sites contributed ten or fewer blogroll out-links, while twenty-one other sites had over 100 equivalent links, with the largest blogroll containing 518 individual out-links. While initial analysis of the total out-links from blogrolls provides some hints of ideological clusters amongst the sites being studied, the network is skewed by the sites with larger blogrolls, which are surrounded by many smaller nodes only receiving one in-link. As such, Figure 1 depicts only those sites that receive at least two in-links, as well as the initial 217 sites. Even if some of the sample sites did not feature blogrolls themselves, if they received at least one in-link they were still included in the blogroll network visualisation for the benefit of categorising sites and identifying clusters ahead of the topical network stage of this study.

The list of links was run through visualisation software *Gephi* using the *Force Atlas* force-directed algorithm developed by the *Gephi* team. This algorithm performs a similar function to other force-directed algorithms, such as the Früchterman-Reingold (1991) algorithm, wherein nodes are arranged through attraction and repulsion forces, but with a view to creating clearer, more compact visual aids for research (Bastian, Heymann, & Jacomy, 2009). In addition to reducing the visual clutter by only including sites receiving multiple in-links, the resultant visualisation shows the presence of inter-linked groups amongst the wider collection of sites. While maps of hyperlink networks and their analysis are restricted by the arbitrary boundaries of research projects (De Maeyer, 2010), visualisation is still useful in providing an overview of large datasets, which can be further investigated through studying specific examples and additional methods.

[IMAGE ONE ABOUT HERE]

Groups found within Figure 1 include self-defined blog communities, such as *les vigilants* - bloggers opposed to French President Nicolas Sarkozy – or sites promoting specific parties or personalities. The presence of a party or politician’s name in the site URL or title was used to identify initial groups of sites dedicated to specific parties. For additional visual clarity, these party-specific sites have been grouped together in Figure 1.

An extreme right-wing cluster, including sites pertaining to the Front National (FN), is easily identified at the top-right of Figure 1. Below that cluster are blogs and sites associated with the centre-right Union pour un mouvement populaire (UMP), and supporting political figures from that party such as Sarkozy, former President Jacques Chirac, and former Prime Minister Dominique de Villepin. Close to the UMP cluster are several official parliamentary sites, including the two French Houses of Parliament and the European Parliament site. To the left of Figure 1 is the group of PS sites and blogs, while to the right are supporters and websites from the centrist parties Mouvement Démocrate and Nouveau Centre. The mainstream media, including the main pages of newspapers such as *Le Monde*, *Liberation*, and *Le Figaro*, as well as television and radio websites, occupy a more central position in the network. However, ‘alternative’ or web-only media websites, including *Rue89*, *Bakchich*, and *Agoravox*, are amongst the most central sites to the network, and receive larger numbers of in-links from blogrolls than the mainstream media sites. Finally, a number of self-defined groups, not solely politically-themed but covering a range of topics and approaches, can also be located within the visualisation, spreading across the network between the PS and centrist clusters. The position of these collectives may be reflective of their political opinions and beliefs, from the anti-Sarkozy *vigilants* to the *Réseau LHC* (*Liberté d'expression, Humanisme et*

Esprit Critique, or Free Expression, Humanism, and Critical Thinking).

Figure 1 shows that alternative media, and some mainstream media, sites are relatively central to the blogroll network, while the party-specific sites form more distinct, inter-linked groups – or, in the case of the FN, form a cluster distant from the wider network. However, while the mainstream media are closely grouped in Figure 1, the position of individual sites such as *Liberation* or *Le Figaro*, which provide different political perspectives on issues, might be expected to change within a topical or time-limited network. Indeed, the relationship between blogs and the mainstream media might not be accurately represented by a blogroll-only network, with inconsistent linking to non-blog sites from the 217 initial nodes. Although blogrolls provide a space for bloggers to state their affiliations and associations, linking to the sites with which they would like to be aligned and from where they might source their information, Hargittai, Gallo, and Kane (2008) note that, 'it would be too simplistic to rely solely on such information for an understanding of bloggers' reading habits' (p85). As such, the importance of the topical networks to studying interactions between blogs and other sites is underlined by the limitations of blogroll networks.

Visualising topical networks

From the seven months of blog activity collected for this project, potential case studies for topical network identification and analysis were evaluated. Following content analysis of blog posts around specific events from the same period (Highfield, 2009b), one of these events, the Obama inauguration, was chosen as a pilot study to test the method for analysing topical networks. The inauguration itself took place on 20 January 2009, rather than being an ongoing, long-term event. However, for the benefit of tracking topical links both before and in response to the event, posts made on surrounding days were also studied. As such, posts published between 16 January and 25

January 2009 were included in this case study. Posts from the chosen date range were extracted from the overall dataset, with the total out-links from blog posts in this period, as well as a further subset of blog posts discussing the inauguration itself, used in the analysis. Internal links, advertising, and 'share this' tools, were removed from the list of out-links prior to analysis.

Looking at the total out-links for the inauguration period, mainstream media sites were among the most frequently referenced sites, occupying more central positions within the in-post citation network than in Figure 1. These sites are some of the largest nodes in the time-limited network, alongside the collaborative encyclopaedia Wikipedia and video-sharing site Dailymotion.

Alternative media and citizen journalism sites, which were central nodes in Figure 1, did not receive as many in-links during this period as the mainstream media. Similarly, the more popular blogs, while still seeing some linking activity, are not as widely cited within the time-limited network as in Figure 1.

Several peripheral clusters were also found around individual blogs, as bloggers linked to unique sites. Such behaviour may suggest that different blogs refer to a variety of sources, with only a few commonly-referenced sites across the sample, such as mainstream media sites. However, this may also be an artefact of focussing on a short period of blog activity, as some blogs only published one post, which could have been on any number of topics.

To further evaluate these observations, posts discussing people and topics surrounding the Obama inauguration were extracted from the total 1779 posts published between 16 and 25 January.

Subjects featured included the presidency of George W. Bush, the internet strategy of both the Obama campaign and the Obama presidency, and the decision to close the Guantanamo Bay detainment camp. 170 posts published over the ten days of the case study were used for this stage of

the analysis. All out-links from these posts were used to generate the visualisation in Figure 2.

While not all of these links were Obama-related, the preliminary analysis provided here gives an overview of the discussions and links around the inauguration and other sites or topics that bloggers mentioned in the same space. For this study, links were truncated to the top-level domain rather than looking at individual stories, to enable an initial analysis of popular references over this period.

[IMAGE TWO ABOUT HERE]

The observations from the time-limited network are supported in the visualisation of the Obama topical network. Mainstream media sites, Wikipedia, and video-sharing site Dailymotion occupy central positions in the network, although there are notable groups of sites providing in-links to different mainstream media sites. For example, the blogs linking to *La Croix* or *Nouvel Observateur* are mostly distinct from the blogs linking to *Liberation* or *L'Express*. While further analysis and additional case studies are required to identify whether different groups of blogs generally refer to different media sources, this case does suggest that mainstream media articles are cited more often than other blogs and alternative media. Indeed, despite the prominence of three alternative and citizen media sites within the blogroll network, only one of these sites, *Rue89*, received any in-links in this particular study.

The network in Figure 2 can also be seen to contain a large central group, comprising the major references for the topic, and several blogs that published multiple posts during the period covered here, such *ilovepolitics.info* and *sarkofrance.wordpress.com*. Around this central group, the more peripheral sites link to several unique sources as well as the more popular references. While this is not unusual, particularly when looking at network visualisations, the position of sites such as *ilovepolitics.info* and their receiving multiple in-links from other blogs in the network suggests that

a model of information flows involving both mainstream media and prominent blogs could be a focal point for further research.

As with the blogroll network, some sites and groups are not strongly connected to the main group. Given the specific focus of the topical network, there are also several pairs of nodes that are entirely separate. In Figure 1, the largest cluster partly separate to the overall network was that aligned with the FN, and in Figure 2 sites from this cluster are disconnected from the rest of the topical network, instead found in floating pairs. However, further case studies are required to test the applicability of these observations to other topical networks.

The international aspect may also account for other features of the topical network. Event-specific sites, such as *barackobama.com* and *whitehouse.gov*, are recipients of several in-links, and international media such as the *New York Times* were also cited. The respective links to video-sharing sites Dailymotion and YouTube might also be a result of the international focus of this study. Dailymotion is the main French rival to YouTube, with both sites launched in 2005 (Yanoshevsky, 2009). While Dailymotion received 150 in-links during the ten days covered in this study, when the data is restricted to posts discussing Obama the site only received ten in-links. Although YouTube, in its *youtube.com* and *fr.youtube.com* forms, received only eleven in-links across the whole period, nine of those in-links are from posts identified in the Obama case study. The inauguration as international event explains some of the links to YouTube, with links to U.S. videos, such as Obama's first presidential press conference, and to official White House and inauguration channels. For events or topics that are directly related to French politics, it might be expected that Dailymotion would receive more in-links.

While not a comprehensive study of topical discussion networks in the French political

blogosphere, the analysis of links made around the Obama inauguration suggests further directions for research into these networks. The number of links towards mainstream media sites, for example, may be compared across different case studies to evaluate whether such referencing is typical within the sample or an artefact of discussions surrounding an international event. The visualisations presented in Figures 1 and 2 provide a starting point for comparative work looking at both semi-permanent and more temporary topical networks across a longer study of blog activity, and the method outlined here may be adapted for research into other social network sites.

Discussion

Types of hyperlink

While hyperlinks are the main data source in attempting to track topical networks within the blogosphere, they also contribute conceptual challenges to the research. An important consideration for this project was how to treat links in terms of type and lifespan. As the most visible means of showing connections or associations between bloggers, whether reciprocated or not, Lovink (2008) describes linking as the 'logic' underpinning blog networks. However, with several different types potentially present on a blog, links can also complicate studies. For example, the different link types and their use by bloggers can influence how blog networks are formed and mapped, from networks of blogroll links (Park & Thelwall, 2008) to a mixture of link forms (Adamic & Glance, 2005).

Although hyperlinks take the same basic HTML form, regardless of their purpose, research projects cannot always treat all links featured in the dataset as being the same in function or intent. As Halavais (2008) notes, while the hyperlink may have been seen primarily as a citation device, it now serves navigational, organisational, and social roles. With hyperlinks becoming ubiquitous, identifying different link types and selecting those which are relevant to the project at hand is a

crucial stage of research in this field.

Bruns, Kirchhoff, Nicolai, Wilson, Saunders, and Highfield (2008) list several link types encountered in blog research, from generic links, including site navigation and advertisements, to links provided by people leaving comments on posts. Other projects may classify links even more specifically, making distinctions, for example, between links made by the blogger both within the post itself, and those they may have added in comments. Methods for determining the semantics of a link on a large scale, deciding whether connections are being made as an indication of the author agreeing or not with the site being linked to, are another aspect of blog network research that is still being developed. Hargittai et al. (2008) describe five types of cross-ideological in-post citations in their study of the U.S. political blogosphere, ranging from agreement to cursory disagreement with another blogger's post. Although Lovink (2008) claims that bloggers will only link to like-minded individuals, studies such as that by Adamic and Glance (2005) have found some links between bloggers across ideological divides, and these may reflect both support and disagreement. While this preliminary study does not take link semantics into account, analysis of these different forms of connection is still an important step for blog research methodologies.

The lifespan of links

Temporal factors may also need to be taken into account when developing the concept of the hyperlink. For how long do links remain as ties between sites, as relevant or current indicators of association? A project may feature several months' of data from a sample of blogs, including posts and links, but should the presence of one individual link from site A to site B over the entire period be treated as evidence of a connection between the two sites? This study has looked at a topical network formed from links made over a ten day period, but should the lifespan of the links included affect how such networks are studied?

Adamic (2008) describes an in-post citation as both a 'permanent and fleeting connection', demonstrating the deliberate choice of a blogger to cite another site, but with its relevance and visibility fading as more posts are made on the blog. Similarly, reciprocity of links and the number of times a link is made may also be factors used to determine the lifespan and strength of blog ties. Hyperlinks might have a kind of half-life, achieving their highest visibility on a blog at the time of publication, and with their relevance to the chosen topic 'decaying' as more posts are made, moving the post featuring the initial link into the blog's archives. The longevity of in-post citations may be dependent on the general posting activity of a blogger – the more active a blogger is, the shorter the time required for a post to move off the main blog page. Eventually, blog posts on a specific topic will not be as noticeable, with the blog archives becoming the record showing that 'a conversation has taken place, but it has moved out of view' (p230).

A link's lifespan also varies due to the type of link. A blogroll link, for example, can be treated as a more permanent, longer-lasting association from blog A to site B than an in-post citation. However, as blogrolls are more generic in their linking, rather than being relevant to a particular topic, and are rarely updated, they are also at risk of becoming 'stale' (Adamic & Glance, 2005). While in-post citations are the most apt source for locating topical networks, even the connections made over the course of these discussions are not necessarily evidence of permanent or strong ties between sites. If a link is made from blog A to site B once over a six month period, then although it may have been an important reference for a topic, the link itself does not suggest a long-lasting connection. On the other hand, if blog A links several times to site B over the period, then the connection may be seen as being renewed and support a view that site B is read often by blogger A. For links made as in-post citations to be seen as strong ties, though, there is a need for repeated references over time, for 'although possibly a stronger indication of interest than a blogroll link, citations, unless periodically

repeated, will become dated' (Adamic, 2008, p230).

Although this research deals specifically with blog data, the lifespan of links as indicators of connections between sites and actors is relevant to other platforms. Indeed, for sites such as Twitter, the longevity and visibility of links may be even more variable than for blogs, due to the character limit in posts leading to the posting of multiple tweets for several links. While the ambiguities in studying hyperlinks may be avoided by looking at other ways of determining association between sites, such as through keywords, the importance of the hyperlink in research, especially to quantitative internet-oriented projects, underlines the need for further development of the concepts surrounding linking.

Limitations

This paper acts as a pilot study for research into topical networks amongst blogs. Focussing on discussion around an international event was an ideal choice for preliminary development of this method as the posts and links concerning the inauguration could be located easily within the chosen period. However, due to its international subject, the topical network located here might not be representative of other topical networks discussing French themes within the same dataset. The topical network discussed here was created by looking at links between sites at the domain level, rather than at individual articles or distinguishing between multiple authors on the one site (further discussed by Benkler & Shaw, 2010). As such, links shown here do not mean that two blogs linking to the same site were citing the same article as each other, but referencing the same domain.

As political blogging is only one genre of blogging, the findings from this study cannot be treated as general conclusions about the sites from which bloggers obtain their information, nor are the

findings of this study representative of French political blogging. The percentage of the original seed sites from which data was successfully collected means that some political groups may be under-represented in this study. For example, Les Verts (the Greens), was only represented by its official site in the blogroll network featured in this paper, in comparison to both the other parties and other studies of the French political blogosphere (RTGI, 2007). While this may be related to restricting the initial sites in this visualisation to those from which posts were obtained during the data collection phase, other sites in the network may also be supporters of Les Verts, but do not explicitly state this in their URL or title. As such, further work in determining, where possible, political ideologies of other sites in the sample may follow, while additional research and expansion of the sample in future projects will help to provide more information on the behaviour of the wider network.

Further directions

This study has presented a preliminary study attempting to track a topical discussion network within long-term data collection and research into the French political blogosphere. In the next stages of the project, additional case studies will be identified and analysed for comparison with both the blogroll network of Figure 1 and the Obama-specific topical network of Figure 2. The aim of such comparisons is to see how topical networks may change between cases and over time. Further work using this dataset may also investigate the presence of clusters of sites acting as information sources for different groups of blogs, as discussed by Rogers (2009) and seen in the study of the Persian language blogosphere (Kelly & Etling, 2008).

The methodology used in this study will be refined over further case studies, developing a more comprehensive method for comparing topical networks with the semi-permanent networks created

from blogrolls or follower/following lists. Further development of the method could also include comments on blog posts, providing an additional angle to research of conversations within the blogosphere. Although the study of links between posts and the topical content of posts during specific periods provides preliminary findings on blog coverage of news themes and the temporary networks associated with these posts, comments arguably allow for greater elaboration on the content of a specific post. While tracking comments from multiple blogs and pseudonyms may provide more methodological challenges, it could be argued that without including comments in the analysis, studying only blog posts does not take into account every contribution to the discussion.

Research into topical discussion networks will also need to take into account different platforms, such as Twitter, and how to collect data from these sites, as analysis looking at only blogs cannot cover the full online discussion of a particular issue. Further work in this field may be best served by tracking multiple platforms to get a wider view of the various topical discussions taking place online (see Bruns, Burgess, Highfield, Kirchhoff, & Nicolai, 2010). However, different social network sites come with their own challenges, from user privacy to changes in site architecture and security. These challenges will need to be taken into account as research projects tracking topical networks progress, while new social network sites might also warrant study as they develop.

Conclusion

Treating the total links produced and received by sites and users within an online social network is one way of gauging the network's level of activity, and has been used in previous research into these systems. However, other approaches can highlight additional information. Rogers (2009) suggests that the participants within a network 'may also be profiled by examining which particular links they give or receive' (p14). This idea has been adapted for this study, focussing on links within

particular periods and around a chosen subject in the identification of topical discussion networks online.

This research has provided a preliminary overview of a method for studying topical discussion networks online, through a case study of French political blogs. Although this approach is still being developed and challenges overcome, the findings from the inauguration-oriented topical network highlight aspects of online communication that may be investigated to a greater extent in future research. Topical networks provide a means for studying how online communication varies over time and subject, and which resources may be widely cited amongst a population of sites. When studied in combination with long-term analysis of a sample of sites, such as blogs, topical networks can provide additional information on how these sites are used. Topical network analysis can then become a complementary method for further research into online communication.

References

- Adamic, L. A. (2008). The Social Hyperlink. In J. Turow & L. Tsui, *The Hyperlinked Society: Questioning Connections in the Digital Age* (pp. 227-249). Ann Arbor: University of Michigan Press and University of Michigan Library.
- Adamic, L. A., & Glance, N. (2005). *The Political Blogosphere and the 2004 U.S. Election: Divided They Blog*. Retrieved from <http://www.blogpulse.com/papers/2005/AdamicGlanceBlogWWW.pdf>
- Allard, L., & Blondeau, O. (2006). Racaille digit@le: Les émeutes de banlieues n'ont pas eu lieu. *Contemporary French Civilization*, 31(1), 197-217.
- Barabási, A. (2003). *Linked: How everything is connected to everything else and what it means for business, science, and everyday life*. New York, NY: Plume.
- Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. In *Third International AAAI Conference on Weblogs and Social Media*. AAAI. Retrieved from <http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154>
- Benkler, Y., & Shaw, A. (2010). *A Tale of Two Blogospheres: Discursive practices on the left and right*. Berkman Center for Internet & Society: Harvard Law School.

- boyd, d., & Ellison, N. B. (2008). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210-230.
- Bruns, A. (2007). Methodologies for mapping the political blogosphere. *First Monday*, 12(5).
- Bruns, A., Burgess, J., Highfield, T., Kirchhoff, L., & Nicolai, T. (2010). Mapping the Australian Networked Public Sphere. (**this volume**)
- Bruns, A., Highfield, T., Kirchhoff, L., & Nicolai, T. (2009). *Critical voices in the Australian political blogosphere*. Paper presented at Internet Research 10.0 – Internet:Critical, Milwaukee, Wisconsin, 7-10 October 2009.
- Bruns, A., Kirchhoff, L., Nicolai, T., Wilson, J. A., Saunders, B. J., & Highfield, T. (2008). *Discussion paper: Network and concept maps for the blogosphere*. Retrieved from <http://eprints.qut.edu.au/archive/00013580/01/13580.pdf>
- Cardon, D., & Delaunay-Teterel, H. (2006). La Production de Soi comme Technique Relationnelle. Un essai de typologie de blogs par leurs publics. *Réseaux*, (138), 15-71.
- De Maeyer, J. (2010). *Methods for mapping hyperlink networks: Examining the environment of Belgian news websites*. Paper presented at International Symposium of Online Journalism, Austin, TX, April 2010.
- Drezner, D. W., & Farrell, H. (2008). The power and politics of blogs. *Public Choice*, 134(1-2), 15-30.
- Echchaibi, N. (2009). From the Margins to the Center: New Media and the Case of Bondy Blog in France. In A. Russell & N. Echchaibi, *International Blogging: Identity, Politics, and Networked Publics* (pp. 11-28). New York, NY: Peter Lang.
- Etling, B., Kelly, J., Faris, R., & Palfrey, J. (2009). *Mapping the Arabic Blogosphere: Politics, Culture, and Dissent*. Berkman Center for Internet & Society: Harvard Law School.
- Forrester Research. (2006). *Profiling Europe's bloggers*. Forrester Research. Retrieved from <http://www.forrester.com/ER/Press/Release/0,1769,1112,00.html>
- Fouetillou, G. (2006a). Ecologie de la Ségosphère. Retrieved from <http://www.observatoire-presidentielle.fr/?pageid=12>
- Fouetillou, G. (2006b). Le web et le débat sur la constitution européenne en France. Retrieved from <http://www.observatoire-presidentielle.fr/?pageid=20>
- Franck, R. (2005). Why did a majority of French voters reject the European Constitution? *European Journal of Political Economy*, 21(4), 1071-1076.
- Früchterman, T. M., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Expertise*, 21(11), 1129-1164.
- Gaines, B. J., & Mondak, J. J. (2009). Typing together? Clustering of ideological types in online social networks. *Journal of Information Technology & Politics*, 6(3), 216-231.

- Granovetter, M. S. (2001). The Strength of Weak Ties. In D. B. Grusky, *Social stratification: Class, race, and gender in sociological perspective* (pp. 447-451). Boulder, Colorado: Westview Press.
- Halavais, A. (2008). The Hyperlink as Organizing Principle. In J. Turow & L. Tsui, *The Hyperlinked Society: Questioning Connections in the Digital Age* (pp. 39-55). Ann Arbor: University of Michigan Press and University of Michigan Library.
- Hargittai, E., Gallo, J., & Kane, M. (2008). Cross-ideological discussions among conservative and liberal bloggers. *Public Choice*, 134(1-2), 67-86.
- Highfield, T. (2009a). *Inauguration from afar: Mapping Obama discussion in the Australian political blogosphere*. Paper presented at ANZCA 09: Communication, Creativity, and Global Citizenship, Brisbane, 8-10 July 2009.
- Highfield, T. (2009b). *Themes and discussions from the French political blogosphere, January-August 2009*. Paper presented at Internet Research 10.0 – Internet:Critical, Milwaukee, Wisconsin, 7-10 October 2009.
- Huberman, B. A., Romero, D. M., & Wu, F. (2009). Social networks that matter: Twitter under the microscope. *First Monday* 14(1).
- Kelly, J., & Etling, B. (2008). *Mapping Iran's online public: Politics and culture in the Persian blogosphere*. Berkman Center for Internet & Society: Harvard Law School. Retrieved from http://cyber.law.harvard.edu/sites/cyber.law.harvard.edu/files/Kelly&Etling_Mapping_Irans_Online_Public_2008.pdf
- Kirchhoff, L., Nicolai, T., Bruns, A., & Highfield, T. (2009). Monitoring the Australian Blogosphere through the 2007 Australian Federal Election. In T. Flew, *Communication, Creativity and Global Citizenship: Refereed Proceedings of the Australian and New Zealand Communications Association Annual Conference, Brisbane, July 8-10* (pp. 982-1005). Brisbane. Retrieved from <http://www.proceedings.anzca09.org>
- Lerman, K., & Ghosh, R. (2010). Information Contagion: an Empirical Study of the Spread of News on Digg and Twitter Social Networks. In *Proceedings of 4th International Conference on Weblogs and Social Media (ICWSM-10)*. Retrieved from <http://arxiv.org/abs/1003.2664>
- Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A., & Christakis, N. (2008). Tastes, ties, and time: A new social network dataset using Facebook.com. *Social Networks*, 30(4), 330-342.
- Lovink, G. (2008). *Zero Comments: Blogging and Critical Internet Culture*. New York: Routledge.
- Park, H. W., & Thelwall, M. (2008). Developing network indicators for ideological landscapes from the political blogosphere in South Korea. *Journal of Computer-Mediated Communication*, 13(4), 856-879.
- RTGI. (2007). *Blogopole [Observatoire Présidentielle 2007]*. Retrieved from <http://blogopole.fr/>
- Rogers, R. (2009). *The End of the Virtual: Digital Methods*. Amsterdam: Vossiuspers UvA.

Thelwall, M. (2009). *Introduction to Webometrics: Quantitative Web Research for the Social Sciences*. San Rafael, CA: Morgan & Claypool Publishers.

Tremayne, M. (2007). *Blogging, Citizenship, and the Future of Media*. New York, NY: Routledge.

Vaccari, C. (2008). Surfing to the Élysée: The Internet in the 2007 French Elections. *French Politics*, 6(1), 1-22.

Vanbreemsch, N. (2009). *De la démocratie numérique*. Paris: Seuil: Médiathèque.

Watts, D. J. (2004). *Six degrees: The science of a connected age*. New York, NY: W.W. Norton.

Yanoshevsky, G. (2009). L'usage des vidéoblogs dans l'élection présidentielle de 2007. Vers une image plurigérée des candidats. *Mots. Les langages du politique*, (89), 57-68.

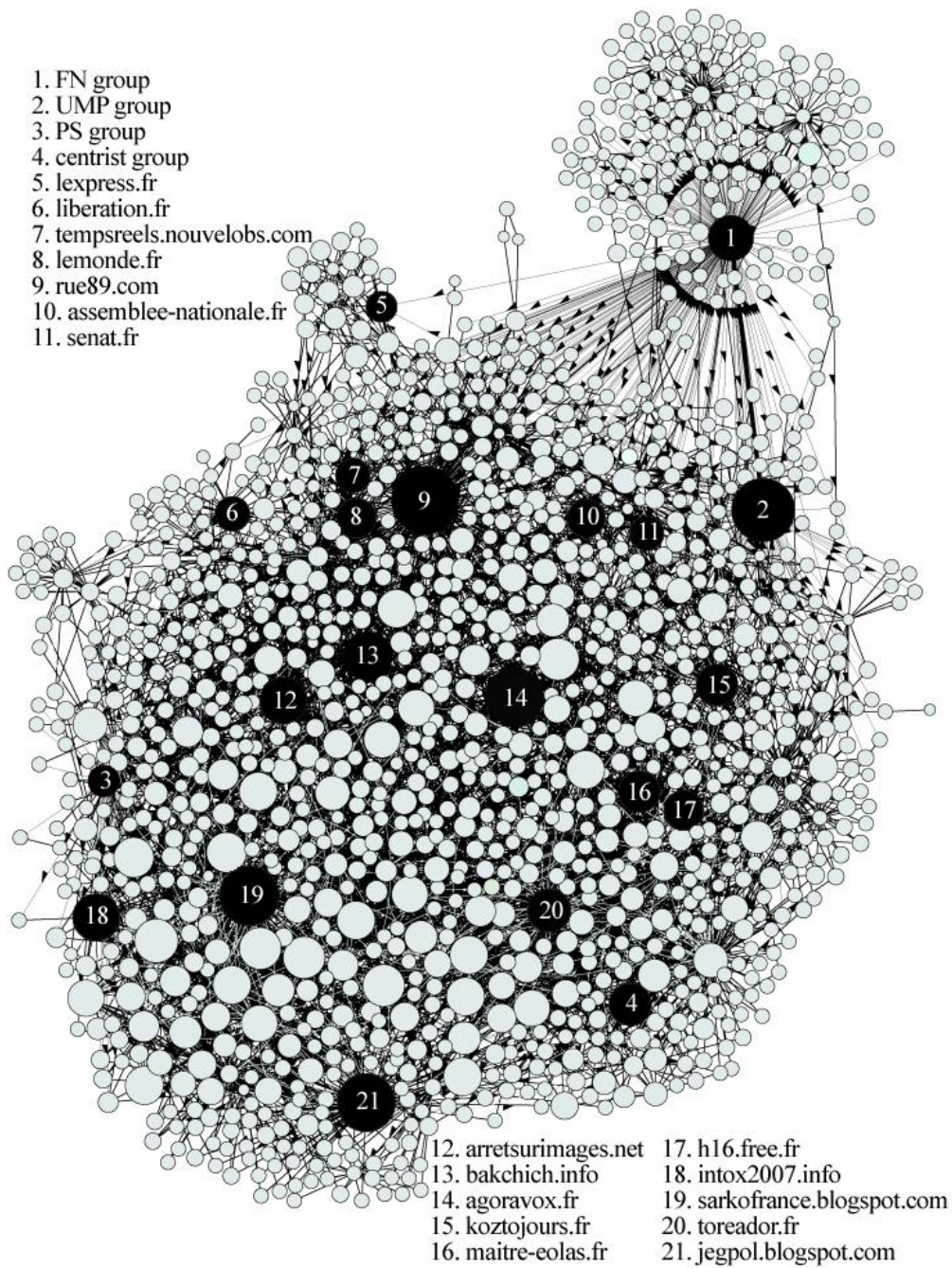


Figure 1: Network formed by out-links from blogrolls on the 217 sites in the sample (limited to sites receiving two or more in-links)

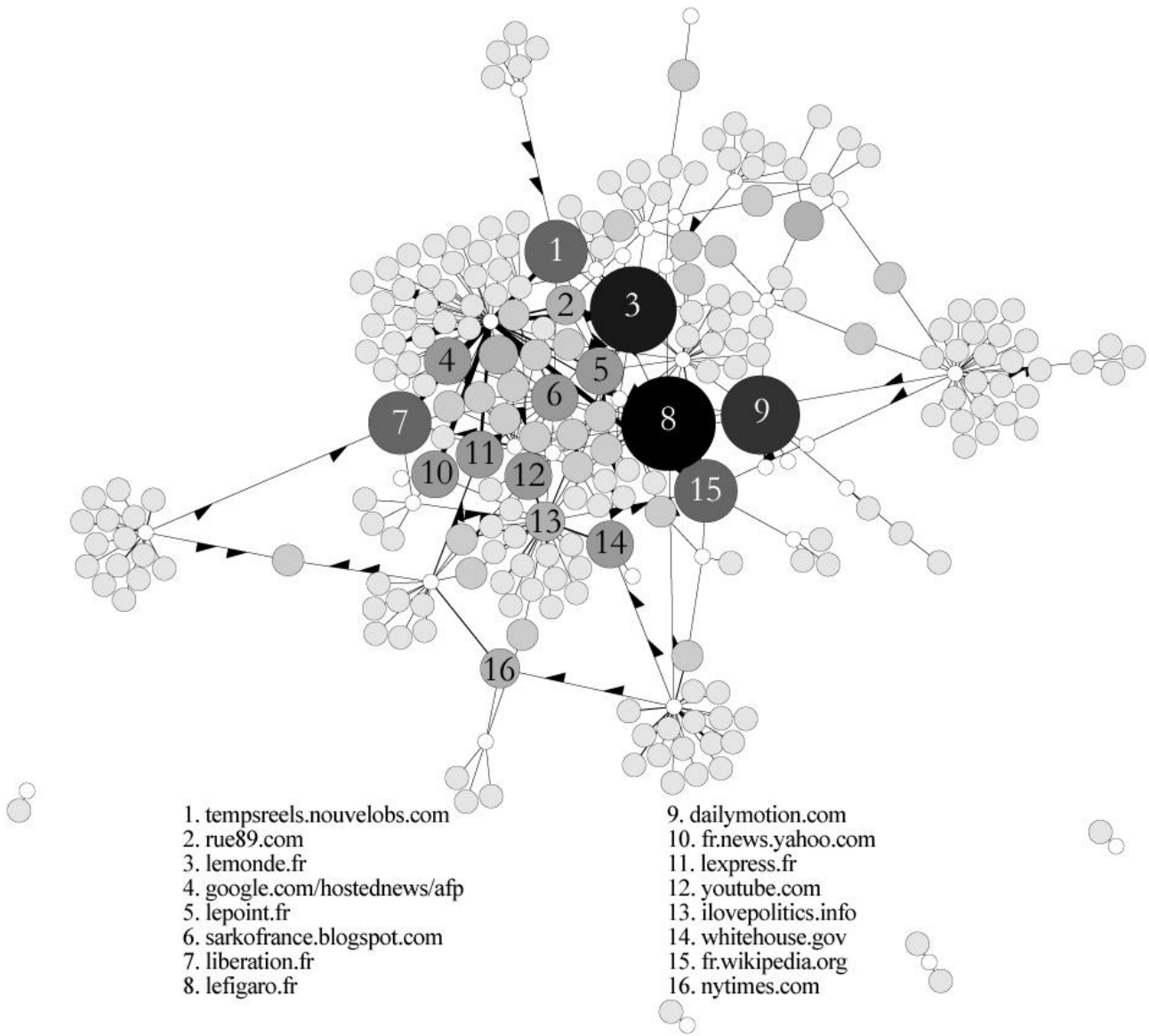


Figure 2: Topical network formed from links made in blog posts mentioning Obama