



UvA-DARE (Digital Academic Repository)

Where do bloggers blog? Platform transitions within the historical Dutch blogosphere

Weltevrede, E.J.T.; Helmond, A.

Published in:
First Monday

DOI:
[10.5210/fm.v17i2.3775](https://doi.org/10.5210/fm.v17i2.3775)

[Link to publication](#)

Citation for published version (APA):

Weltevrede, E., & Helmond, A. (2012). Where do bloggers blog? Platform transitions within the historical Dutch blogosphere. *First Monday*, 17(2). DOI: 10.5210/fm.v17i2.3775

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <http://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.



Where do bloggers blog?
Platform transitions within the
historical Dutch blogosphere

by Esther Weltevrede
and Anne Helmond

Abstract

The blogosphere has played an instrumental role in the transition and the evolution of linking technologies and practices. This research traces and maps historical changes in the Dutch blogosphere and the interconnections between blogs, which — traditionally considered — turn a set of blogs into a blogosphere. This paper will discuss the definition of the blogosphere by asking who the actors are which make up the blogosphere through its interconnections. This research aims to repurpose the Wayback Machine so as to trace and map transitions in linking technologies and practices in the blogosphere over time by means of digital methods and custom software. We are then able to create yearly network visualizations of the historical Dutch blogosphere (1999–2009). This approach allows us to study the emergence and decline of blog platforms and social media platforms within the blogosphere and it also allows us to investigate local blog cultures.

Contents

[Introduction](#)
[Characterizing Dutch blogs: Where do bloggers blog?](#)
[Reconstructing the blogosphere](#)
[Defining the actors](#)
[The Dutch blogosphere in transition](#)
[Concluding discussion](#)

Introduction

The blogosphere has played an instrumental role in the transition and evolution of linking technologies and practices, such as the introduction and development of the trackback, pingback and RSS and their use by bloggers to develop a culture of blogging as a distinct online culture. Important research in this area has been practice, event or issue based, trying to capture an otherwise fleeting phenomenon in realtime, before it is deleted, overwritten or no longer available. Now that the blogosphere has reached maturity, the first historical accounts are being created. This study seeks to contribute to this body of literature by investigating the blogosphere's more structural platform and software infrastructure. More specifically, we seek to contribute to the empirical research of the national blogosphere by proposing new methods to explore transitions in the historical Dutch blogosphere. In addition, we will consider the implications of these methods when thinking about transitions in blogs and the local blogosphere.

Working from method, our contribution will be theoretical, as choices in method shape the definition of blogs and the blogosphere. This paper addresses methodological questions related to the empirical research of the national historical blogosphere, and present the outcome of preliminary research into the Dutch blogosphere [1]. The proposed approach combines techniques that are also used by search engines and Web archive crawlers — to discover and analyze the content of the entire Web — with editorial techniques commonly used in

Reading Tools

Where do bloggers...

Weltevrede, Helmond

[Abstract](#)

[Review policy](#)

[About the author](#)

[How to cite item](#)

[Indexing metadata](#)

[Supplementary files](#)

[Print version](#)

[Notify colleague*](#)

[Email the author*](#)

RELATED ITEMS

[Author's work](#)

[Government policy](#)

[Book](#)

[Book reviews](#)

[Dissertations](#)

[Online forums](#)

[Quotations](#)

[Resources](#)

[Media reports](#)

[Web search](#)

SEARCH JOURNAL

All

 CLOSE

* Requires [registration](#)

humanities and social sciences. While preliminary, the research both acts as a proof of concept and as a model for studying national and historical blogospheres, and provides new insights into the shape of the Dutch blogosphere and its interconnections.

The blogosphere is often studied by mapping and visualizing the interconnections between blogs, in order to make the blogosphere tangible and visible. In other words, to become visible, the image of the blogosphere must be constructed, either by blogosphere related services such as directories, Web rings and blog search or by academic network visualizations. By means of similar techniques such as contemporary blog related services, network visualizations may be constructed by employing RSS feed-crawlers to fetch the content — current and newly updated blog posts and their links — of blogs using their feeds (Bross, *et al.*, 2010) or by using Web crawlers for network analysis. Crawling and network analysis may be used for widely varying analytical purposes, such as the IssueCrawler for issue network analysis to track conversation patterns in the blogosphere (Bruns, 2007); crawling the front page of blogs to reflect blogroll communities (Adamic and Glance, 2005); and, large scale grouping of linked blogs to define clusters of shared informational worlds (Kelly and Etling, 2008). Although different tools and methods produce different network visualizations, they all provide graphical representations of interconnections and insights into the overall structure of the blogosphere and its actors (Highfield, 2009). We uphold that choices in method do not only shape the blogosphere but also shape the definition of blogosphere and blogs.

Historical blogosphere research mainly consists of ethnographic research, providing personal stories and anecdotes (Blood, 2004; Rosenberg, 2009), in addition to empirical work, such as Michael Stevenson's (2010) research on the early A-list blogosphere, Rudolf Ammann's (2009) research project on the birth of the blogosphere and Ravi Kumar, *et al.* (2004) researching the structure and evolution of LiveJournal blog space. Kumar, *et al.* suggest a method based on time stamps — in addition to other features such as those present in profile pages — to map a blog space over time. More generally, the Internet Archive allows the study of previous states of the Web by providing time-stamped snapshots. Although the single-site history is preferred — as only single URLs can be retrieved — Internet Archive data may be used in a variety of ways. Ammann studies the emerging blogosphere by mapping linking patterns of early blogs on the basis of the Internet Archive and Stevenson outlines a method to re-purpose the Internet Archive to create a custom archive by using the early blog index EatonWeb as a historical resource to recreate the blogosphere. Our research builds on the above-mentioned methods and tools and has developed a number of novel techniques and methods.

First of all we will investigate the historical blogosphere by making snapshots of the Dutch blogosphere with specific attention to actor definitions and interlinking practices by introducing fine-grained URL and source code analysis. After that we seek to widen the historical blogosphere analysis beyond the Anglo-American context by specifically focusing on the Dutch blogosphere. A contribution to the definition of a "national blogosphere" is proposed by investigating the Dutch character of top level domain, blog software and platform use. Finally, we aim to contribute to hyperlink network and issue network analysis research by redefining what is considered an actor in the blogosphere.



Characterizing Dutch blogs: Where do bloggers blog?

How to formally define the nationality of an online site? The Web archiving community often tried to answer this question with locative technical indicators such as the IP address or top level domain (TLD). Indicators for location on the Web, however, are always ambiguous and their usefulness highly depends on the purpose and application. For the purpose of saving digital heritage for future posterity, the Dutch Web archiving institution formulated three defining characteristics, including language, TLD, and subject matter "aimed at the Netherlands" which is rather difficult to automate (Weltevrede, 2009). When defining a Dutch blog in this research project, in first instance we rely on authoritative sources, with their selection criteria for including blogs in their lists. In a second step, the question: "what is a Dutch blog?" evolves into "where do Dutch bloggers blog?" in order to enrich and refine the understanding of the location of Web content.

The collection of blogs in our corpus is retrieved from a 2001 database dump — containing 631 unique blogs — from *Loglijst*, an early Dutch blogosphere indexing initiative. In addition, we compiled expert lists from interviews, books and authoritative lists found on the Web and in the Internet Archive. These experts lists include long list nominations for the Dutch blog awards, the Dutch Bloggies from 2001–2008, all blogs mentioned in two seminal pieces on the history of the Dutch blogosphere by Dutch blogosphere historians Frank Schaap (2005) and Frank Meeuwse (2010) and finally a list citing "Weblogs that really matter" in a December 2010 blog post by Bert Brussen, blogger for the famous Dutch 'shocklog' *Geenstijl* [2]. Relying on these sources to provide us with a collection of Dutch blogs led us to include a small number of Belgian (Dutch language) blogs that our sources considered to be part of the Dutch

blogosphere [3].

We queried the Internet Archive's new Wayback Machine for each blog's URL and selected the result dated closest to the middle of each year under investigation. From the 2,500 URLs requested we were able to retrieve just under 1,000 blogs from the Internet Archive. This method yielded a collection of archived copies of historical Dutch blogs for each year with a timestamp near the middle of the year. Only blogs with a copy in the Internet Archive were retained for further analysis. The following table represents the number of blogs per year serving as starting points:

Table 1: Starting points retrieved from the Internet Archive per year.

1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
24	138	456	816	778	863	850	788	717	860	723

Blog software features have previously been studied in relation to popularity and the success of Weblogs (Du and Wagner, 2006) and in relation to blogging practices which enable or restrict certain actions (Schmidt, 2007). In addition, we wanted to interrogate our starting points to look into the question 'where' bloggers blog by analyzing the TLDs, platforms and the software used. As far as we know this question has so far been understudied and, if at all addressed, the demographics of bloggers have been analyzed, such as the location provided in user profiles of blog platforms (Kumar, *et al.*, 2004). This information, however, is optional and limited to some platforms only. Our study focuses on online culture — national digital culture — investigating a further specificity in Dutch online practice, *i.e.*, software and platform usage as well as applications that persist, despite Twitter, Technorati and other dominant services from the U.S. When describing Dutch blogging practices we therefore pay specific attention to 'where' Dutch bloggers blog. In this study the question of 'where' is three-fold and includes the analysis of TLDs, platforms and self-hosted blog software.

TLD analysis

The top level domain (TLD) analysis presented here is part of a larger series of URL analysis methods discussed in this paper. As a first step we counted the TLDs of our starting points per year by entering URLs in batches corresponding to a single year using the TLD Count tool [4]. Figure 1 shows the relative distribution of TLD usage over time. The Dutch blogs in our collection favor the .nl domain over all other domains throughout the years. Moreover, a significant increase in the .nl domain becomes apparent, whereas the .com domain is steadily losing share over time; the results in the next section show that Dutch bloggers move away from .com blogging platforms such as Blogger's Blogspot to go to Dutch .nl blogging platforms.

The Dutch .nl domain is one of the top five country code top level domains (ccTLDs) in the world [5], which is also reflected in the Dutch blogs. It is however remarkable that the .nl domain has been dominant from the beginning, since the .nl domain only became available to private individuals since 2003. As a forerunner, since 2000, individuals were allowed to register third-level domains such as jansen.123.nl [6] but these domains were rather rare and are absent from our collection of blogs. As stated before, the Dutch blog collection contains a number of .be blogs, steady from 2000 onwards. Furthermore, 2002 presents a peak of .tk. Dot.tk "Renaming the Internet" offers free domain names and includes URL redirection and forwarding services. Lastly, a number of domains is unconventionally used for "commercial or vanity" purposes, including .nu (country code for Niue), marketed as 'now' in Dutch and .is (country code for Iceland), which is used as the verb 'to be' [7].

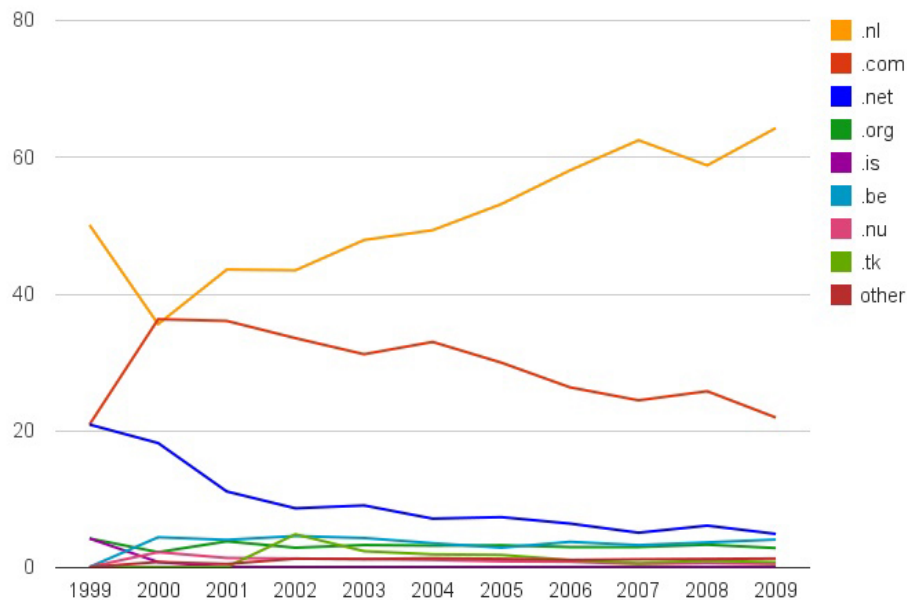


Figure 1: Relative distribution of Top Level Domains (TLDs) in the Dutch blogs over time.

Platform analysis

A second way to answer the question ‘Where do bloggers blog?’, complementing the TLD analysis, is by visualizing the variety and proportion of blog platforms used in the Dutch blogosphere. This requires basic background knowledge of blog platforms. With the use of Google Refine, “a power tool for working with messy data” [8] we ‘coded’ each of the blog platforms in GREL (Google Refine Expression Language) to automatically search, transform and count the platforms in our set of URLs. The results are presented in Figure 2, a custom made visualization combining the blog platform analysis with the self-hosted software analysis as discussed in the next section.

The graph shows the rise and popularity of Blogger’s platform, Blogspot, in the beginning of 2000. The decline of Blogspot coincides with the rise of the Web-Log.nl blogging platform, and other Dutch blog platforms such as BlogNL, Bloggo, Blogse, Punt and Blogeland. Figure 2 clearly shows how from 2004–2005 onwards Dutch bloggers — except for a relatively small number of Blogspot and WordPress.com users — shift to Dutch platforms, which are orange color-coded. Only a few bloggers remain on legacy platforms such as Pitas, which no longer accept new members but are still functional for old members.

Dutch software and platforms play an important role in the Dutch blogosphere and between 2004 and 2009 over 40 percent of all bloggers use Dutch blog software or Dutch blog platforms. When zooming into the use of platforms only, in 2009 almost all bloggers on blog platforms make use of Dutch platforms (see Figure 3) [9].



Figure 2: Relative distribution of self-hosted blog software & blog platforms in Dutch blogs.

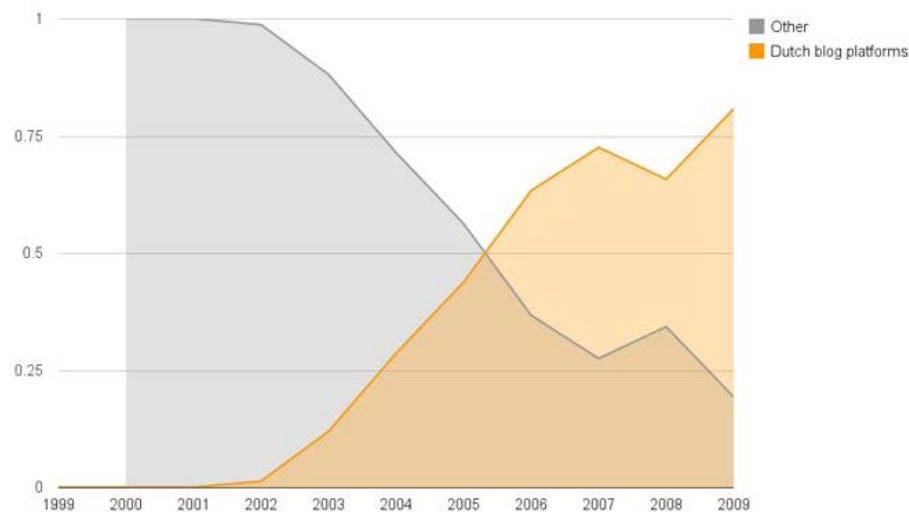


Figure 3: The relative amount of Dutch blog platforms over time compared to other blog platforms.

Self-hosted software analysis

URLs were analyzed to investigate the distribution of TLDs and platforms used in the Dutch blogosphere. The outcome suggests that the early Dutch bloggers did not use blog platforms. In general, they preferred to manually create their blogs, written in HTML or they used specifically designed self-hosted blog software. In HTML, the reverse-chronology, which is considered to be a key characteristic of blogs (Blood, 2004; boyd, 2006), had to be manually enforced in order to place the latest blog post on top. In order to include these kinds of blogs in our analysis we developed a method going beyond the blog's URL. We searched within the page's source code to look for the URL referencing the software powering the blog to create an accurate list of blog software.

Initially, the list was compiled by analyzing maps (see Section 4) and then refined with newly discovered blog software throughout the research project. To compile the list of self-hosting software, we used the reflexivity of bloggers. Typically, bloggers tend to analyze and describe the practice of blogging (Hourihan, 2002; Blood, 2002). When researching our initial list of software, we found blog posts comparing or mentioning different types of software (for an example, see Figure 4). For each year we searched the source code of the collection of archived blog front pages for the presence of the blog software types with the Source Code Search tool [10]. The results were editorially checked to establish whether the reference to the software implied that the blog was indeed running on it. Especially in the beginning, references to self-hosted blog software were not standardized. In later years the 'powered by' button in the side bar or footer became standard for most self-hosting software.



Figure 4: "Not tonight love/I'm busy playing around with Weblog software".

Contrary to the blog platform counts, the self-hosted blog software results suggest that the Dutch blog software Pivot/PivotX has been powering Dutch blogs from the start; it appears to have been the most frequently used software in the heydays of Dutch blogging. The decline of Blogger, the first blog platform used by Dutch bloggers, coincides with the rise of Blogspot — Blogger's platform. Furthermore, the bar graph shows a boost of blogs powered by WordPress.org in the blogosphere from 2006 onwards. Movable Type and the Belgian Nucleus have a small but loyal share of bloggers running the software.

In terms of blog software and blog platforms, the peak of Dutch blogs was around 2005 for platforms and 2006 for software. Notably, the share of self-hosted software exceeds one-click publishing platforms, which even the bloggers themselves had not expected. A number of posts from early bloggers express fear that soon everybody will be blogging; some others voice rivalry between self-hosting bloggers and platform bloggers (for an example, see [Figure 5](#)). A next step in further developing this methodology is to formalize the various types of references to software, throughout the years, and design queries to automate the collection and analysis process of the results.

Ach, de gratis weblogdiensten. Hoe wij, de webloggers
van het eerste uur, met een eigen domein en een
zelfgeknutselde site ze verachtten, de diensten waar
je met een paar klikken een weblog op kon zetten.
Kijk hem, hij heeft een blogspot, of erger nog, een
web-log.nl, wat wij smalend web-streepje-log
noemden.

Figure 5: Early Dutch blogger about the rise of free blog platforms [11].

In this first part we focused on designing methods to address the question ‘Where bloggers blog?’ in order to enrich current methods determining the nationality of blogs by complementing a TLD analysis with a platform and software analysis. In the following part we will look into the interconnections between these blogs and propose a method to create historical blogospheres.



Reconstructing the blogosphere

In 1999, Brad L. Graham coined the term ‘blogosphere’ to mark the end of cyberspace: “Goodbye, cyberspace! Hello, blogiverse! Blogosphere? Blogmos?” (1999); William Quick revived the word in 2001 as “the intellectual cyberspace we bloggers occupy”, explicitly mentioning that the blogosphere is a space for serious discourse. Echoing the idea of the blogosphere as a discursive space “the imagined public sphere” (boyd, 2006) was presented alongside the idea of blogs as a reaction to mainstream media (Lovink, 2008). Besides the notion of the blogosphere as a space for discourse, other definitions stress the formalistic characteristics of the blogosphere as an interlinked set of blogs which “allows for the networked, decentralised, distributed discussion and deliberation on a wide range of topics” (Bruns, *et al.*, 2009). A complimentary approach to the blogosphere as an interlinked set of blogs looks at how blogs are “embedded into a much bigger picture: a segmented and independent public that dynamically evolves and functions according to its own rules and with ever-changing protagonists, a network also known as the ‘blogosphere’” (Bross, *et al.*, 2010). Following this line of thinking, *i.e.*, blogs which are embedded in a larger networked ecology with shifting protagonists, the blogosphere may also be defined by including the actors they link to in their networked ecology: “The notion of a mini-blogosphere additionally rests on the extent to which the set of blogs doing an issue are interconnected by links and/or by textual referencing. Blogs also make [sic] be ‘connected’ together through common references to a third party, *e.g.*, all blogs linking to or referencing a particular piece in the *New York Times*” (Rogers, 2005). Although these two dominant approaches to research the blogosphere have different objects of research, they do not exclude each other, as demonstrated by Benkler and Shaw’s (2010) U.S. political blogosphere research.

Although highly formal, the blogosphere has more of a cultural than a technical meaning, because as illustrated in the previous section, the many different blog platforms and software types permit the blog’s custom use. At first glance our approach might appear formalistic, because the definition of the blogosphere follows from the outlined method based on link analysis, see below. However, by mapping the formal changes in linking patterns and URLs over time, we can propose results about specific local cultures of use.

The annual blogospheres were created from a collection of blogs retrieved from the Internet Archive by means of custom tools. One of the consequences of studying transition with the Internet Archive is that only research on front–page level and not on a post level is possible. Hence this method may be viewed as a more structural ‘blogosphere’ analysis instead of an ‘issue’ or ‘event’ analysis. Although fully aware that our choice of starting points shapes the Dutch blogosphere, the methodology used only retains blogs deemed relevant by other blogs. It is a co–link analysis as used by the IssueCrawler [12]. This co–link analysis is performed in two steps: first, for each blog all links on front–page level are extracted (one depth) and subsequently, in Gephi, only nodes receiving at least two links from the starting points are maintained in the network visualization (one iteration). The resulting network map thus retains only co–linked actors, those receiving at least two links from the starting points. This implies that the starting points themselves might drop off the map and that Dutch blogs, which are not available in the Internet Archive, might reappear in the network. It thus also acts as a validation of our expert lists.

Whereas the co–link analysis is an analysis module most successful for locating issue networks, in our case, the result of the co–link analysis is that issue or event–based links are excluded from analysis. This has three main reasons. First, the starting points are not chosen because they share an issue or an interest in an event, but rather the practice of blogging in the Dutch Web space. Second, only front–pages are crawled, which means that the more structural links are followed, such as links in blog rolls, and links to blog related services and blog software. In other words, these links are the stable variable in the analysis, whereas links in posts are only taken into account if present on the front–page. Third, the time frame of each network is one year. Combined with the previous point, that links from posts are only crawled one level deep, it results in links to versatile issues dominating the Dutch blogosphere for a short period of time being excluded, only the more structural issues prevailing. When studying a structural blogosphere blogs are assumed to be embedded in a larger networked ecology created by bloggers through their linking practices, including other actors than blogs, such as blog portals, Web rings, news Web sites and social media platforms.

In what follows, we will further describe how we constructed the Dutch blogosphere on the basis of the Internet Archive and prepared it for further analysis. Specific attention will be paid to the construction process by reconfiguring actor definitions and reconsidering interlinking practices. This approach gives us novel insights into the composition of the blogosphere and its actors. We have further developed methods to study transitions in the historical blogosphere with the Internet Archive. Our method consists of two strands: first we refine the network analysis by defining the actors using Gephi and G–Atlas software [13], and then we complement the network analysis by color–coding the platforms present in the blogosphere.



Defining the actors

As previously described, we retrieved snapshots of our blogs from 1999 to 2009 from the Internet Archive, extracting their outlinks on a front–page level and putting the results in Gephi’s GEXF format. In Gephi, a simplified version of IssueCrawler’s co–link analysis is performed so that only blogs with more than two links from our starting list are retained. Co–link is performed on a ‘by site’ level, as it is more indulgent than the ‘by page’ option because it counts all links from site to site. In other words, co–link analysis is performed on the hosts and not on the deep pages.

A common problem in online network visualizations is that big platform nodes take a prominent position in the graph. Analysis of these maps often suggests that the debate is moving elsewhere (*i.e.*, to social media). In an attempt to untangle the big platform nodes in the Dutch blogosphere, we propose to redefine the nodes of the network to actors. Most network analysis software treats the host and in some cases sub–host as the actor. However, in our case the ‘actor’ or blogger is often defined after the slash, like the early bloggers that started blogging from their personal homepage or the recent microbloggers on Twitter. A similar approach is being developed by Medialab Science Po who have defined the concept of ‘Web entities’ to unravel pages grouped by domain name (Girard, 2011). Also Benkler and Shaw (2010), in their work on the U.S. political blogosphere, stress the importance to analyze what is inside the large network nodes in order to specify their internal differences.

To identify nodes in the blogosphere as actors, we redefined ‘actors’ on a URL level. This did require an additional analysis step because not all URLs follow the same pattern. With most Web sites ‘actor’ equals ‘host’ (*e.g.*, example.com) while actors on blog software usually are defined before the host on a subdomain (*e.g.*, example.blogger.com), actors on personal homepages are often defined by their – after the slash (*e.g.*, xs4all.nl/~example) just like microbloggers on Twitter (*e.g.*, twitter.com/example). In the actor definition project we sought to formalize ‘URL patterns’ in the network.

The Dutch blogosphere in transition

Mapping the outlinks of the blogs we retrieved from the Internet Archive from 1999 until 2009 allows us to go back in time and study how and where the Dutch blogosphere originated. Using the fine-grained actor definition, the network is visualized with Gephi for each year. [Figure 6](#) shows the rise, evolution and first signs of decline of the Dutch blogosphere, grey depicting the hyperlink network of all years together and red the blogosphere of a particular year. The first Dutch bloggers starting mid 1999 were not interlinked into a 'sphere', so we can trace back the beginning of a structural Dutch blogosphere to 2000.

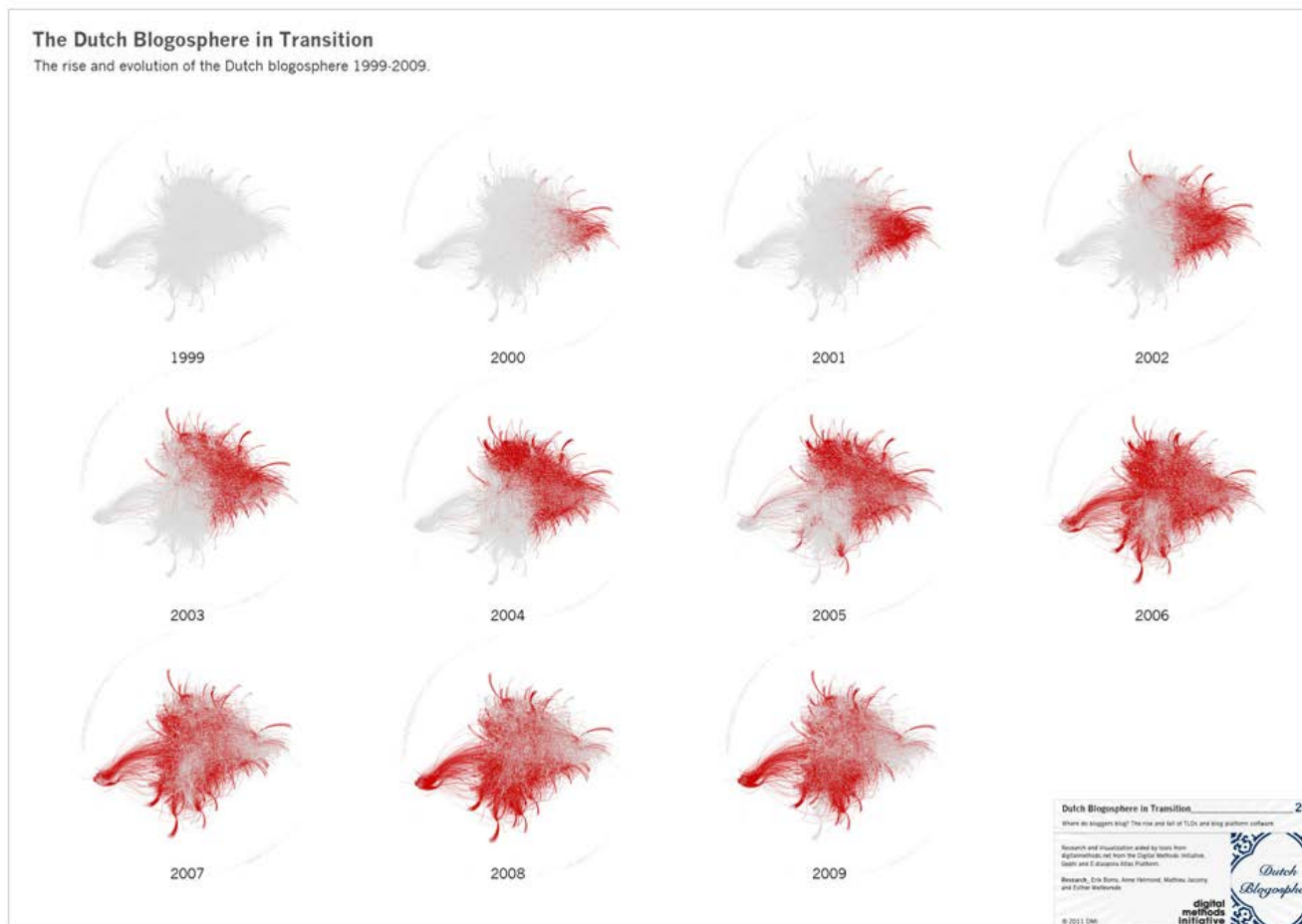


Figure 6: The Dutch blogosphere in transition.

In 1999 the map (not displayed) only shows four nodes, not linking to each other but present because they receive at least two links from our selected starting points. The four nodes are Nedstat, Nedstatbasic, *Wired* and a Dutch blog by Wessel Zweers, a.k.a. -wzweers. A familiar node on the map is *Wired*, a technology magazine also prominent in the American early blogosphere. The only Dutch blogger on the map is hosted on one of the oldest Dutch hosting services providing free personal homepages, "De Digitale Stad" (DDS, Digital City). Well-known Dutch blogs from that period, like Sikkema, Prolific and Alt0169 are notably absent because they do not receive two links from our starting list's blogs. [Figure 7](#) shows that some

of the well-known Dutch bloggers, as mentioned in Meeuwse (2010), together with less well-known bloggers, are present but do not form a blogosphere yet. Most notably Alt0169, ~wzweers and ~onnoz reach out to other Dutch blogs and may be seen as an effort to establish a community between blogs. Exemplary are links to blogs that list blogs, like <http://beboo.org/metalog>, listing the top 50 (international) blogs.

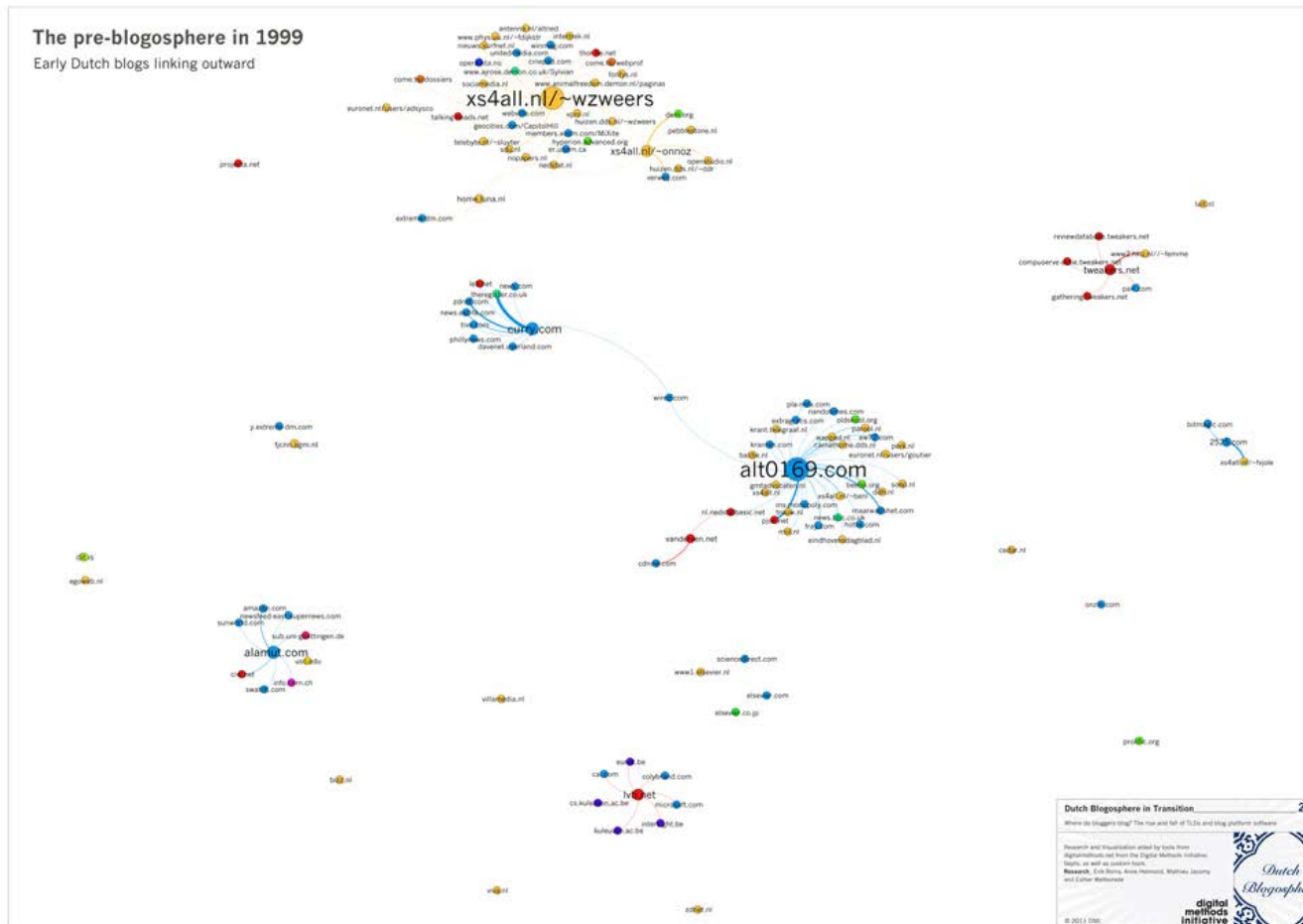


Figure 7: The pre-blogosphere in 1999. Early blogs linking outward.

Cluster analysis over time

In 2000 the Dutch blogosphere is dominated for the first time (see Figure 8) by bloggers on personal homepage providers (blue) and student pages (pink). The right side of the blogosphere shows a cluster of Dutch homepages (~) and student homepages. The free homepage provider DDS and Dutch Internet service provider XS4ALL are the most prominent providers. The larger nodes in the center are the founding blogs of the Dutch blogosphere, such as Alt0169, Sikkema, S-Ir, Smoel, Rikmulder, Tonie, Prolific, Pjoe, Stronk, Ben Bender, Vandenb, Retecool. They are actually a closely linked cluster. Alt0169.com, a heavy linker in 1999 but without receiving any links back, is a central node in 2000. Figure 9 shows the Dutch marketing cluster, which emerged in 2005 and still a very dominant cluster in the Dutch blogosphere. Another distinct cluster in the later blogosphere is the Blog.nl cluster. Blog.nl has a very distinct shape because all Blog.nl blogs list and link the other blogs on that platform as can be seen on the right in Figure 11.

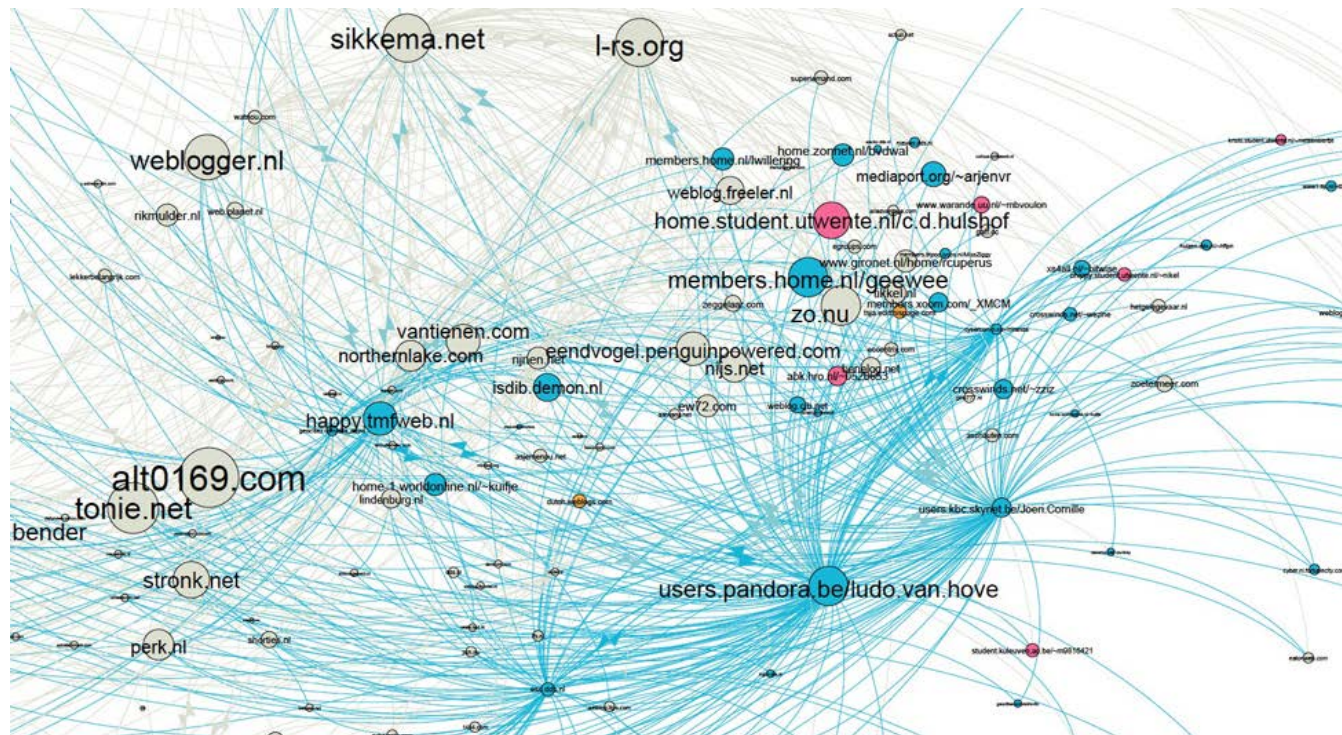


Figure 8: The Dutch blogosphere in 2000. Note: Blue: personal homepages; Pink: student pages; Yellow: blog platforms.

(Baym, 2007), or the distributed self across social media platforms (Helmond, 2010). Whereas in the mid and late 1990s the self was defined on the personal homepage and later on the blog, nowadays the self is also defined and performed on social networking sites and content platforms. Blog software popularized the creation of the widgetized self with its easy drag and drop widgets that allowed bloggers to easily embed content from their other platforms into their blog via the sidebar. The sidebar is no longer only used to link to other bloggers, using the blogroll, but also to link to the self on other platforms such as Last.fm for music, Flickr for photos and YouTube for videos. As our method collects outlinks from the front–page and subsequently performs a co–link analysis, the widgetized self in the sidebar on the front page can be captured as a new actor in the blogosphere. In traditional hyperlink analysis social media nodes are disproportionately large as all references are collapsed into one node.

Comparing the 2009 blogosphere *with* and *without* our custom actor definition (Figure 11), it becomes apparent that the social media platforms privilege a more fine–grained analysis. Social media are the big nodes in the network without actor definition; however, with actor definition the social media platforms seem to lose prominence in the blogosphere.

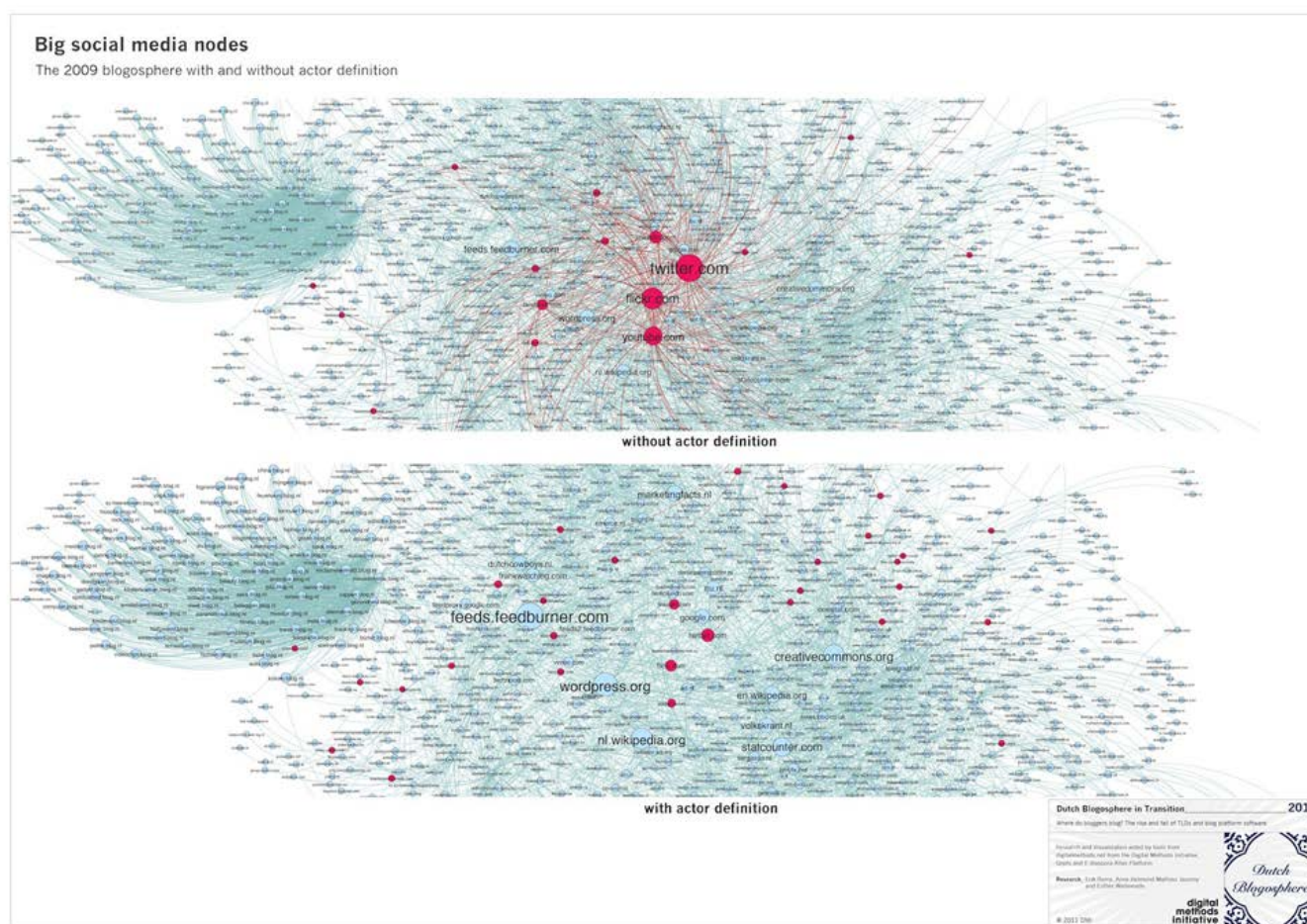


Figure 11: Big social media nodes. The 2009 blogosphere with and without actor definition. Note: Social media platform nodes are highlighted in magenta.

The question then arises what do people link to in social media: to user pages or to content (e.g., video, photo, status update)? Figure 12 shows the large social media platform nodes, containing smaller nodes. Comparing the various social media platforms, the results suggest that some platforms can be defined as 'media sharing' platforms, such as YouTube and Flickr,

which mainly consist of embedded content links in blogs. In the blogosphere map with actor definition, these nodes decrease in size. Facebook is a relatively small node in the Dutch blogosphere and the links it receives dissolve into a diverse set of profiles, pages, apps, events, and groups. Hyves — the Dutch social network that, arguably, still outnumbers Facebook in the Netherlands [14] — is one of the smallest social media references. Although the Dutch blogosphere prefers Dutch software and platforms, this is not reflected in social media platform links. Twitter, the largest node in the network is a platform mainly receiving links to user pages. This means that bloggers refer to themselves or to friends on the micro-blogging platform.

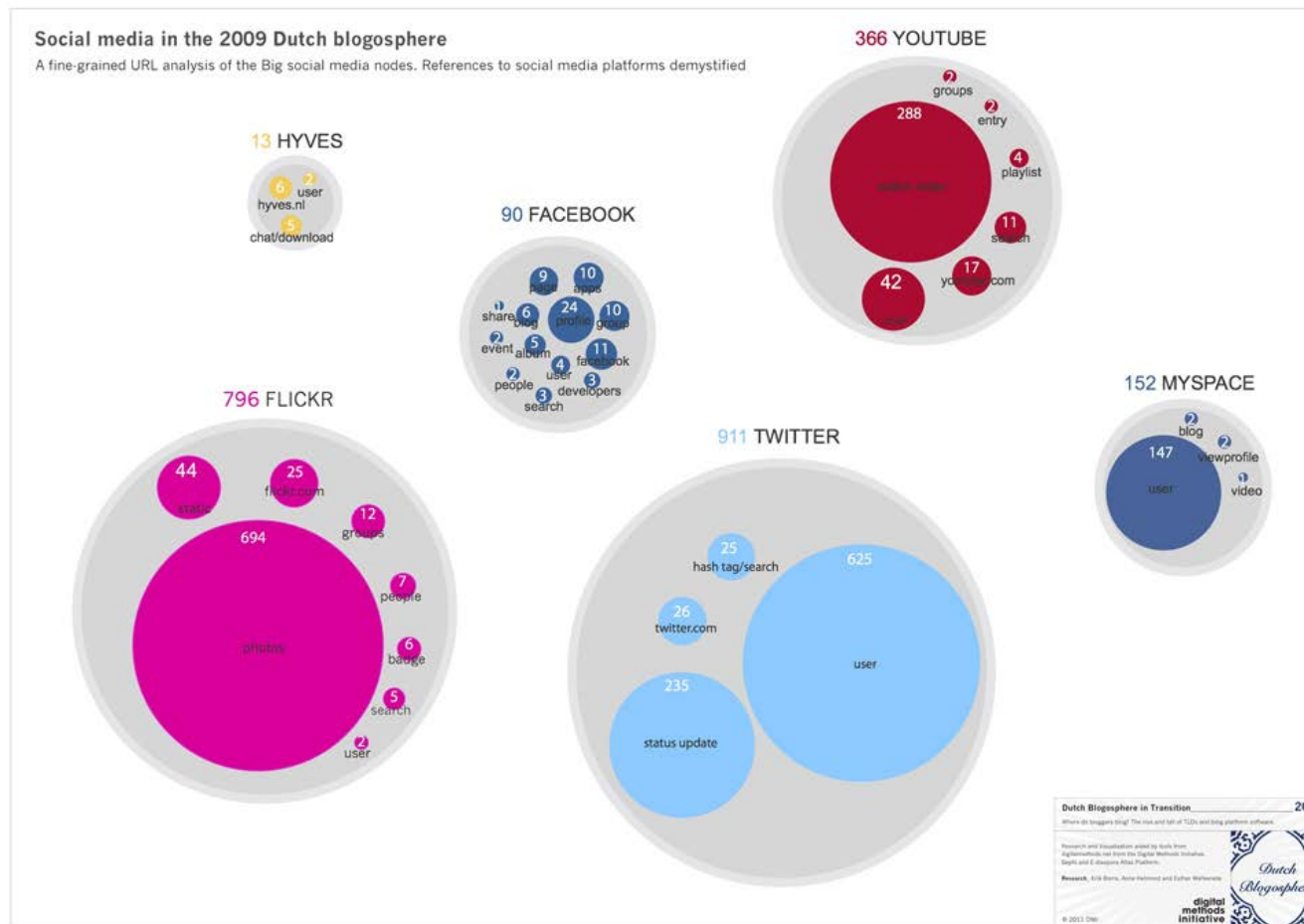


Figure 12: Social media in the 2009 Dutch blogosphere. A fine-grained URL analysis of the big social media nodes. References to social media platforms untangled.

Traditional link analysis has its limitations when analyzing the share of social media in blogosphere networks. Our study suggests that the uniform large platform nodes are misleading. We found that link analysis zooms out to look at platforms as a whole and treats the entire platform domain as the node; in doing so the individual content link and the individual author disappear. The platform nodes require a more nuanced exploration.



Concluding discussion

The results of our research indicate that the methods proposed for investigating the national historical blogosphere appear to be overall useful. This paper aimed to contribute to the growing body of literature on blogs and the blogosphere by proposing new methods to empirically investigate transitions in the historical blogosphere over time. Hence a method was developed and described to create a so-called structural blogosphere on the basis of the medium specific characteristics of the Internet Archive, allowing for the re-construction of a blogosphere on domain level and not on post level. The advantage of this method is that it allows for a 'structural' blogosphere analysis instead of an 'issue' or 'event' analysis. With a structural analysis software and platform analysis may be performed; in this case it was used as a new way to study the nationality of blogs. We sought to contribute to the 'where' question of Web content by asking "where do Dutch bloggers blog?" and looked into TLD, platform and software usage. Our study suggests that Dutch bloggers increasingly blog on in the .nl space despite the more general trend of software concentration and domination of actors like Blogger and WordPress.

This paper further developed three analytical techniques and methods to study the national historical blogosphere: URL analysis, source code analysis and hyperlink analysis. URLs are very rich sources of information often following a certain syntax, which makes them very suitable for analysis. Here we used URL analysis in two ways: TLD analysis and platform analysis. With source code analysis we contribute to the study of software in general and, more specifically, the study of national software. The method developed provides insight into the software powering a blogosphere. Further research may include a fine-grained feature analysis over time, placing special emphasis on collaborative and discursive features such as the comment, plugins and the permalink. Our contribution to link analysis considers ways to treat the large platform nodes in network visualizations. We propose two methods, the first being an actor definition and the second a fine-grained social media analysis. Whereas traditionally the host is considered to be the actor, when dealing with platforms the actor, or blogger in this case, is often defined after the slash. By detecting URL patterns, new actor definitions may be implemented *before* co-link analysis. Fine-grained social media analysis is similar in technique, but instead of only looking for actors, it is aimed at distinguishing actor and content links. The analysis is performed *after* co-link analysis.

Future research might cover in how far blog-native software features such as the permalink, trackback or RSS feed contributed to the construction of the early blogosphere and its subsequent transitions; in other words, how do changing linking practices relate to new features introduced by blog software? We would also like to further develop our hyperlink analysis by looking into different types of hyperlinks, beyond the <a href> link here retrieved from the front pages, and by distinguishing between 'traditional' hyperlinks, embed codes and social buttons and plugins in a hyperlink analysis [15].

Additionally, we also like to further explore the definition and circumscription of national blogospheres by enriching our research with content analysis, with a special interest in language transitions in the Dutch blogosphere. Content clusters do not only arise from linking practices but may also be defined through their common language. Moreover, the Dutch blogosphere may be analyzed on the basis of 'Web words' used, choosing the most distinctive and significant ones as points of departure, like *Retecool's* jargon or the specific language used by *GeenStijl*; two leading Dutch blogs. [16]

About the authors

Anne Helmond is a Ph.D. candidate with the Digital Methods Initiative, the New Media Ph.D. program at the Department of Media Studies, University of Amsterdam. In her research she focuses on software-engine relations in the blogosphere and cross-syndication politics in social media. She also teaches new media courses in the Media Studies Department.
E-mail: anne [at] digitalmethods [dot] net

Esther Weltevrede is a Ph.D. candidate with the Digital Methods Initiative, the New Media Ph.D. program at the Department of Media Studies, University of Amsterdam, where she also teaches. Esther's research interests include national Web studies as well as platform and engine politics. Additionally Esther has been coordinating the DMI Summer schools and is also a member of Govcom.org, a foundation dedicated to creating political Web tools.
E-mail: esther [at] digitalmethods [dot] net

Acknowledgements

We would like to express our sincere thanks to Erik Borra for developing custom tools, discussing methods and probing sharp questions, Mathieu Jacomy for his help with creating

Gephi maps and giving access to the G–Atlas tool for analyzing maps and Jan–Willem Hiddink and Robert–Reinder Nederhoed for providing a database dump from “Loglijst”. Last, but not least, we would like to thank Marguerite Lely and Anneke Agema for their editorial advice.

Notes

1. The project page of this empirical research including data, tools and visualizations is located at <http://dutchblogosphere.digitalmethods.net>.
2. <http://www.dejaap.nl/2010/12/28/verplicht-in-uw-rss-reader-Weblogs-die-er-echt-toe-doen/>, accessed 1 September 2011.
3. Belgium borders the Netherlands and the two countries share a common language: Dutch (Flemish).
4. The TLD Count tool is located at: <http://tools.digitalmethods.net/beta/tldCounts/>.
5. http://www.sidn.nl/fileadmin/docs/PDF-files_NL/SIDN_Jaarverslag_2010.pdf, accessed 1 September 2011.
6. <https://www.sidn.nl/nieuws/nieuwsbericht/article/sidn-kondigt-uitfasering-persoonsdomeinnamen-aan/>, accessed 1 September 2011.
7. http://en.wikipedia.org/wiki/Country_code_top-level_domain#Commercial_and_vanity_use, accessed 2 February 2012.
8. Google Refine is located at <http://code.google.com/p/google-refine/>, accessed 1 September 2011.
9. In 1999, none of the blogs were located on blog platforms because the first platforms were introduced around this time. For example, Pitas (later known as Blogger) was created in July 1999 and Diaryland in September 1999 (Helmond, 2008).
10. The tool can be found at <http://tools.digitalmethods.net/beta/sourceCodeSearch>.
11. Translation: “Ah, the free blog services. How we, the bloggers of the first hour, with our own domain and a self–made site, despised them, the services that allowed you to put up a blog in a few clicks. Look at him, he has a blogspot, or worse, a web–log.nl, which we scornfully called a web–dash–log.”
12. A software tool locating and visualizing networks on the Web, at <http://issuecrawler.net>.
13. Gephi is open source software for visualizing and analyzing large networks graphs, located at <http://gephi.org>, and the G–Atlas software is developed by TIC Migrations in Paris, located at <http://ticmigrations.fr/>.
14. http://www.comscore.com/Press_Events/Press_Releases/2011/4/The_Netherlands_Ranks_number_one_Worldwide_in_Penetration_for_Twitter_and_LinkedIn, accessed 1 September 2011.
15. Some of these ideas are currently being implemented in the Tracker tracker tool, developed by the Digital Methods Initiative. The tool is located at <http://tools.digitalmethods.net/beta/trackerTracker>.

References

- Lada Adamic and Natalie Glance, 2005. “The political blogosphere and the 2004 U.S. election: Divided they blog,” *LinkKDD '05: Proceedings of the Third International Workshop on Link Discovery*, pp. 1–16, at <http://dl.acm.org/citation.cfm?id=1134277>, accessed 1 February 2012.
- Rudolf Ammann, 2009. “Blogosphere 1998: Analysis,” at <http://tawawa.org/ark/2009/11/5/blogosphere-1998-analysis.html>, accessed 1 September 2011.
- Nancy Baym, 2007. “The widgetized self,” at <http://www.onlinefandom.com/archives/the-widgetized-self/>, accessed 2 February 2012.
- Yochai Benkler and Aaron Shaw, 2010. “A tale of two blogospheres: Discursive practices on the left and right,” *Berkman Center for Internet and Society Working Paper Series*, at http://cyber.law.harvard.edu/publications/2010/Tale_Two_Blogospheres_Discursive_Practices_Left_Right, accessed 2 February 2012.
- Rebecca Blood, 2004. “How blogging software reshapes the online community,”

Communications of the ACM, volume 47, number 12, pp. 53–55.

Rebecca Blood, 2002. "Weblogs: A history and perspective," In: John Rodzvilla (editor). *We've got blog: How Weblogs are changing our culture*. Cambridge, Mass.: Perseus.

danah boyd, 2006. "A blogger's blog: Exploring the definition of a medium," *Reconstruction*, volume 6, number 4, at <http://reconstruction.eserver.org/064/boyd.shtml>, accessed 1 September 2011.

Justus Bross, Matthias Quasthoff, Philipp Berger, Patrick Hennig and Christoph Meinel, 2010. "Mapping the blogosphere with rss-feeds," *Proceedings of the 24th IEEE International Conference on Advanced Information Networking and Applications* (Perth, Western Australia), pp. 453–460.

Axel Bruns, 2007. "Methodologies for mapping the political blogosphere: An exploration using the IssueCrawler research tool," *First Monday*, volume 12, number 5, at <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1834/1718>, accessed 1 September 2011.

Axel Bruns, Lars Kirchhoff and Thomas Nicolai, 2009. "Mapping the Australian political blogosphere," *Proceedings of the WebSci'09*, at <http://snurb.info/files/WebSci%20Poster.pdf>, accessed 2 February 2012.

Helen S. Du and Christian Wagner, 2006. "Weblog success: Exploring the role of technology," *International Journal of Human-Computer Studies*, volume 64, number 9, pp. 789–798.

Paul Girard, 2011. "HyperText Corpus Initiative: How to help researchers sieving the web?" Proposal for the "Using Web Archives" panel, *Out of the Box Conference* (9 May), at <http://www.medialab.sciences-po.fr/publications/Girard-HCI.pdf>, accessed 1 February 2012.

Brad L. Graham, 1999. "Friday, September 10, 1999," at http://www.bradlands.com/weblog/comments/september_10_1999, accessed 1 September 2011.

Anne Helmond, 2010. "Identity 2.0: Constructing identity with cultural software," paper presented at the DMI mini-conference at the University of Amsterdam (20–22 January), at <http://www.annehelmond.nl/2010/01/21/essay-on-identity-2-0-constructing-identity-with-cultural-software/>, accessed 2 February 2012.

Anne Helmond, 2008. "Blogging for engines. blogs under the influence of software-engine relations," M.A. thesis, University of Amsterdam, at http://www.annehelmond.nl/wordpress/wp-content/uploads/2008/09/helmond_mthesis.pdf, accessed 2 February 2012.

Timothy Highfield, 2009. "Which way up? Reading and drawing maps of the blogosphere," *Ejournalist*, volume 9, number 1, pp. 99–114.

Meg Hourihan, 2002. "What we're doing when we blog," at <http://www.oreillynet.com/pub/a/javascript/2002/06/13/megnut.html>, accessed 1 September 2011.

John Kelly and Bruce Etling, 2008. "Mapping Iran's online public: Politics and culture in the Persian blogosphere," *Berkman Center for Internet & Society, Research Publication*, number 2008–01, at http://cyber.law.harvard.edu/publications/2008/Mapping_Irans_Online_Public, accessed 1 February 2012.

Ravi Kumar, Jasmine Novak, Prabhakar Raghavan and Andrew Tomkins, 2004. "Structure and evolution of blogspace." *Communications of the ACM*, volume 47, number 12, pp. 35–39.

Geert Lovink, 2008. *Zero comments: Blogging and critical Internet culture*. New York: Routledge.

Frank Meeuwse, 2010. *Bloghelden*. Utrecht: LeV.

William Quick, 2002. "Tuesday, January 01, 2002," at <http://www.dailypundit.com/backupcreateblogoarchiveposts/backup000229.php>, accessed 1 September 2011.

Scott Rosenberg, 2009. *Say everything: How blogging began, what it's becoming, and why it matters*. New York: Crown.

Richard Rogers, 2005. "Old and new media: Competition and political space," *Theory & Event*, volume 8, number 2, at http://www.govcom.org/publications/full_list/rogers_old_new_media.html, accessed 2 February 2012.

Frank Schaap, 2005. "Links, lives, logs: Presentation in the Dutch blogosphere," In: Laura Gurak, Smiljana Antonijevic, Laurie Johnson, Clancy Ratliff and Jessica Reyman (editors). *Into the blogosphere: Rhetoric, community and culture of Weblogs*, at <http://blog.lib.umn.edu/blogosphere/>, accessed 1 September 2011.

Jan Schmidt, 2007. "Blogging practices: An analytical framework," *Journal of Computer-Mediated Communication*, volume 12, number 4, at <http://jcmc.indiana.edu/vol12/issue4/schmidt.html>, accessed 2 February 2012.

Michael Stevenson, 2010. "The archived blogosphere: Exploring Web historical methods using the Internet Archive," paper presented at Digital Methods mini-conference, University of Amsterdam (January); see <http://www.annehelmond.nl/2010/01/20/dmi-mini-conference-day-1-michael-stevenson-on-the-archived-blogosphere/>, accessed 2 February 2012.

Esther Weltevrede, 2009. "Thinking nationally with the Web: A medium-specific approach to the national turn in Web archiving," M.A thesis, University of Amsterdam, at https://wiki.digitalmethods.net/pub/Dmi/DmiSummer09/weltevrede_national_webs.pdf, accessed 2 February 2012.

Editorial history

Received 20 September 2011; accepted 23 January 2012.



"Where do bloggers blog? Platform transitions within the historical Dutch blogosphere" by Esther Weltevrede and Anne Helmond is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License](https://creativecommons.org/licenses/by-nc-nd/3.0/).

Where do bloggers blog? Platform transitions within the historical Dutch blogosphere by Esther Weltevrede and Anne Helmond
First Monday, Volume 17, Number 2 - 6 February 2012
<http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/3775/3142>
doi: 10.5210/fm.v17i2.3775