

Tracing Eurosceptic Party Networks via Hyperlink Network Analysis and #FAIL!ng: Can Web Crawlers Keep up with Web Design?

Michael Bossetta
University of Copenhagen
Øster Farimagsgade 5
DK-1353 Copenhagen K
+45 50 35 63 17
mjb@ifs.ku.dk

Anamaria Dutceac Segesten
Lund University
Centre for European Studies
22100 Lund
+46 (0)46 222 88 21
anamaria.dutceac.segesten@slav.lu.se

ABSTRACT

This #FAIL! paper is the result of our experience with the hyperlink analysis software 'Issuecrawler' (www.govcom.org) whilst writing the paper "The Europeanization of Eurosceptics? A Hyperlink Network Analysis of the Sweden Democrats" for the 2015 European Consortium for Political Research (ECPR) Joint Sessions. In attempting to visually map the online hyperlink network of the Sweden Democrats, a Swedish far-right Eurosceptic party, we found the Issuecrawler's inability to read JavaScript a critical limitation to our study. Knowing that other scholars from a range of disciplines are currently using the Issuecrawler and largely unquestioning its generated data, we produced this #FAIL! paper to educate other researchers about our experience and invite their opinions about the analytical utility of this digital method for future research. Can the software be improved to accompany dynamic programming languages? Or, is its inability to read JavaScript (a limitation shared by all current web crawling software) a reason to fundamentally question its utility as a digital method?

Keywords

Euroscepticism, Europeanization, Hyperlink Network Analysis, Issuecrawler, Digital Methods, Social Media

1. OUR STUDY'S INITIAL AIM

Theoretically, our study sought to explore the level of Europeanization amongst Eurosceptic parties; that is, we wanted to see if national Eurosceptic parties were affiliated with those in other Member States based on their anti-EU platforms. If such a network were to be detected, we could then pose the argument that Eurosceptic parties are, paradoxically, driving Europeanization, understood as a process of cross-border communication about Europe at the level of national politics. Using the Sweden Democrats (SD) as our test case, we posed the following research questions:

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

#FAIL! Workshop at the *Websci'15 Conference*, June 29, 2015, Oxford, UK.

- Which actors are connected to the Sweden Democrats and how are they networked online?
- Does SD's social network reach outside the national borders of Sweden and if so, who are their international connections?
- What role do Social Networking Sites (SNSs) play in linking the Sweden Democrats with other actors and do they facilitate a transnational communicative network of Eurosceptics?

To operationalize our research questions, we performed a series of hyperlink network analyses using the webcrawling software Issuecrawler. We grounded our approach mainly on three points suggested by existing research. The first is that hyperlinks are publicly available and socially significant markers of communication and/or coordination between actors. Second, as the 'structural glue' of the Internet, hyperlinking is a universal method of connectivity between actors online – regardless of language or geographical location. Finally, hyperlinking, like other forms of digital communication, can circumvent traditional power-relations and are therefore particularly suitable for the study of political challengers (like the Sweden Democrats). In sum, we consider hyperlink networks as proxies for offline actor networks, where linking or not linking suggests association and non-association respectively.

After a number of failed attempts in our learning of the software, we generated four completed webcrawls with the Issuecrawler to test two different, yet interrelated, dimensions of the Sweden Democrat's online networks. First, we crawled directly outwards from the SD's homepage and other closely affiliated sites in order to map their immediate hyperlink network. The aim was to see how SD is connected within Sweden by uncovering hidden actors not traditionally associated with the party, as well as explore whether SD's network extends beyond Swedish borders to non-Swedish actors. Second, we crawled the homepages of all Eurosceptic parties with MEP's in Brussels after the 2014 EP election to see if and how SD fits into a larger, pan-European network of Eurosceptics.

2. WHAT WENT WRONG

Our first set of webcrawls, i.e. those mapping SD's immediate network, began from a set of starting URL's comprising SD's homepage and those of its affiliates (e.g. its youth organization and newspaper), as well as the social media and blog links from its leading politicians. Our first crawl was a 'co-link' analysis,

where the Issuecrawler follows the hyperlinks outwards from the starting points and retains only those URLs shared by at least two of the starting points. The visual map generated by the co-link analysis showed SD at the periphery of the network, with mass media and Twitter holding the network together. Interestingly, the webcrawl revealed a number of alternative media sites (news aggregators and blogs) sympathetic to SD’s immigration-critical message. However, the Sweden Democrats were not shown to actively ‘out-link’ to any other source, which was strange to us considering we could clearly see hyperlinks on their homepage.

In an attempt to make our crawl more inclusive, we crawled the same set of starting points (minus the social media sites – to avoid their overrepresentation in the map) with a ‘snowball’ analysis. A snowball analysis, unlike a co-link analysis, does not look for shared linkages between the starting points; instead, it captures and retains all links from the starting points. The resulting map showed alternative media sites in the center of the map and traditional media clusters further from the center than in the co-link analysis. We did detect some international news sources, like FoxNews and CNN, demonstrating that indeed the snowball analysis found a broader scope of actors. Primarily, though, the uncovered international actors were based in Denmark. This can be explained by the fact that SD’s primary constituency is located in Southern Sweden, having historically has close linguistic and cultural ties to neighboring Denmark. Like our first crawl, we did not see the Sweden Democrat’s linking outwards to any actors, and therefore we did not find much evidence for Europeanization in the hyperlink analysis of SD’s immediate network.

Continuing our search for the Europeanization of SD, we decided to take a new approach inspired by the literature on the collaboration of Eurosceptic parties in the parliament. For example, SD is a member of the European Parliament party group ‘Europe of Freedom and Direct Democracy’, so at the very least SD should be linked to other members within this party group. To test this, we ran a co-link analysis of all 47 Eurosceptic parties that are represented in the European Parliament after the 2014 elections (Figure 1):

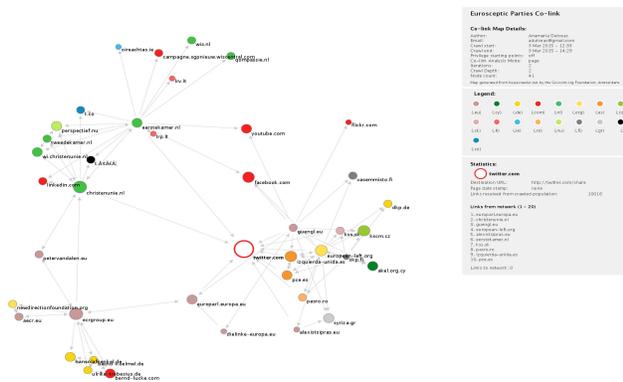


Figure 1. Co-Link Analysis of Eurosceptic Party Homepages

The nodes of the map were polarized by ideology (right ideologies on the left side of the map, and left-leaning ideologies on the right side). We found a strong clustering of the far-left, suggesting that they are closely linked through the European Left party group. In the middle of the map and connecting the two sides were the social media sites YouTube, Facebook, and Twitter. This suggests that social networking sites are common forums for being linked to, and thus supports the notion of these sites as arenas for pan-

European debate and the development of a European public sphere. Still, though, the Sweden Democrats were not represented in the visualization.

Finally, we ran a snowball analysis on all the Eurosceptic party homepages, expecting a huge map based on the number of starting points and the analysis’ focus on retaining all links found during the crawl. The resulting map had only a few nodes and was by-and-large unreadable. However, we saw that 27,297 URLs had been generated by the crawl, so we exported the data into Gephi in order to better visualize the network. We generated a map (Figure 2) that showed heavy national clustering, but little horizontal linkages across countries. In the center and holding the network together, as in the other crawls, were the SNSs Facebook, Youtube, and Twitter. Like in the other crawls, the Sweden Democrats were largely absent from the map, with only their initial homepage retained in the Gephi visualization, reproduced below:

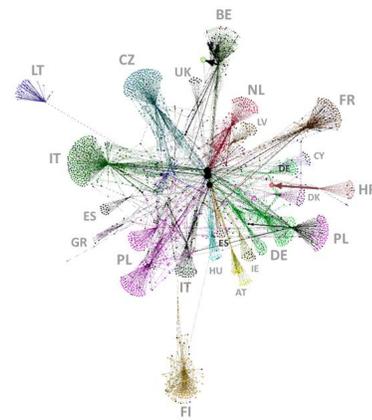


Figure 2. Gephi Visualization of Snowball Analysis of Eurosceptic Party Homepages

From the web crawls and visualizations, we can make a number of generalizable inferences based on the data: there is little evidence for Europeanization online, the Sweden Democrats are not highly networked, the far-left is better networked than the far-right online, Swedish alternative media dominate coverage of the Sweden Democrats, social media sites are integral lynchpins in connecting actors across ideologies and parties, etc.

3. THE NATURE OF THE PROBLEM AND WHAT OTHERS CAN LEARN FROM IT

However, through our research we uncovered a major limitation of the software that undermines the validity of our data: the Issuecrawler – like all other webcrawling softwares - does not read JavaScript. Although this limitation is stated in the Issuecrawler’s [FAQ](#), we had not come across this limitation stated anywhere in previous studies using the software.

Using the ‘Show Page Source’ function of the network browser Safari, we brought up the coding language of the Sweden Democrat’s homepage. Indeed, the Sweden Democrat’s homepage, and many of those from our original starting points, are written (or partially written) in JavaScript. This means that the Issuecrawler cannot crawl through SD’s homepage; furthermore, it does not find the links of other actors whose sites are written in dynamic programming languages. Today, there are many other (and more advanced) coding languages, allowing for hyperlinking

in newer forms than mere in-text linking. Missing these links construes an incorrect picture of online networks. For example, in our Gephi visualization (Figure 2), Eurosceptic parties from Finland and Lithuania are shown as outliers because they supposedly do not link to the SNSs in the middle of the map. Looking at the webpages of these parties, however, we found that they were indeed linking to SNSs, albeit through a widget that the Issuecrawler did not detect.

Based on this limitation, it is likely the case that other studies using webcrawling data have incomplete or skewed data sets (e.g. the aforementioned example of Finland and Lithuania). Although, as we showed in our paper, inferences can be drawn from the maps generated by the Issuecrawler, the big question is whether these inferences can be made with certainty taking into account the software's limitations. Although hyperlinking remains the fundamental and universal method of connectivity between websites, the rapid development of website design and the codes the web pages are written in seem to have outpaced the capabilities of hyperlink analysis software.

4. WAYS FORWARD

The ideal solution would be to update the Issuecrawler or develop new software that has the capabilities to map the hyperlinks of web pages written in JavaScript and other dynamic programming languages. We understand that this solution may be very time-consuming and costly, but unless academics also dedicate resources for developing, science will have to suffer. Perhaps through cooperation with/sponsorship by commercial outlets like Google, the necessary resources will be made available also for academic study.

At a fundamental level, we consider it necessary that social science researchers work in closer cooperation with computer scientists and developers to address this limitation, and/or find other methods of analysis to measure connections online (e.g. social media visualization methods).