3 Political news in search engines

Exploring Google’s susceptibility to hyperpartisan sources during the Dutch elections

Guillén Torres and Richard Rogers

Abstract
The research enquires into the susceptibility of Google’s search engine to provide users with questionable information when querying political parties and their issues during the run-up to the Dutch provincial and European parliamentary elections. Which rankings has the search engine assigned to problematic sources when querying political parties and their issues? Are there particular political issues and party spaces where these sources are prevalent or entirely absent? Do the ranks and amounts increase as the elections draw near? In all, it was found that hyperpartisan sources are rather pervasive in the search-demarcated political space, but far more so for certain actors and their issues on the far right of the political spectrum.

Keywords: Google Web Search, search engines, elections, social issues, digital methods

Introduction: Search engines as junk source space

As key entry points to the web, search engines serve as a site for the consumption of information, including political information, and as such are a relevant space for the study of both the presence of disinformation and junk news as well as approaches to combat it (Bowden, 2016). Although they are described in the industry as ‘organic’, the output of search engines could be

---

1 The research reported here was undertaken in collaboration with Anja Duricic, Lisa Fluttert, James Ingleby and Ziwen Tang.

Rogers, Richard, and Sabine Niederer (eds), The Politics of Social Media Manipulation. Amsterdam, Amsterdam University Press 2020
doi: 10.5117/9789463724838_CH03
termed manufactured hierarchy (Hindman, 2008; Halavais, 2017). Sources are ranked per query, and as such certain ones are offered as more relevant than others, as if naturally. Such ranking practices are often considered inscrutable, since search engines generally do not provide a means to save and study query results, e.g., through offering an API that enables it or terms of service that allow it. In fact, the ranking algorithms are trade secrets jealously guarded by corporations.

Since 2009 Google and subsequently other search engines have personalized results, be it for the individual or the place where the search has taken place (Pariser, 2011; Puschmann, 2018). Increasingly engines are thus both providing ranked political information but also tailoring it to user preferences and/or location (Martens et al., 2018).

When it comes to events, such as elections, search engines become providers not only of topical but also of timely information. These are particularly poignant moments to study the presence of disinformation and junk news. As cases in point, there have been occasions when sources that were otherwise insignificantly ranked rose to the top of engine returns during the ‘breaking news’ period of an event, such as in the immediate aftermath of the Las Vegas shootings in 2017, when a 4chan post misidentifying the shooter rose to the top of the results (Robertson, 2017). Google results also prominently linked to rumours about the identity of a Texas shooter in 2018 as ‘a Muslim convert, member of Antifa or Democrat supporter’ (Lomas, 2018). One could point to artificial manipulation, such as search engine optimization, as triggering the unexpected rankings and sudden presence of dubious information. With respect to the 4chan post, the gaming of the engine (if that were the cause) also may have been maliciously playful, introducing misinformation as an act of trolling. In the case of the rumours about the Texas shooter, the manipulation appears to have been hyperpartisan. Both spates of false news were not ‘corrected’ in the editorial sense of an erratum notice; rather, the dynamically published results are continually algorithmically tweaked so ‘good information’ is said to ultimately prevail (Waters, 2017).

In the study of engine returns and hierarchies (through manual capture techniques) it is often pointed out that top placement matters, since engine users over the years have been browsing fewer and fewer result pages (Jansen and Spink, 2003; Dan and Davison, 2016). Thus, in the above examples, the significance of the location of misinformation, rumour and extreme results relies on findings about how users gravitate to the top results, making them the most consumed and thereby particularly worthy of study. An additional research strategy for inquiries into junk news presence concerns anticipatory search, also known as autosuggestion, which drives the user to particular search
terms. It also has been studied for the offensive associations made by engines, such as the completion of ‘are Jews’ with ‘evil’ (Cadwalladr, 2016). Misogynistic autosuggestions also were documented in earlier cases which led to a UN campaign in 2013 discussed in a longer study of ‘how search engines reinforce racism’ (Noble, 2018). Other extremist content has been similarly documented towards the top of Google results for the query ‘holocaust’ (Hern, 2017).

The discovery by The Guardian journalist of the offensive associations with the word ‘Jews’ and the resulting sites that surfaced (such as the neo-Nazi website, The Daily Stormer) have led to discussions of not just how Google’s algorithms could be tweaked, but also the reach of the sources producing and driving such information in the first place. Their presence (and top placement) could be interpreted as a proxy for the significance and audience of such material online, or as others have argued as evidence of a culture war, driven by the online boosting tactics of ‘culture hackers’ (Albright, 2016; Confessore and Wakabayashi, 2017). Whilst they may seek to correct the autosuggestions (and perhaps remove religious ones all together), companies such as Google are hesitant to delist such extreme websites, given free speech concerns, which also may arise if they are nudged downwards.

The present chapter studies the susceptibility of Google’s search engine to provide users with questionable information sources in the results for queries related to Dutch political parties during the Dutch provincial and European parliamentary elections of 2019. Our goal is to identify the presence of dubious sources in the results for political queries. Thus, the research questions read as follows. Which rankings has the search engine assigned to false and junk sources when querying political parties and their issues? Are there particular political issues and party spaces where junk news is prevalent or entirely absent? Do the ranks and amounts increase as the elections draw near?

We have divided the chapter into six sections: a brief reflection about the methodological challenges of studying search engines, the methodology for building our dataset, three sets of findings, and a discussion of the limitations and further steps. In all it was found that junk news, specifically of the hyperpartisan variety, is rather pervasive in the search-demarcated political space, but far more so for certain actors and their issues on the far right of the political spectrum.

Studying personalization, junk news, or both?

The extent to which autosuggestions are personalized is understudied, but the personalization of results more generally has been the subject of
numerous inquiries and methodological innovations that work around Google's inscrutability through selectively scraping results or soliciting data donations. The findings from scraping batch-queried results have shown relatively low amounts of results affected by personalization in the Google search engine (Feuz et al., 2011), and the same held for Google News (Haim et al., 2018), suggesting that original concerns regarding the ‘filter bubble’ may not be as well founded any longer. Where the second method is concerned, Algorithm Watch, the German NGO, created a browser extension for users to install that regularly would make political queries, such as for [“Angela Merkel”] (Puschmann, 2017). The results would be donated by the users to Algorithm Watch so as to enable a larger number of those under study than is normally the case. Here again the findings have shown low levels of personalization, but the study of the presence of certain junk sites (be they disinformation or another genre) could be pursued further. Another technique, discussed below, is to selectively scrape results in a manner that seeks to minimize personalization effects, thereby concentrating on the overall presence of junk sites rather than on whether particular users, in a filter bubble, are seeing more of them.

To begin to understand the amount and placement of junk news in search engine space, be it around events or even after an algorithmic tweak, a query routine is designed, and a window of activity is chosen. (Longer-term studies also may be undertaken, as in the Issuedramaturg project that followed 9/11 query results for years, but Google often changes its output formats, breaking automated tools (Rogers, 2013).) In order to reduce personalization, a research browser may be deployed, which is a clean instance of a browser with the user not logged in. City-level geographical personalization may be avoided through the use of advanced settings, choosing a particular region, such as the Netherlands. For projects as ours, geographical personalization is not viewed as a disadvantage in the sense of creating the conditions for a filter bubble to materialize.

A brief mention should be made of the search engine under study. Among them Google is the most popular, with the largest market share of users in most countries, certainly in the Netherlands. As mentioned above, recently, the company has become entangled in the fake news debate through the appearance not of Russian disinformation sources (though that to our knowledge has not been studied in great detail), but owing first to the appearance of misogynistic and extremist content that the company previously defended as ‘reflective’ of societal concern rather than the product of algorithmic error or ‘culture hacking’. If one were to expand the number of search engines under study (to include Bing and Yahoo!,
for example), one could triangulate results, and inquire further into the normalcy and regularity of misogynistic and extremist content present in the top results, though one could not control for algorithmic concentration or the extent to which the big engines’ algorithms are anyways similar. The extent to which the results reflect societal concern would remain an open question.

Studying Google results

The presence of junk news within Google’s search engine results is a multi-causal phenomenon that may be credited to a number of factors. Among others, Google’s algorithm reacts and learns from users’ own consumption of junk sources. It is trained using varied datasets, and content producers’ attempt to game the search engine via search engine optimization tactics (Finkel et al., 2017). Given the inscrutability of Google’s tools, it is difficult to determine what could be causing the presence of junk sources in the Dutch web sphere or others. Here, rather than attempting an explanation for the presence of junk news, we conduct a test of the engine’s susceptibility to connect politically relevant queries with junk sources.

As noted, the investigation relies on scraping as a method (Marres and Weltevrede, 2013), and takes as its point of departure the question of junk news in search returns rather than the effect of personalization in the creation of a filter bubble. The research seeks so-called junk news in search engine results, which has been defined as ‘extremist, sensationalist, conspiratorial and masked commentary’ (Howard et al., 2017: 1). In keeping with Buzzfeed News’ definition of fake news (Silverman, 2016), we also seek (foreign) disinformation, hyperpartisan sources as well as clickbait, which itself may be extreme. In order to do so we rely on a list of sources expertly curated by other researchers in the project (see the Appendix 6.2 in Hagen and Jokubauskaite, this volume).

Generally, the research employs the ‘source distance’ approach, inquiring into how far from the top of the returns are the offending results (Rogers, 2013). More specifically, we investigate how false and junk webpages are positioned in the first twenty Google.nl results of various queries of political parties and their most significant issues during the 2019 provincial elections campaign as well as those of the European parliamentary elections. Thus, this case examines the susceptibility of search engine results to junk news, as defined above, rather than exploring the issue of falseness and junk in themselves or the effectiveness of countermeasures.
Engine returns as political spaces

In order to demarcate a political space in search engine results, we designed a query protocol based on combining the names of political parties with specific issues associated with their respective political agendas. In that sense, the underlying assumption is that junk news may have a more significant impact when discussed in a specific political context, such as election campaigns, when voters gather information to guide their choice. To be able to collect the results of a large number of queries, we used the Search Engine Scraper by the Digital Methods Initiative, a tool that allows one to scrape search engine results for a given query and commit them to a database for further scrutiny through visualization.

The first step of the methodology consisted in constructing a list of Dutch parliamentary parties and locating their websites as well as Facebook pages (see Table 3.1). Another list was created pertaining to the parties participating in the European parliamentary elections. In the next step we identified the relevant keywords to build the queries; in the case of the Dutch provincial elections, these were sourced from both the party webpages and the party Facebook pages. For the parties competing in the European parliamentary elections, only the parties’ own websites were consulted. The party webpages provided the parties’ issue keywords. The Facebook pages furnished a more vernacular set of issue terms, as they contain issue keywords from users or citizens in the comment space. The aim of sourcing these two sets of keywords is to enable us to capture and compare the results for both official as well as more popular issue language.

The lists of party keywords were built by collecting the platform standpoints (standpunten) on the party websites. There is one list for the provincial elections and another for the European parliamentary elections. Most political parties mention between five and ten keywords on their platforms, and all were collected. A few parties (e.g., the Staatkundig Gereformeerde Partij, SGP) offer longer lists which were shortened on the basis of their key issues. In all, the political party issue space consisted of 158 keywords across the thirteen parties for the provincial elections, and four keywords across fourteen parties for the European parliamentary elections. The vernacular list was made through a close reading of the comments made under the posts of the Facebook pages of each political party. To build this list, the most commented posts around the days of the elections were close-read, and the most representative keywords related to the views expressed by the commenters were chosen. Identifying the most relevant issues in the comment space on Facebook proved to be problematic, given that the comments
were often polarizing and emotive, especially those relating to the elections themselves. This citizen-enriched political issue space consisted of five keywords for each of the thirteen parties, making a total of 65.

Table 3.1  List of Dutch political parties under study

<table>
<thead>
<tr>
<th>Dutch Provincial Elections</th>
<th>European Parliamentary Elections</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name of the Party</td>
<td>Abbreviation</td>
</tr>
<tr>
<td>Volkspartij voor Vrijheid en Democratie</td>
<td>VVD</td>
</tr>
<tr>
<td>Partij voor de Vrijheid</td>
<td>PVV</td>
</tr>
<tr>
<td>Christen-Democratisch Appèl</td>
<td>CDA</td>
</tr>
<tr>
<td>Democraten 66</td>
<td>D66</td>
</tr>
<tr>
<td>GroenLinks</td>
<td>GL</td>
</tr>
<tr>
<td>Socialistische Partij</td>
<td>SP</td>
</tr>
<tr>
<td>Partij van de Arbeid</td>
<td>PvdA</td>
</tr>
<tr>
<td>ChristenUnie</td>
<td>CU</td>
</tr>
<tr>
<td>Partij voor de Dieren</td>
<td>PvdD</td>
</tr>
<tr>
<td>50Plus</td>
<td>50plus</td>
</tr>
<tr>
<td>Staatkundig Gereformeerde Partij</td>
<td>SGP</td>
</tr>
<tr>
<td>Denk</td>
<td>DENK</td>
</tr>
<tr>
<td>Forum voor Democratie</td>
<td>FvD</td>
</tr>
<tr>
<td>De Groenen</td>
<td></td>
</tr>
</tbody>
</table>

The three lists of keywords were inputted in the Search Engine Scraper along with the name of each party. The results from the parties’ own websites were excluded. For example, for the political party D66 and the keyword onderwijs (education), the following query was made: [onderwijs d66 -d66.nl]. Using the advanced search features of Google, maximum results were set to 1,000, and each day of our periods of interest (13-22 March and 22-24 May 2019) was queried separately. The date ranges included the run up to the provincial elections on 20 March and a short election aftermath period, and the days before and after the European Parliamentary elections on May 23. The searches were conducted in a clean browser, in the Dutch Google.nl domain, in the Dutch language, and in the Netherlands region (through the advanced settings). The keyword and party were queried together so that the scraper tool delivered results that are related to election politics, rather than a general overview by querying each keyword in isolation. The keywords derived from Facebook were queried in the same format, using the same settings and date range.
The outputs of the Scraper tool are the top twenty ranked Google.nl results per query. The URLs in our lists were then truncated to their host names so that they could be cross-checked with the list of known false and junk websites curated by other researchers in the project. Here a formula was used that effectively linked the search engine results spreadsheet to that of the expertly curated list of junk sites. The question for each source concerned its ranking per query and its presence or absence in the expert list. All query results (per party and per official or vernacular language type) were marked as junk sites or not and listed in the order they were returned.

Table 3.2 List of categories and political keywords used in the study

<table>
<thead>
<tr>
<th>Category</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>Party platforms</td>
<td></td>
</tr>
<tr>
<td>Foreign affairs</td>
<td>Europese Unie, Europa, EU, Nederland en Europa, buitenland, internationale zaken</td>
</tr>
<tr>
<td>Polarizing topics</td>
<td>Islamisering, Islam</td>
</tr>
<tr>
<td>Health</td>
<td>zorg, menselijke zorg</td>
</tr>
<tr>
<td>Environment</td>
<td>klimaat, natuur, milieu, dierenrechten, natuur en milieu</td>
</tr>
<tr>
<td>Finances</td>
<td>belasting, economie, inkomens, pensioen, werk en inkomens, schone economie, eerlijk delen, werklozen, economisch beleid, overheid en bestuur</td>
</tr>
<tr>
<td>Safety and security</td>
<td>veiligheid, privacy, defensie, criminaliteit, rechtsstaat, terrorisme</td>
</tr>
<tr>
<td>Society</td>
<td>waarden, gezin, respect, familie, samenleving, burgers, democratie, ouderen, onderwijs, goed onderwijs voor iedereen, vrijheid, verantwoordelijkheid, drugs</td>
</tr>
<tr>
<td>Future</td>
<td>innovatie, duurzaamheid, schone energie, energie</td>
</tr>
<tr>
<td>Migration</td>
<td>immigranten, migratie</td>
</tr>
<tr>
<td>Facebook</td>
<td></td>
</tr>
<tr>
<td>Foreign affairs</td>
<td>Europa, EU, referendum</td>
</tr>
<tr>
<td>Polarizing topics</td>
<td>Islam, Moslim, racisme, discriminatie(^2)</td>
</tr>
<tr>
<td>Environment</td>
<td>milieu, klimaat, kernenergie, energie</td>
</tr>
<tr>
<td>Finances</td>
<td>belasting, bezuinigingen, pensioen, onderwijs</td>
</tr>
<tr>
<td>Society</td>
<td>samenleving, democratie, toekomst crisis, vrouwen, vrijheid, Nederland, armoede</td>
</tr>
<tr>
<td>Faith</td>
<td>Islam, Moslim, Christendom, Christenen, geloof</td>
</tr>
<tr>
<td>Migration</td>
<td>migranten, immigranten, gelukszoekers, migranten</td>
</tr>
</tbody>
</table>

\(^2\) Islam is placed in both faith as well as polarizing topics categories, given how it is discussed as shorthand for a social issue.
We zoomed in on those queries in which junk news showed up consistently, that is, for a minimum of four days within our period of interest for the provincial elections, and two days for the European parliamentary elections. The keywords that produced junk news websites in their search results in the first case were then grouped thematically in the following categories: ‘foreign affairs’, ‘polarizing topics’, ‘health’, ‘environment’, ‘economy’, ‘safety and security’, ‘society’, ‘future’, ‘migration’ and ‘faith’, the most salient of which are described in some detail below (see Table 3.2). For the case of the European parliamentary elections, the four keywords common to all parties were queried: Europese Unie [“European Union”], klimaat [“climate”], migratie [“migration”] and economie [“economy”].

Political parties and issue keywords

Before analyzing the presence and positioning of junk news in Google web search, we would like to discuss briefly the keywords obtained for the provincial election campaign from the official websites and Facebook pages of the political parties. Comparing the composition of the categories that emerged from each of the two political spaces allows for showing differences between the matters of concern as expressed by political parties and citizens or social media users. Whereas political parties included keywords that could be grouped under the categories, ‘future’, ‘security’, and ‘health’, that was not the case for the Facebook users. In contrast, ‘faith’ was present in the Facebook comment space, whereas it was largely absent from the party platforms (except for the SGP, with its long list). There are also matters of concern common to citizens and political parties alike, such as ‘foreign affairs’, ‘economy’, ‘society’, and ‘environment’.

Within the shared concerns there are still differences between the way each political space is constructed by political parties or citizens. For example, within the ‘foreign affairs’, ‘economy’ and ‘society’ categories, parties tend to refer to a wider variety of issues in comparison to the concerns expressed by citizens, which are mostly focused on the European Union and the referendum. In the economy cluster, political parties address ten issues, whereas citizens are concerned with far fewer. The same holds roughly for the society cluster. Interestingly, this trend reverses in the environment cluster, where users tend to express concerns about nuclear energy, while

3 Islam was present in the party platforms (largely the PVV and SGP) but discussed in terms of a social issue.
the topic is not present in parties' platforms (again, with the exception of SGP). Lastly, even though the usage of some keywords is more or less the same in party platforms and the Facebook comment spaces, for others it differed, as is the case for onderwijs (education). By the political parties it is framed as a societal issue, whereas in the discourse on Facebook it is discussed in terms of citizens’ ability to afford it.

Visualizations are made to facilitate the analysis; they show at a glance the presence and ranked position of junk news for each query over time. The columns represent the days in the timeframes studied (13-22 March or 22-24 May 2019). Red cells indicate the source as marked as junk news. One also may read the placement and distribution of it over time.

**Political party standpoint space**

The study of the political party standpoint space found overall that all false and junk webpages that appeared can be subcategorized as hyperpartisan. The one exception fell within the environment cluster in the form of a conspiracy website in the top twenty. Three specific websites make up for the biggest amount of junk: *De Dagelijkse Standaard, Stop de Bankiers,* and *Opiniez.* A second general observation is that queries related to right-wing parties returned hyperpartisan sources in a greater proportion than queries related to parties with other political orientations. In particular, the queries related to the FvD were the most populated by hyperpartisan sources. Thirdly, queries related to parties located at the centre of the political spectrum seem to produce results with less questionable sources. In those instances where hyperpartisan websites appear among the top twenty results for centre-oriented parties, the sites mainly do not make it to the top positions (though there were exceptions).

In the political party standpoint space, most junk appears to be associated with queries related to keywords within ‘foreign affairs’ and ‘polarizing topics’. The keywords within the foreign affairs cluster mainly relate to the European Union and the Dutch relationship with it. Issues related to political parties from the centre of the political spectrum seem to be least connected to junk, as may be noted for 50Plus, Partij voor de Dieren or D66 (see Figure 3.1). Contrariwise, parties that position themselves strongly against the European Union are linked to a high concentration of junk results. For example, when looking at FvD and PVV, we can see that hyperpartisan sites account for 37% and 47% respectively of all the returns discussing the European Union. Moreover, the hyperpartisan sources are among the top five results throughout almost the entire time period.
Regarding the keywords within the ‘polarizing topics’ cluster (Figure 3.2), two related issues in the political party-demarcated space lead to search engine results with a large amount of hyperpartisan sources: Imam and Islamisering (Islamization). These keywords are only discussed by two right-wing parties, PVV and SGP, the latter from the religious right. In particular, when PVV is queried together with the keyword Islamisering (Islamization), hyperpartisan websites appear at the top of the results throughout the entire time span, occupying even the highest positions. This changes only for three days (16, 19 and 20 March), when, however, the amount of junk increases overall.

The queries for environmental keywords (Figure 3.3) also lead to significant quantities of junk. First, it is of note that the amount decreased as
the elections neared, though junk websites still maintained a prominent position among the first 10 results. Once again, queries mentioning right-wing parties such as FvD are more prone to produce junk sources within their results. The site, ninefornews.nl, was identified here as the only conspiracy website in the entire data set. It emerged in the 8th position when [milieu FvD -www.fvd.nl] was queried, meaning that it appears on the first page of Google results (with default settings at 10). The centre party, D66, also registered significant amounts of junk for the query climate (klimaat), as did the centre-left labour party (PvdA).

The cluster related to economic issues, the second largest in our set of results, contains nine keywords such as belasting (taxes), eerlijk delen (fair sharing) and werklozen (unemployed) (see Figure 3.4). Queries including FvD are once again among the most prone to produce hyperpartisan results, particularly with the keywords economie (economy), pensioen (pension) and belasting (taxes). When examining the remaining parties (with the possible exception of the PVV), the number of junk results is relatively low overall, and they rarely occupy the first position of the search engine results.

The ‘society’ cluster is the largest; twenty issues were identified in this cluster, ranging from concerns about warden (values) to drugs (see Figure 3.5). Social issues produced a varied distribution of hyperpartisan sources over-time, with hyperpartisan websites appearing less prominently. Only three queries, two including the FvD and one the VVD, stand out in this cluster as junk-ridden: drugs, democracy and responsibility (verantwoordelijkheid).

Lastly, the thematic cluster, ‘future developments’ (Figure 3.6), with such issues as duurzaamheid (sustainability), innovatie (innovation), and schone energie (clean energy), are addressed by parties located at the extremes of the political spectrum and the query results are populated by hyperpartisan sources. This is most evident in the case of FvD, where questionable sources appear nearly every day, and in four instances are returned in the top two positions, albeit not in close vicinity of election day, which remains largely unaffected by junk sources in relation to these terms.

Vernacular political issue space

Based on the keywords gleaned from the Facebook pages of the Dutch political parties, one new category was created (‘faith’) on top of the other six from the previous exercise. Generally, the results were similar. Queries mentioning right-wing parties such as PVV and FvD returned more hyperpartisan
Figure 3.3  Presence of junk news in Google.nl search engine results for political queries related to the environment, 13-22 March 2019
Figure 3.4 Presence of junk news in Google.nl search engine results for political queries related to the economy, 13-22 March 2019
Figure 3.5  Presence of junk news in Google.nl search engine results for political queries related to societal issues, 13-22 March 2019
sources in the top 20 Google results, and the positions of these junk sources tend to be higher than in queries related to other parties.

In the vernacular political issue space, the foreign affairs cluster received the most junk news returns; nearly 25% of the top 20 results are considered hyperpartisan websites. Migration and polarizing topics have the second and third highest percentage of junk websites, with 19% and 17% of junk, respectively. Results associated with keywords such as *immigranten* (immigrants) and *Islam* were significantly populated with hyperpartisan sources. Issues related to environment, society, and finance obtained fewer junk returns (less than 10%). The faith cluster had the fewest hyperpartisan returns.4

The environment cluster (which for many terms could have been merged with the future innovations one) is the largest one in the vernacular issue space. In it the keyword *klimaat* (climate) was discussed by Facebook users on the pages of five different parties, and *milieu* (environment) on three. For Groenlinks, D66, PvdA and VVD, the keyword *klimaat* features hyperpartisan sources in the top results (see Figure 3.7). From the election day onwards, fewer junk websites showed up in the top 20 results. The same pattern was observed in the ‘foreign affairs’ cluster (see Figure 3.8). Hyperpartisan websites occupied the first position for five days during our 10-day research period for queries related to FvD and PVV. Queries for FvD and EU were the most likely to return junk webpages.

Regarding the keywords grouped under the ‘polarizing topics’ cluster, *Islam* was brought up by Facebook users in the pages of four political parties:

---

Figure 3.6 Presence of junk news in Google.nl search engine results for political queries related to future innovation, 13-22 March 2019

---

4 Here again, Islam is excluded, because it is considered a social issue, given the manner in which it is discussed in the vernacular issue space.
Figure 3.7  Presence of junk news in Google.nl search engine results for political queries related to the environment, using language from the Facebook comment space of the political parties, 13-22 March 2019
PVV, SGP, FvD and DENK. Among them, results associated with FvD received the most junk webpage returns in the top 20 results, and hyperpartisan results maintained the first position for seven days (Figure 3.9). On election day, however, all the parties returned few to no such results in the cluster of ‘polarizing topics’, with the exception of the FvD. The number of junk sources and the rankings dropped dramatically on that day. Concerning DENK, although queries including it did not return many junk websites overall,

For the political party, DENK, Islam, while discussed as a social issue, also could be categorized as ‘faith’.
they were found in relation to the issues *Moslim* (Muslim), *discriminatie* (discrimination) and *racisme* (racism).

The migration cluster (Figure 3.10) was more prominent in the discussion space on Facebook pages than in the platforms of the political parties. Keywords as *migranten* (migrants), *migratie* (migration), *immigranten* (immigrants) and *gelukszoekers* (fortune seekers) were often mentioned in the comments on the Facebook homepages of political parties. (*Gelukszoekers* could be said to be a pejorative term for economic migrants.) Results for three parties in particular (FvD, PVV and VVD) had junk webpages in the top 10 Google search results, and they were spread evenly over the ten-day research period. It is also noteworthy to see that there appears to be a decreasing tendency of junk news from the election day onwards in this cluster, both in terms of the amount as well as the rankings.

**European Parliamentary Elections**

The results for the political party standpoint space during the European Parliamentary elections also show a consistent presence of junk news. Three major findings are worth mentioning. Firstly, as figures 11 and 12 make evident, the presence of junk sources in the Google search engine results was lower during the European parliamentary elections than during the provincial elections. In many cases, our queries combining issues with parties did not produce links to hyperpartisan material or only did so for one day. *Only 25%* of our queries returned junk for more than one day during the three-day period under research. Of the four keywords queried, the one that produced the least amount of junk in combination with party
names was *economie* (economy). In contrast, *migratie* was the most prone to returning junk, with queries related to five parties consistently returning junk websites between May 22 and 24 (see Figure 3.11).

Queries in combination with the political party DENK were particularly junk ridden. On the day of the election and the day after, almost half of the results provided by the Google search engine is problematic. It is particularly relevant that during the 24th of May, hyperpartisan websites occupied the top seven positions.

A second relevant finding is that the presence of hyperpartisan resources is more prevalent on the day after the election than the day before, in contrast to what happened during the provincial elections. Although this is the case for all parties and keywords where junk sources were identified, it was especially prominent for the issues of migration, economy and climate (keywords *migratie*, *economie* and *klimaat*). In most cases, junk sources also occupy the first positions in the results during the 24th of May. We cannot answer the question of whether this phenomenon can be credited to the Google search algorithms reacting to an increase in searches related to the elections, to a surge in the activity of hyperpartisan websites after they took place, or to some combination. A close reading of the results in the highest positions, however, shows that the hyperpartisan sources behave as one would expect from any information provider during election season, if only keeping their radical tone; before the election they provide predictions about the results, on the day of the election they invite citizens to vote, and on the day after they discuss the results. An article by the *De Dagelijkse Standaard* that consistently occupied the top result for various keywords and parties is a reflection about how the FvD and leader Thierry Baudet need to tone down their radical discourse in order to become a more powerful political force.

Lastly, whereas during the Dutch provincial elections the queries related to right-wing parties were more strongly connected by the search engine with junk sources, during the European elections this is not the case. For example, although queries performed in May involving FvD also produced results pointing to hyperpartisan websites – similarly to the results obtained in March – it was those related to DENK which, in aggregate terms, produced more junk (i.e., 25 for FvD and 27 for DENK). However, the case of DENK is difficult to assess given that, apart from the name of a party, it is also a common Dutch word (‘think’, in English). Although this does not change the fact that people looking for information about this party would likely be exposed to junk sources, the content may not specifically relate to DENK. In fact, the highest-ranking result for the query [migratie DENK -bewegingdenk.
Figure 3.11  Presence of junk news in Google.nl search engine results for political queries related to migration and European Union issues, 22-24 May 2019

Figure 3.12  Presence of junk news in Google.nl search engine results for political queries related to climate and economic issues, 22-24 May 2019
nl] is an article in the hyperpartisan website *De Dagelijkse Standaard* that makes no specific mention to this party, but instead generally reflects on the results of the elections and specifically on the demise of the PVV.

**Conclusions: Junk news in search engine results**

The goal of this research is to locate the presence and ranking of junk websites within the first twenty Google.nl results of queries concerning Dutch political parties and their most significant issues during the 2019 provincial and European parliamentary elections. The keyword queries were built by combining political party names with keywords retrieved from political party platforms and party Facebook page comment spaces (in the case of the provincial elections). We clustered the keywords into categories to enable a comparative analysis. At the outset the research questions were as follows: Which rankings has the search engine assigned to junk sources when querying political parties and their issues? Are there particular political issues and party spaces where junk news is prevalent or largely absent? Do the ranks and amounts increase as the elections draw near?

Our results indicate that the junk websites present in the results of our Google.nl political queries are almost exclusively hyperpartisan, rather than sources spreading disinformation, conspiracy theories, or clickbait. Three websites, namely *De Dagelijkse Standaard*, *Stop de Bankiers* and *Opiniez*, account for the largest portion of the junk sources identified. We did not find fake advocacy groups or foreign disinformation operatives. Furthermore, we found that queries involving right-wing parties were more prone to result in exposure to hyperpartisan sources than those associated with centre-left or left-wing parties. For most keywords, hyperpartisan websites appeared in the top positions, and certainly always within the first page of results. Our findings suggest that on Google.nl there is a considerably high probability that junk news is outputted on the first page of results when the queries concern right-wing parties and their issues.

Another finding is that hyperpartisan results spread unevenly during our research period. For the case of the Dutch provincial elections, they are more present before the day of the election and drop their presence and ranking, in some cases dramatically, on election day and in the immediate aftermath, for instance in the case of migration issues in the vernacular issue space. In contrast, during the European parliamentary elections this phenomenon reversed, and junk sources were more prominent the day after the elections.
Concerning the comparison between the two spaces, the vernacular space had the highest percentage of junk news returns, especially in the categories of foreign affairs and migration. In the political party standpoint space, the differences among themes is smaller. In other words, the queries designed with the political language of Dutch Facebook users were more likely to result in hyperpartisan results than the queries built with the standpoint language of political parties. Although more research is needed in this regard, such a finding suggests that the discourse of normal citizens, or those commenting on party Facebook pages, is more politically contentious than that of political parties.

The results are indicative of the amount of hyperpartisan material in political space in Google.nl rather than conclusive, for they derive from a particular query strategy and not from multiple strategies and are only a snapshot from a particular event-related timeframe. The data set we built could also be read more closely, and additional junk sources could be found, meaning that we could have undercounted (rather than overcounted).

Given that our intention is to determine the susceptibility of the Google search engine to junk news, the question remains whether the location of the results of politically charged queries can be credited to an optimization effort on the side of hyperpartisan content generators, an overall susceptibility of Google’s search algorithm to provide questionable content to its users, consumers’ preference for low quality information, or some combination of the three.

While discussing the two latter hypotheses would require more space, a few words can be said about the first. A possible way to detect search engine optimization strategies consists of using one of the many online services providing SEO analysis. Given the proprietary nature of their methodologies, however, the results should be interpreted with caution. We submitted the three most recurrent junk websites we found during our research to the service SEO Tester Online, a tool that measures a website’s readiness to achieve top positions within search results. This tool provides analysis in four different categories: basic (related to the overall online presence of the website), content (measuring the richness of the keywords that trigger the website to pop up in search results), web performance (indicating how fast the website can be rendered in mobile and desktop devices), and social (providing information about the website’s engagement with users through social networks). De Dagelijkse Standaard obtained a score of 56/100, faring the lowest in their web performance, and the highest in its social engagement. Opiniez obtained a score of 62/100, faring the best in web performance and the lowest in content, although it obtained an excellent
score in the number of unique keywords. *Stop de bankiers* obtained the highest score, with 74/100, including a 100/100 in the assessment of keywords which may lead to the site through search engines. It also fared the best in social engagement and the lowest in web performance.

The reports produced by *SEO Tester Online*, which for all websites included a considerable amount of suggestions about how to improve the websites' code, suggest that, at least currently, content producers are not actively seeking to artificially alter the performance of their sites. Consequently, the presence of these junk sources in our analysis is more likely related to an inherent susceptibility of Google's search engine and/or an organic result of users' preferences.

Further research is necessary to paint a clearer picture regarding the increased consumption of junk news. One could repeat the work for longer periods of time in order to ascertain the extent to which the presence of these or other dubious source types is becoming more widespread or even disappearing in the top results for political queries, as is currently the case with clickbait. Furthermore, specific sources could be monitored over time to track the performance of their content (and the changes to their code), in order to detect attempts to artificially increase relevance and thus the ranking assigned by search engine algorithms.

References


Lomas, Natasha (2018). “Google is surfacing Texas shooter misinformation in search results — thanks also to Twitter”, TechCrunch, 11 June.


Robertson, Adi (2017). “After its 4chan slip-up, is it time for Google to drop Top Stories?”, The Verge, 3 October.


About the authors

Guillén Torres is a Ph.D. researcher at the University of Amsterdam, within the DATACTIVE research group, and an information activist at ControlaTuGobierno, a Civil Society Organization based in Mexico. His work focuses on how data and information mediate the interaction between citizens and the state.

Richard Rogers is Professor of New Media & Digital Culture at the University of Amsterdam and Director of the Digital Methods Initiative, the group responsible for social media research tools. Among other works, Rogers is author of Information Politics on the Web (MIT Press, 2004), Digital Methods (MIT Press, 2013), and Doing Digital Methods (Sage, 2019).