



To Polarize or Not:

Comparing Networks of News Consumption

Susan Banducci, Lorien Jasny, Iulia Cioroianu, Travis G. Coan*, Daniel Stevens**, Ian Weaver and Hywel Williams***

* Exeter Q-Step Centre, University of Exeter, UK; S.A.Banducci@exeter.ac.uk, L.Jasny@exeter.ac.uk, i.cioroianu@exeter.ac.uk, t.coan@exeter.ac.uk;

** Politics (Cornwall), University of Exeter, Exeter, UK; d.p.stevens@exeter.ac.uk;

***Computer Science, University of Exeter, Exeter, UK, isw1g10@soton.ac.uk, h.t.p.williams@exeter.ac.uk;

Abstract: *We use individual data on browsing histories combined with survey data to examine whether online news exposure exhibits signs of segregation and selectivity. By using online news behaviour combined with survey reports of attitudes, we can capture exposure to both traditional news sources and news shared via social media platforms. Most importantly, we can also examine what types of individuals (e.g. partisans, educated) are more likely to exhibit selective tendencies. We find, consistent with recent empirical work, the extent of segregation in exposure may be overstated. Furthermore, the degree of segregation and selectivity varies across groups that are defined by holding shared political preferences. For example, in the case of Brexit, those who supported the 'Leave' side were more selective in their news exposure. Our approach allows comparison of news exposure patterns by domains versus news exposure to topics. To our knowledge, this is the first analysis to allow this comparison.*

Keywords: polarization, segregation, clickstream, online news consumption, Brexit, survey, political attitudes

Acknowledgement: Paper prepared for submission to Political Polarization Workshop - Social Media and Political Polarization in Non-U.S. Contexts to be held at the Conference on E-Democracy and Open Government Asia 2018 (CeDEM Asia 2018) July 12-13th, 2018, Yokohama, Japan. This work was supported by the Economic and Social Research Council ES/N012283/1 "Measuring Information Exposure in Dynamic and Dependent Networks (ExpoNet).

Normative theories of democracy suggest that, to function effectively, citizens should have access to a variety of sources of information (Downs, 1957). Ideally, citizens form their preferences and understand the preferences of others through exposure to information from a variety of viewpoints. As the dominant source of political information for citizens, there seems to be little

question that the media matter as providers of information in politics in general and in elections in particular. However, the media landscape has changed –

more people are accessing news online (both web-based versions of legacy news sources and ‘digital native’ sources), social algorithms direct users to news and information (e.g. filter bubbles), and news and information is shared via social media.

As the internet has become a dominant platform for disseminating news and information, individuals are faced with a wider range of options (from social and traditional media), new patterns of exposure (socially mediated and selective), and alternate modes of content production (such as user-generated content) (Valkenburg, Peter, & others, 2013). One of the important questions that researchers are grappling with is whether this development has increased exposure to different viewpoints or whether it has led citizens to a narrower range of viewpoints. While initially thought to democratize access to information, there is now conflicting evidence on how the use of the internet and social media influence public opinion (Bode, 2015; Boulianne, 2015; Sudulich, 2011). The empirical evidence about selectivity and about the impact of online news consumption on voters’ attitudes and behaviour is far from conclusive, and scholars have been cautious in making claims about the impact of the internet on increased selectivity (Quintelier & Vissers, 2008; Tucker et al., 2018). The conclusion that “people restrict themselves to their own points of view” (Sunstein, 2004, p. 4) may be overstated. Despite the lack of consensus over the direction and size of effects, there have been advances in terms of conceptualizing the nature of media influence (Bennett & Iyengar, 2008; Iyengar & Simon, 2000) and the methods used to study them.

Our analysis of patterns of news exposure proceeds as follows: We first situate our contribution in the ongoing debate about segregation in online news exposure. We then present details on data collected around the EU Referendum vote in Great Britain in 2016. In order to assess the attitudinal basis of selectivity and segregation, we combine the online browsing histories with an online panel survey. We are then able to analyze how segregated the news consumption is of “Leave” and “Remain” voters by using indicators of the content of news domains and news stories. Our findings suggest that predispositions and preferences are linked to different news consumption patterns, with some citizens exhibiting a greater degree of segregation. However, the news consumption patterns exhibit a great deal of overlap. In particular, public service broadcasting (BBC) represents a shared news experience.

1. Echo Chambers or the Marketplace of Ideas: Online News Consumption

One of the central questions about the effect of the technological changes on how citizens access news is whether the developments have contributed to increased polarization or segregation (i.e. exposure to a narrower range of views) in the citizenry. On the one hand, many have argued along with Sunstein (2009) that when given more options, individuals will choose to consume content that fits with prior beliefs. Thus, they are creating “echo chambers” – views reflecting the narrow range of views to which they choose to be exposed.¹ The lack of exposure to a variety of viewpoints might

¹ V.O. Key in *The Responsible Electorate* describes public opinion as but an “echo...the output of an echo chamber bears an inevitable and invariable relation to the input. . . the people’s verdict can be no more than a selective reflection from the alternatives and outlooks presented to them” (Key, 1966). The term is now used to describe the process by which individuals ‘echo’ views that are

increase ideological segregation in news consumption and lead to polarization in the electorate (Tewksbury & Riles, 2015).

The 'motivated reasoning' and 'confirmation bias' paradigms suggest that citizens will tend to resist information that is inconsistent with prior beliefs and values, and seek out information that confirms them (Kunda, 1990). Individuals, in controlled experiments, choose news articles aligned with their political interests (Iyengar & Hahn, 2009). They are also more likely to share information that conforms to their opinion in their networks and to interact with others that share their opinions (Schkade, Sunstein, & Hastie, 2007). There is evidence that these biases are replicated in online behaviour. Quattrociocchi (2016), for example, attributes decisions to promote a particular narrative on social media to confirmation bias. Other studies concur that the internet and social media reinforce predispositions through selective exposure. Individuals prefer or opt for partisan sources over sources that offer a variety of views (Mutz, 2001) and seek out websites that advance their views (Bimber & Davis, 2003). Aside from selective exposure, algorithms used by search engines, news aggregators, and social media sites 'personalize' a user's experience, and promote information and opinions that conform to and reinforce opinions creating 'filter bubbles' (Dillahunt, Brooks, & Gulati, 2015).

For others, the internet represents the marketplace of ideas or the public sphere realized online. By removing media conglomerates as the gatekeepers of what is newsworthy, the internet and social media open up the potential for citizens to access a wider range of views. Exposure to this wider range of viewpoints may be intentional or accidental. For example, accessing news and information online leads to inadvertent exposure to diverse and sometimes contradictory viewpoints, which, in turn, contributes to heterogeneity in political discussion (Brundidge, 2010). Intentional exposure to more diverse views depends on existing networks and predispositions. Individuals who have more heterogeneous networks seek out more information on diverse topics (Scheufele, Hardy, Brossard, Waismel-Manor, & Nisbet, 2006) and are more aware of opposing viewpoints (Mutz, 2002). This is in contrast to the extreme positions adopted by those in homogeneous networks (Mutz, 2001; Sunstein, 2009).

Thus, there seems to be sound arguments and evidence to support both sides of the debate about the effect of online news consumption on democratic deliberation. However, there is a growing body of evidence to suggest that not all online news consumption is alike. First, many of the studies that demonstrate echo chambers are based on analyzing exposure to and sharing of information on social media platforms. Based on examining patterns of clicking through to and sharing stories on Facebook, Bakshy et al. (2015) conclude that user selectivity, when compared to stories appearing on a news feed determined by algorithms, had a bigger impact on narrowing exposure. Likewise, echo chambers have been demonstrated on Facebook, Twitter, and blogs across a range of topics (Adamic & Glance, 2005; Colleoni, Rozza, & Arvidsson, 2014; Halberstam & Knight, 2016; Quattrociocchi et al., 2016; Williams, Cioroianu, & Williams, 2016). However, there is persuasive evidence that 'weak ties' (connections to those outside ones immediate network of friends), which have not been examined in previous research work, expose users to more heterogeneous views (Barberá, 2014). The extent of agreement among even those within immediate networks may be overestimated, with agreement resulting from projection rather than true similarity (Goel, Mason,

similar to their own, as no other views are selected to enter the 'chamber'.

& Watts, 2010). Further, analysis across a number of countries shows that online news communities are not any more fragmented than other online communities (Fletcher & Nielsen, 2017).

Second, despite this evidence for echo chambers, the question remains: What is the influence of social media on echo chambers relative to other types of online news consumption (directly accessing a news source, search engines, and news aggregators). Furthermore, there is recognition that those that rely on social media are not representative of the general population, thus the extent to which social media echo chambers are representative of news exposure in the population is questionable (Barberá, Jost, Nagler, Tucker, & Bonneau, 2015; Mellon & Prosser, 2017).

Comparing online and offline news consumption, Gentzkow and Shapiro (2011) demonstrate that selectivity leading to ideological segregation is stronger in online news than in offline news consumption, but weaker relative to face-to-face communication (or what those relying on social media refer to as echo chambers). In order to make these comparisons, Gentzkow and Shapiro (Gentzkow & Shapiro, 2011) assess the extent to which browsing histories (data comprised of the URLs users click during online browsing) of users are balanced in terms of the domains that are accessed compared to individual level data from surveys about offline news consumption and face-to-face interactions. There are a number of forces that limit selection into ideological echo chambers: Most online news consumption is limited to centrist domains and users visit many sites (combining centrist and extreme in some cases) (p 1802).

Flaxman et al. (2016) also using detailed information on browsing histories, are able to provide a more nuanced understanding of the conditions under which online news consumption contributes to echo chamber effects. Similar to Gentzkow and Shapiro (2011), they find that most online news consumption results from individuals directly accessing mainstream news outlets. Thus, given that offline media is comprised mostly of mainstream news outlets, online patterns of news consumption are remarkably similar to offline patterns (Flaxman et al., 2016). These direct means of access display less segregation than news reached via social media or search engines. It is unclear whether the relatively high degree of segregation via search engines results from algorithmically generated news feeds or user selectivity. However, the conclusion is that search engines, by whatever process, do contribute to increased segregation to a similar extent as news sharing on social media.

These two studies analyzing browsing histories (clickstream data) focus on an analysis of the domains visited to measure ideological segregation and echo chambers. Both studies (Flaxman et al., 2016; Gentzkow & Shapiro, 2011) rely on a similar method for classifying the ideological orientations of these domains—the average ideological orientations of those visiting the URLs. While this allows ideological classification of a large number of URLs that would otherwise be impossible, we argue that to capture possible segregation or, on the other hand, homophily among news consumers it is necessary to look beyond the domain. Individuals consume stories rather than domains. And so, the pattern uncovered, that a majority of individuals visit a variety of domains, may reflect the behaviour that users follow a story rather than visiting specific domains. Furthermore, the measure of ideology used for URLs may be generated by the browsing behaviour itself rather than reflecting the ideological content of the news stories. Gentzkow and Shapiro (2011), in a robustness check, suggest that segregation at the story level does not vary much from segregation at the domain level (p. 1827-28), However, this is based on examining patterns of URL views during top news events.

Thus, we argue that to fully understand news consumption it is important to take into account the following: 1) capture online browsing rather than social media sharing due to the

unrepresentativeness of a social media only sample; 2) account for the content of news stories consumed rather than rely exclusively on domain views; and 3) have independent survey-based measures of attitudes. In the below section, we describe the study design we employ that integrates respondent browsing history data with a three-wave survey panel. Our analytical strategy is to describe patterns of news exposure among groups of respondents identified by their expressed preference on the EU referendum. Thus, we take a descriptive account of news consumption patterns that we believe is also in the spirit of the earlier work seeking to understand variation in news consumption patterns across platforms, means of access, and subgroups of the population.

2. The Case of Brexit

We examine news consumption during the UK's 2016 EU Referendum. Public debates and first academic contributions have converged, at least initially, on a dominant interpretation of the Brexit vote as a vote of: the 'left behind', the 'outsiders', and, overall, an outcome of the losers from globalization (Antonucci, Horvath, Kutiyski, & Krouwel, 2017; Goodwin & Heath, 2016; Hobolt, 2016). These studies have found that while the "Leave" vote reflects the lack of opportunities across the country, the profile of "Leave" voters is not homogeneous, both with respect to education (Goodwin & Heath, 2016) and to socio-economic conditions (Swales, 2016). This vote appears to be less socially uniform than the popular coverage would concede. However, the socio-economic and psycho-social factors that made voting "Leave" appealing for an ample segment of the British population are still being explored (Vasilopoulou & Wagner, 2017).

Research on the media in the referendum campaign also suggests that understanding the interplay between and consumption of both traditional and social media is necessary. Media coverage during the referendum campaign was characterized as 'divisive and acrimonious', dominated by party leaders, with the economy as the most frequently mentioned topic, and populist undertones used by outlets that supported the "Leave" side (Moore & Ramsay, 2017). On social media, Twitter analysis showed the "Leave" side was more prominent than the voice of "Remain" (Llewellyn & Cram, 2016).

There is ongoing research on the links between the coverage of the campaign in the traditional media, social media, and vote choice that draws on the relationship between media coverage and attitudes toward Europe. Based on this research, there are at least two ways by which media coverage could influence Brexit preferences. First, there is evidence that media coverage can directly influence support for EU integration (Vliegenthart, Schuck, Boomgaarden, Vreese, & H, 2008). Second, there may have been indirect effects of the media. Leader evaluations, in particular evaluations of the leaders of the "Leave" campaign (Nigel Farage and Boris Johnson) that were prominently featured (Clarke, Goodwin, & Whiteley, 2017), influenced the vote, and these evaluations tend to be driven partially by media coverage (Stevens & Karp, 2012).

Our analysis makes three important contributions: First, in terms of the specific case of Brexit, our analysis will further understanding of how the news consumption patterns varied among "Leave" supporters, "Remain" supporters, and those who were undecided; Second, where previous studies of online news consumption have focused on the domain level, we extend our understanding of online news consumption beyond the domain to examine variation at the story level; Third, and more broadly, our analysis links online news exposure to respondent-reported attitudes in order to assess how polarization in news consumption is linked to attitude polarization.

Before moving to our results, we provide detail on the data collection, processing of our web browsing history data, and analysis of news stories.

3. Data and Methods

a. Measuring News Exposure with Surveys and Web Browsing Histories

Some studies rely on self-reported news exposure via various platforms, and then compare how reported exposure to online, social media, and offline media exposure influences various outcomes (for example, (Mellon & Prosser, 2017)). However, self-reported measures tend to over-report some types of news exposure (Prior, 2009) and do not always capture news exposure via social media or searches. Other studies rely on social media exclusively to generate patterns of exposure and draw conclusions about selectivity (for example, (Bakshy et al., 2015)). However, while social media has grown in terms of its prominence as a news platform, the majority of citizens still rely on traditional sources for news and increasingly reach these online (Perrin, 2015).

Our data starts with ICM Research's Reflected Life panel (as it offers distinct advantages for understanding networked online activity), allowing us to track online behaviour and model connections to both media content and other actors on the web. During the UK's 2016 EU Referendum campaign, in collaboration with ICM, we conducted a large-n online survey of close to 1,000 respondents over three waves. The survey was conducted by ICM Unlimited over three waves in 2016, February (3-8), April (22-6), and June (23-24), and included a total number of 1154 respondents.² The use of the survey data permits analysis of subgroups of the electorate based on attitudinal or socio-demographic measures.

In addition to the three-wave panel, our research design includes capture of the digital footprint of respondents over 12 weeks (coinciding with survey field work periods) before the referendum. Members of the panel sample download an app that then tracks the URL of all websites visited, including via Facebook, which we code for content. In the online surveys, we also elicit self-reported social media use. Therefore, the panel allows us a detailed view of a large sample's activity on social media over time, including exposure to information about the EU referendum, while also permitting us to compare online news consumption to expressed attitudes on the referendum.

The clickstream data was collected between February 17 (three days before the EU referendum was announced) and June 23, 2016 (the day of the vote), for a total number of 3,310 users, 959 of which were also present in at least one of the survey panels. So, our analysis is based on the 959 users for which we have survey data and online browsing histories. The data collection periods for the clickstream data are: 7-26 February, 15-30 April, and 1 May thru 23 June (all in 2016), which coincide with the fieldwork for the survey panel waves.³ For the analysis presented in this paper, we combine the three waves of the panel to classify respondents on the basis of referendum preferences.

²In the collection of the survey data, given panel attrition, ICM supplemented each wave with a new sample of respondents. Thus, the analysis is based on respondents who may have been in one, two, or all three of the waves.

³The fieldwork dates for the survey are as follows: 3-8 February, 22-26 April, and 23-24 June. Thus, clickstream data was generally collected following February survey fieldwork and overlapped with the April and June survey fieldwork periods.

b. Clickstream/Browsing History Data Processing

The web browsing data includes all the URL requests made by the user's computer. Opening a new page generates a request which shows up as one URL in the data, but this request is accompanied by multiple others, generating URLs that correspond to ads, widgets, and trackers which are not relevant for our analysis. In fact, for every loaded page containing relevant information (such as a newspaper article), there are on average at least 10 other URLs loading at the same time in the clickstream data, which are irrelevant. Many of these URLs are on the same domain as the page of interest, so the challenge is being able to identify pages that contain articles and distinguish them from irrelevant URLs on the same domain.

To achieve this goal, we started from a list of the most popular news domains, as identified by Alexa (an Amazon company that ranks websites by traffic and classifies them into multiple categories based on content). Within the "News" category, we selected websites which were placed in the "Newspapers", "Analysis and Opinion", "Breaking News", "Current Events", "Extended Coverage", "Internet Broadcasts", "Magazines and E-zines", "Journalism", and "Weblogs" sub-categories, and then selected the top 400 domains in each news category as well as the top domains categorized by UK region. The total number of news domains considered was 4,179. Out of these, 750 domains appeared in our web browsing history/clickstream data.

We then visited each of these domains, manually coded whether the website contained news and information articles or not, and what the structure of URLs pointing to articles on the website was. Knowing the structure of links that point to articles allows us to write regular expressions that match all the possible articles on a domain, while excluding any other types of pages on that domain. Most news websites have a clear subdirectory structure that can be used for this purpose. For example, article pages on the Guardian website have the following structure:

`www.theguardian.com/section/year/month/day/article-title.`

We can, therefore, identify all the Guardian articles that show up in a user's browsing history (and only articles) with the following general regular expression: `www.theguardian.com/.\d{4}/\w{3}/\d{2}/.+`. The coded news domains were further pruned to eliminate those that only included weather and other procedurally generated articles (such as traffic information, sports results, TV programming guides, stocks monitoring pages), news aggregation websites (such as Google and Yahoo News, Flipboard, etc.), videos without an attached article or description, and guides and how-to pages (recipes, reviews, self-diagnosis, travel guides, etc.), which left a total number of 508 domains.

Finally, we merged the identified domains and articles on these domains with the survey data (i.e. the news stories clicked on by our sample of users). The resulting dataset includes a total number of 672 users and 332 domains, which correspond to 5,6289 user-article observations and is a subset of the 959 users for whom we have browsing histories. The original sample of 959 users with browsing histories is reduced by 287 respondents, because these respondents did not visit any of the news and information domains we identified in the initial step of our data processing. While respondents who did not view any news online are of interest when examining offline versus online news consumption, that comparison is beyond the scope of our current analysis. For comparison purposes, of those who viewed no news and information URLs 16% were undecided on Brexit and 35% were "Leave" voters, while 16% of those who did view news URLs were undecided and 43% had a preference for "Leave".

c. Topic Models

To identify topics in our corpus of news stories, we rely on the latent Dirichlet allocation (LDA) model developed in Blei et al. (2003). LDA is a hierarchical Bayesian model of text that assumes that each text in a corpus is a combination of a specific number of topics, and each topic is modelled as a distribution of words over the fixed vocabulary of the corpus. We estimate the model for 50 topics using Gensim (Řadim & Sojka, 2010) and compute average topic probabilities for each article.

d. Full corpus vs. Brexit only

We present results for a sub-corpus of articles that mention the EU Referendum in the text or the title (using multiple phrases and keywords to identify them, such as “Brexit”, “EU referendum”, “UK referendum”, etc.). In our corpus, 1,720 unique articles mention the referendum, and appear 3,679 times in the user-click data. We estimate topic models for each of these two corpuses separately and report average topic probabilities for each article.

4. Results

a. Online News Consumption

Our approach is to provide a descriptive account of how segregated news consumption was during the Brexit campaign and to examine whether news selection generated echo chambers using web browsing data. In the current analysis, we are testing hypotheses about how these patterns of news consumption may have influenced preferences. Our analysis and results proceed in three steps: first, we examine the characteristics of the URLs visited during the campaign; second, we examine the topics of the news stories visited during the campaign by comparing topics of stories shared on social media to topics viewed by our users; third, we examine the network structure of the topics and domains.

In Figure 1, we give a general description of the most visited domains during our final data collection period (1 May through 23 June, 2016) that corresponds to the seven weeks prior to the referendum vote. By a wide margin, with almost 2.5 times more visits than the next most visited site (in particular when combined with BBC.com), the BBC was the most visited online news domain during the campaign period. When examining the stories themselves, the most visited stories were the summary updates of the EU referendum campaign. Thus, the BBC provides a common source of information and is an important force against segmentation and polarization in news consumption behaviour. The other most visited sites represent offline news brands and the only online only (digital native) domain to fall in the top 20 sites is HuffingtonPost.com. Therefore, online versions of offline news outlets are heavily favoured. There are also a number of local news domains that appear in the list of top sites. The Manchester Evening News (part of the Guardian family) features prominently in the list, surprisingly, with more views than the Guardian. We have dropped a single user who was an outlier, and verified that Manchester as a region was not overrepresented in our online survey. Upon further checking, it appears as if the traffic to the Manchester paper is partly driven by links to the Guardian.

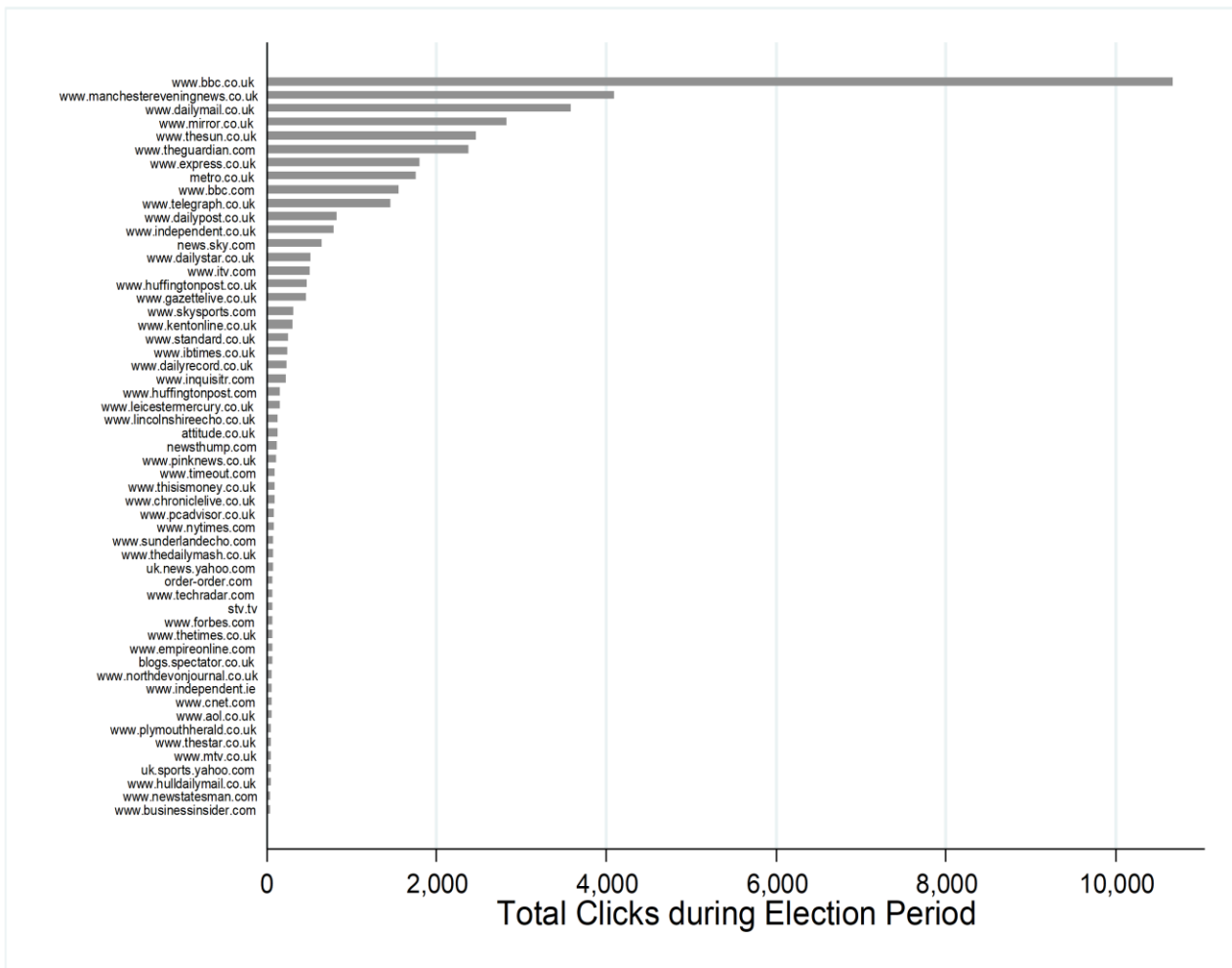


Figure 1. Total Traffic to Online News Domains During EU Referendum Campaign. These data are based on total traffic to identified news/information domains during the 7 Weeks prior to Referendum Vote (1 May - 23 June, 2016) n = 672 users.

One of the other features of the browsing history data we are using to measure online exposure allows us to evaluate the domains visited by other characteristics. One of these features of concern is whether the news and information domains can be classified as "fake news". Prior to reducing the domains to known news and information URLs, we checked the list of all domains against a database of known fake news sites.⁴ Contrary to impressions regarding the spread of fake news via social media and online sharing, there were only four instances of visits to one of the listed fake news sites. The phenomenon of "fake news" may be more prominent in the USA.

Given that public broadcasting consumed a large share of news browsing, does other news consumption show patterns of ideological segregation? We initially used a method of identifying domains as left-leaning or right-leaning as employed by Gentzkow and Shapiro (2011) and Flaxman

⁴The list we relied on was compiled by Lisa Friedland, Kenny Joseph, Nir Grinberg, and David Lazer from NULab at Northeastern University. The list includes almost 500 "fake news" sources (websites) which have been aggregated from existing lists (Hoaxy, BuzzFeed, Media Bias Fact Check, etc.) and verified. The list is updated regularly.

et al.(2016). In order to adjust for the British media market, we used the BBC as our benchmark for a centrist position. We coded domains that had a higher proportion of Labour users than the BBC as to the left and those domains with a lower proportion of Labour voters as to the right of the BBC. However, even after this adjustment, this coding procedure produced some anomalies in our data. The Telegraph, using the measure of partisan leaning based on the partisan makeup of visitors, is to the left of the BBC, as 55% of users who clicked on the Telegraph were Labour voters in GE2015. While The Telegraph is a broadsheet and often considered a 'newspaper of record', its editorial board has links to the Conservative party, and it has been loyal to the Conservative Party in terms of its editorial stance and election endorsements. Therefore, the fact that it would be considered to the left of the BBC on this partisan measure suggests our scale does not reflect content. One of the other main Conservative tabloids, The Sun, is also coded as a left-leaning online source as 59% of its viewers voted for Labour in 2015. Thus the readership, or visitors to news domains, does not reflect the editorial leaning of the paper and what we have observed in past studies about the content. The finding that online visitors to newspaper sites do not reflect the established editorial stance of the newspaper is particularly surprising in the UK, where there is a strong tradition of a partisan press. Consequently, utilizing users of online news domains to identify the 'ideology' of a domain may present issues in other contexts as well.

Therefore, rather than relying on user-defined measures, we have classified domains by their endorsements of political parties in the past general election (GE2015). Using this measure does limit the number of domains for analysis, as we can classify only a subset of newspapers.⁵ However, we do see in Figure 1 that most of the top visited sites were newspapers. Furthermore, using coding of newspaper URLs by endorsements allows us to examine any potential polarization in news consumption by referendum position. We report the average number of clicks on right-leaning and left-leaning news URLs in Figure 2.

We have classified respondents in our survey as either supporting "Leave" or "Remain" (or undecided) based on responses to a question about intended vote in the EU referendum (or how they would vote if a referendum were held which was asked in February 2016 prior to announcement of the referendum). If a respondent appears multiple times in the survey we have taken the response most proximate to the referendum elections. Among "Remain" supporters, averaging the number of views across the right and left leaning newspaper URLs, left-leaning newspaper URLs are only slightly more visited than right-leaning. However, among "Leavers" there is a larger gap between clicking on left versus right-leaning newspaper URLs—showing a higher level of news consumption segregation by "Leavers". They are significantly more likely to view right-leaning newspaper URLs. Among those who remained undecided, we see that they were also more likely to visit right-leaning newspaper URLs, but visited them less frequently than "Leavers'.

We do note that more newspaper URLs fall into the right-leaning group, which contributes to the pattern that all groups of respondents visited right-leaning newspaper URLs at approximately the same rates. Where the difference lies is in how frequently the groups visited left-leaning newspaper URLs, with "Leavers" and "Undecided" respondents having less balanced consumption patterns. We should note we are not making claims about the nature of news consumption on attitudes about

⁵We classify the Sun, Daily Mail, The Telegraph, The Daily Express, and The Standard as right-leaning, and The Guardian and the Daily Mirror as left-leaning.

the EU referendum, but only showing how news consumption patterns vary across users grouped by referendum preferences.

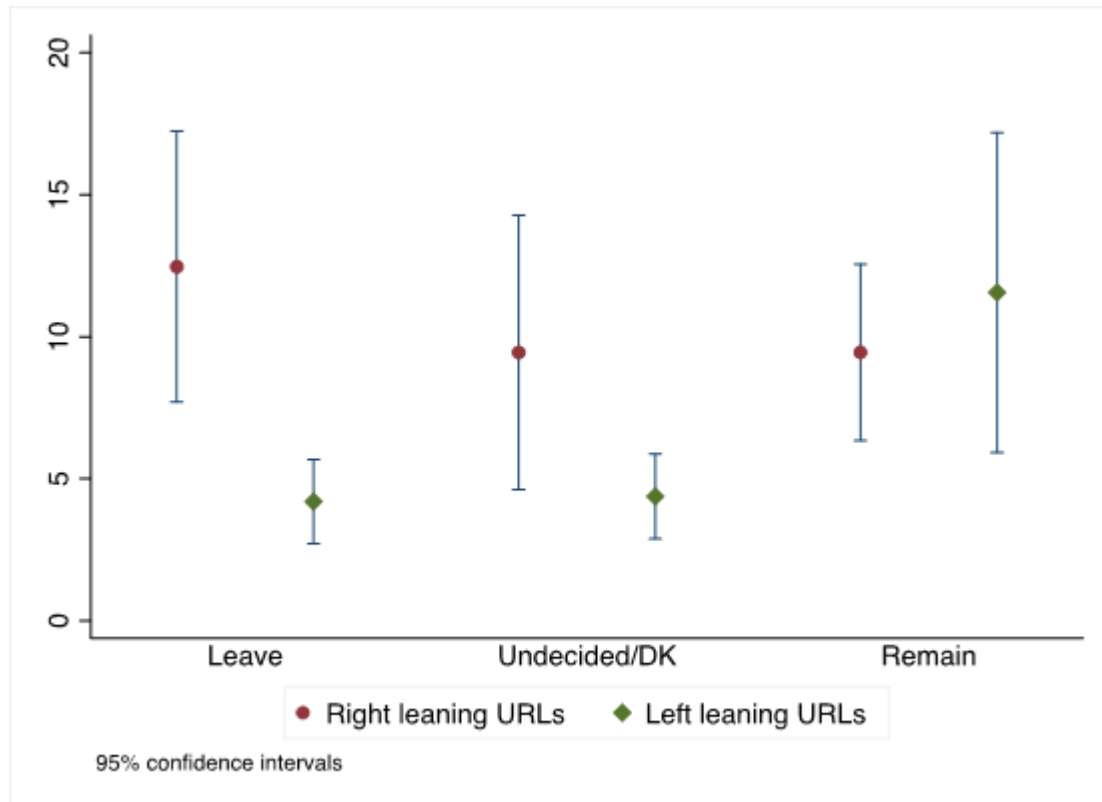


Figure 2. Average Number of Visits to Newspaper URLs by Partisan Slant (identified by endorsements. Note that we identified The Sun, The Daily Mail, Telegraph, The Express as right-leaning, and the Guardian and the Daily Mirror as left-leaning. Y-axis is the average number of visits per user.)

b. Topics at the Domain versus Story Level

The above analysis is only based on newspaper domains that have identifiable editorial stances and could be coded as left- or right-leaning. However, beyond the news domains, we have also coded the topics of the news stories read (or clicked on) by respondents in our survey. Past studies, as in our above analysis, have relied on domain visits rather than actual news stories to measure exposure. But as we argue above, it is important to understand segregation in news topics. Our analysis of the domains clicked on during the campaign show that some domains, such as the BBC, have a broad reach across all respondents. However, the question remains as to whether these users are reading stories about the same topics. For example, in the Brexit debate, the “Leave” campaign tended to focus on the negative impact of EU membership on immigration and transfers to the EU. Did “Leave” voters also tend to read stories about these topics? In Figure 3, we utilize the data generated from extracting topics from the news stories that our respondents clicked on during our period of capturing browsing histories. Figure 3 shows the distribution of topics viewed by our survey respondents.

To think about clustering, we first look at which topics were given more weight by each group of “Leave”, “Remain”, and “Undecided” respondents. As a result, we are able to detect any segregation in the topics of the news stories to which individuals selected to view, rather than relying on the domains to detect segregation. Of the 50 topics extracted from news stories mentioning Brexit, 29

topics had higher associations with “Leave” voters, 21 with “Remain”, and no topics were associated with “Undecided”. Those respondents who supported the “Leave” side had a higher probability of viewing news stories where immigration was a dominant topic. “Remain” voters, on the other hand, had the highest probability of viewing stories where Jo Cox featured (a pro-Remain Labour MP who was murdered by a man associated with the far-right, anti-immigrant organization Britain First). While we have identified which topics tend to be associated with the “Leave” or “Remain” supporters, one of the main conclusions to be drawn is that there is still a great degree of overlap in topics associated with both groups of supporters. In contrast, immigration, single market/trade, internal Conservative Party divisions on Brexit, economic impact, and the polls are five topics where there are the biggest differences between the two groups’ average probabilities of viewing stories with those topics.

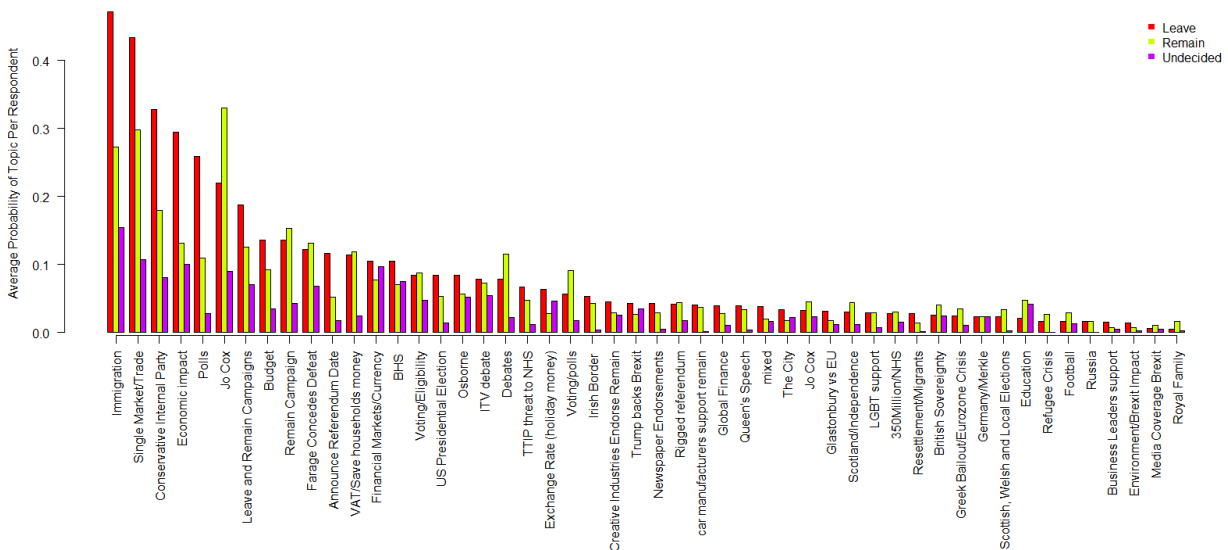
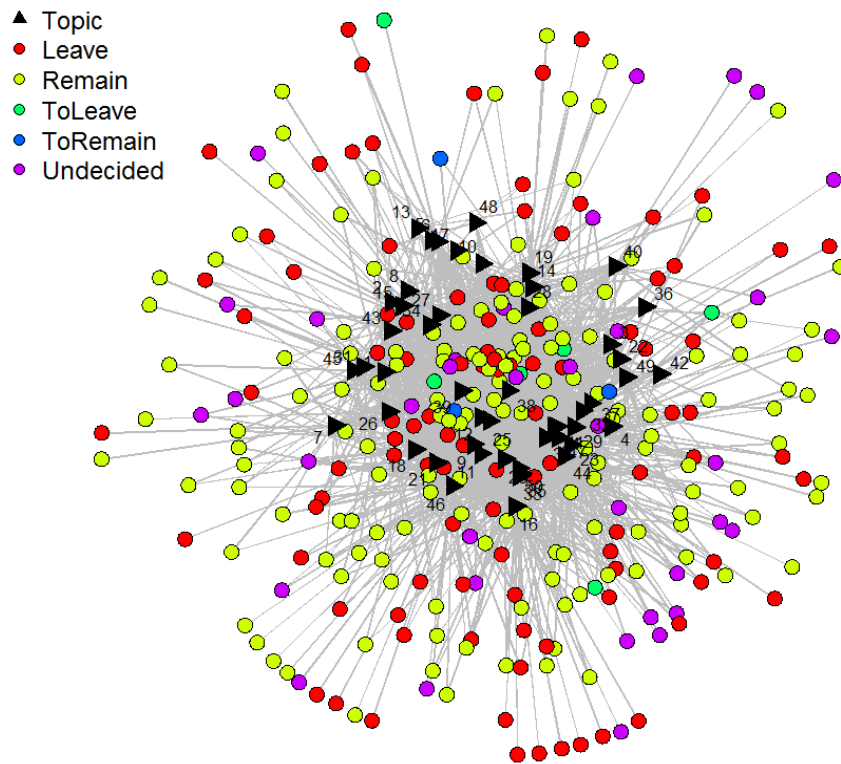


Figure 3. Topic popularity among Leave, Remain, and Undecided Respondents

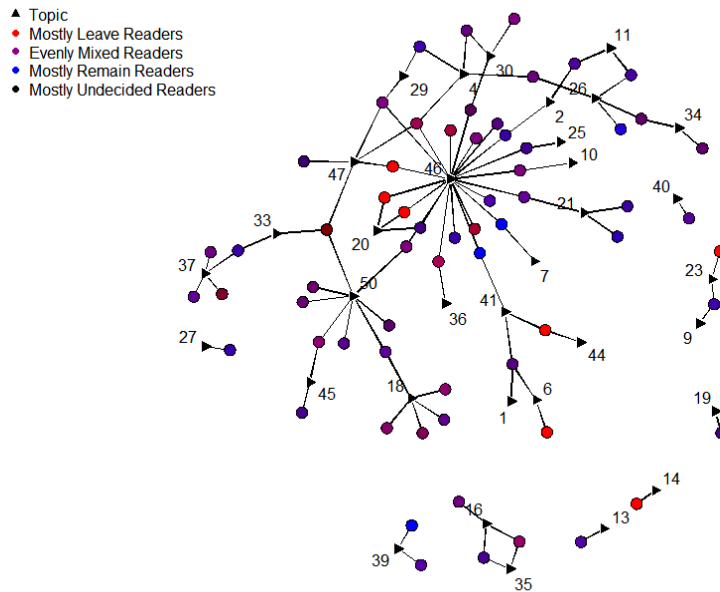
c. Structure of Users: Domains and Topics

In order to compare the degree of segregation at the story and domain levels, we turn next to the relationships among the respondents, the domains they clicked on, and the topics in the articles sourced from the domains. Figure 4 shows the same sets of topics as in Figure 3, but using a network analysis to show the relationships of the topics to the individual user (a) and to the domain (b). Thus in (a) we can examine whether groups of users as identified by EU referendum preferences are clicking on the same topics, and in (b) we can examine whether news domains, with similar profiles of readers, are reporting on the same topics. To make the images readable, all ties to topics that were lower than .2 (a 20 percent confidence level from the model) were removed. Additionally, any topics, respondents, or domains without any resulting ties were also removed. The respondents are coloured by their support for “Remain” and “Leave”, and the domains by the percentage of their supporters in favour of “Remain” or “Leave”. Topics are represented by triangles. If Leavers were all reading the same topics, for example, we should see in (a) the red circles (“Leavers”) linked to the same topics (triangles) with no yellow circles (“Remainers”) connected to these triangles. Likewise, in (b) if domains read mostly by “Leavers” were focused on a set of topics, then we should see the blue circles clustered around a set of triangles. What is evident in Figure 4 for the respondents

shown in (a) is that there is no clustering of topics by whether they supported “Leave”, “Remain”, or were “Undecided”. Wherever multiple domains or multiple readers are linked to a particular topic, that topic has a mixed readership without any clear signs of polarization.



(a) Respondents



(b) Domains

Figure 4. Networks of topics by survey respondents (a) and domains (b)

If there was a clear clustering of topics by vote, we would expect that “Leave” voters mostly read those topics associated with ‘Leave’, and the same for “Remain” voters. We can express this as a block model [?], where each ‘block’ consists of “Leave”, “Remain”, and “Undecided” voters in the rows, and then “Leave” or “Remain” predominantly associated topics in the columns. This idealized image is shown in Figure 5 (a), with the empirical values of the association of the average reader to each type of topic (b). The z-scores in (b) are estimated with a permutation test conditioning on the distribution of emphases on topics. Thus, compared to a world where people randomly clicked on topics (with the same weight as in the empirical data), we see some different results based on “Leave”, “Remain”, and “Undecided” categories. “Leave voters significantly click on predominantly ‘Leave’ topics more than expected (a z-score more than 2 or less than -2 is considered significantly positive or negative, respectively), and were significantly less likely to click on a ‘Remain’ topic—showing evidence of clustering. In the case of “Remain” readers, they were not significantly associated with clicking on either ‘Leave’ or ‘Remain’ associated topics. ‘Undecided’ respondents were significantly less likely than our randomized null model to click on either type of topic. This reinforces what is clear in Figure 4; some topics are vastly more popular among “Leave” voters without the same being true of “Remain” supporters. Even though immigration was a likely topic for both sides and a larger percentage of “Remain” supporters clicked on articles that were associated with this topic (24% vs 18%), once weighting was taken into account—in other words the strength of the association between each article and topic—“Leave” then had a much higher association with the topic of immigration than “Remain” voters (47% tie on average to 27%).

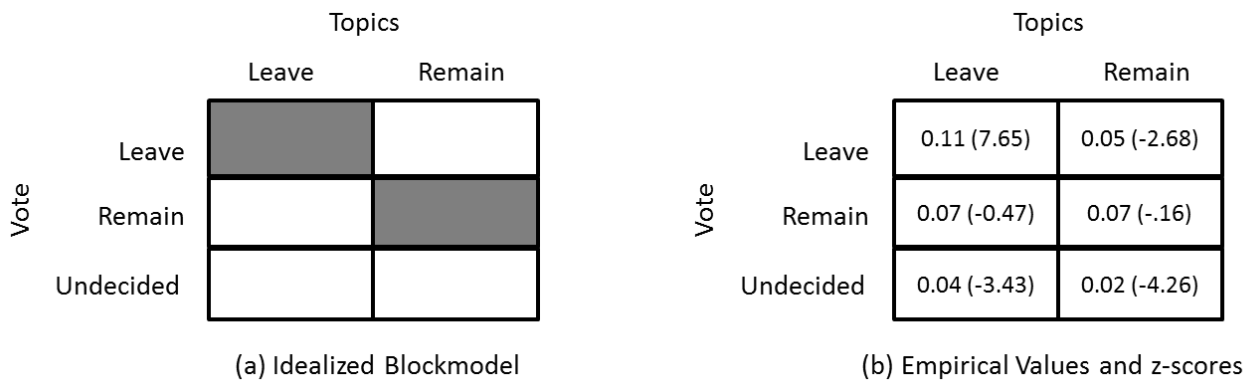


Figure 5. Comparing Clustering: Block Models

5. Conclusions

One of the most salient questions facing established democracies is how the internet has disrupted the flows of communication that keep democracies vibrant (Sunstein, 2004). We address this question by seeking to understand online news consumption by directly observing how individuals select and view articles online, rather than relying on the patterns of sharing news stories on social media platforms. These direct observations have already yielded important insights about the relatively polarizing effects of directly accessing news domains versus social media sharing (Flaxman et al., 2016), and online versus traditional media versus face-to-face contact (Gentzkow & Shapiro, 2011). Our contribution moves the debate a step further by drilling down to the story level and linking this information to direct observations of attitudes collected via an online panel survey.

Our results highlight three conclusions. First, and similar to the recent research in the USA, we show a great degree of overlap in news consumption rather than segregation. In terms of popular URLs, we see that, for example, the BBC – a public broadcasting outlet – was the most visited URL during the Brexit campaign. Therefore, there is a great deal of shared information. Furthermore, by linking web browsing histories with survey responses on political preferences, we see that it was the “Leave” voters who demonstrated a more segregated pattern of news consumption than either “Remain” voters or “Undecided”. This provides a useful condition to our understanding of the polarizing effect of online news consumption; it may only be present for segments of the population.

Second, the online news environment has the potential to stem polarizing forces in the structure of the media ecosystem. One of the defining features of the UK media system is the strength of the partisan press. Indeed, in studies of voting behaviour, a citizen’s chosen newspaper is a strong predictor of partisanship and vote choice. However, through examining online news exposure we see that current online news communities show a great deal of overlap – i.e. left-leaning (or “Remain”) voters click across a range of domains including a number of right-leaning (and pro-“Leave”) online newspapers (The Sun, The Telegraph). For this reason, we also note that measuring URL ideology by user ideology does present some concerns outside of the USA. This measure of the partisan slant of URLs – estimated through the proportion of users who are Labour supporters – yielded results that were inconsistent with content-based measures of ideological slant. However, generally our research is consistent with the conclusion that online news exposure alone does not drive ideological segregation.

Third, our methodological approach shows the advantages of extending the analysis of web browsing history data to the news story level. Previous studies into the extent of polarization and segregation in news consumption have relied on analyzing domains, and this may underestimate the extent of segregation in patterns of news consumption. As above, there is a great deal of overlap in domains accessed. While we find a great deal of commonality in the networks of news exposure even at the story level, there appears to be more segregation at the story level than the domain level.

References

- Adamic, L. A., & Glance, N. (2005). The Political Blogosphere and the 2004 U.S. Election: Divided They Blog. In *Proceedings of the 3rd International Workshop on Link Discovery* (pp. 36–43). New York, NY, USA: ACM. <https://doi.org/10.1145/1134271.1134277>
- Antonucci, L., Horvath, L., Kutiyanski, Y., & Krouwel, A. (2017). The malaise of the squeezed middle: Challenging the narrative of the ‘left behind’Brexit. *Competition & Change*, 21(3), 211–229.
- Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130–1132.
- Barberá, P. (2014). How social media reduces mass political polarization. Evidence from Germany, Spain, and the US. *Job Market Paper, New York University*, 46.
- Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber? *Psychological Science*, 26(10), 1531–1542. <https://doi.org/10.1177/0956797615594620>
- Bennett, W. L., & Iyengar, S. (2008). A New Era of Minimal Effects? The Changing Foundations of Political Communication. *Journal of Communication*, 58(4), 707–731. <https://doi.org/10.1111/j.1460-2466.2008.00410.x>
- Bimber, B., & Davis, R. (2003). *Campaigning online: The Internet in US elections*. Oxford University Press. Retrieved from https://books.google.co.uk/books?hl=en&lr=&id=JGpMCAAAQBAJ&oi=fnd&pg=PT6&dq=Bimber+and+Davis+2003&ots=tf4aShyxq5&sig=COfbfrUrwZsw7PAm0BD_3nYW528
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan), 993–1022.
- Bode, L. (2015). Political news in the news feed: Learning politics from social media. *Mass Communication and Society*, 1–25.
- Boulianne, S. (2015). Social media use and participation: a meta-analysis of current research. *Information, Communication & Society*, 18(5), 524–538. <https://doi.org/10.1080/1369118X.2015.1008542>
- Brundidge, J. (2010). Encountering “Difference” in the Contemporary Public Sphere: The Contribution of the Internet to the Heterogeneity of Political Discussion Networks. *Journal of Communication*, 60(4), 680–700. <https://doi.org/10.1111/j.1460-2466.2010.01509.x>
- Clarke, H. D., Goodwin, M., & Whiteley, P. (2017). Why Britain Voted for Brexit: An Individual-Level Analysis of the 2016 Referendum Vote. *Parliamentary Affairs*, 70(3), 439–464. <https://doi.org/10.1093/pa/gsx005>
- Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal of Communication*, 64(2), 317–332.
- Dillahunt, T. R., Brooks, C. A., & Gulati, S. (2015). Detecting and Visualizing Filter Bubbles in Google and Bing. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 1851–1856). New York, NY, USA: ACM. <https://doi.org/10.1145/2702613.2732850>

- Downs, A. (1957). An Economic Theory of Political Action in a Democracy. *Journal of Political Economy*, 65(2), 135–150.
- Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, 80(S1), 298–320.
- Fletcher, R., & Nielsen, R. K. (2017). Are News Audiences Increasingly Fragmented? A Cross-National Comparative Analysis of Cross-Platform News Audience Fragmentation and Duplication. *Journal of Communication*, 67(4), 476–498. <https://doi.org/10.1111/jcom.12315>
- Gentzkow, M., & Shapiro, J. M. (2011). Ideological Segregation Online and Offline. *The Quarterly Journal of Economics*, 126(4), 1799–1839. <https://doi.org/10.1093/qje/qjr044>
- Goel, S., Mason, W., & Watts, D. J. (2010). Real and perceived attitude agreement in social networks. *Journal of Personality and Social Psychology*, 99(4), 611.
- Goodwin, M. J., & Heath, O. (2016). The 2016 Referendum, Brexit and the Left Behind: An Aggregate-level Analysis of the Result. *The Political Quarterly*, 87(3), 323–332. <https://doi.org/10.1111/1467-923X.12285>
- Halberstam, Y., & Knight, B. (2016). Homophily, group size, and the diffusion of political information in social networks: Evidence from Twitter. *Journal of Public Economics*, 143, 73–88.
- Hobolt, S. B. (2016). The Brexit vote: a divided nation, a divided continent. *Journal of European Public Policy*, 23(9), 1259–1277. <https://doi.org/10.1080/13501763.2016.1225785>
- Iyengar, S., & Hahn, K. S. (2009). Red media, blue media: Evidence of ideological selectivity in media use. *Journal of Communication*, 59(1), 19–39.
- Iyengar, S., & Simon, A. F. (2000). New Perspectives and Evidence on Political Communication and Campaign Effects. *Annual Review of Psychology*, 51(1), 149–169. <https://doi.org/10.1146/annurev.psych.51.1.149>
- Key, V. O. (1966). *The responsible electorate*. Belknap Press of Harvard University Press.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480–498. <https://doi.org/10.1037/0033-2909.108.3.480>
- Llewellyn, C., & Cram, L. (2016). Brexit? Analyzing Opinion on the UK-EU Referendum within Twitter. In *ICWSM* (pp. 760–761).
- Mellon, J., & Prosser, C. (2017). Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users. *Research & Politics*. <https://doi.org/10.1177/2053168017720008>
- Moore, M., & Ramsay, G. (2017). *UK media coverage of the 2016 EU referendum campaign*. King's College London.
- Mutz, D. C. (2001). Facilitating Communication across Lines of Political Difference: The Role of Mass Media. *American Political Science Review*, 95(1), 97–114.
- Mutz, D. C. (2002). Cross-cutting Social Networks: Testing Democratic Theory in Practice. *American Political Science Review*, 96(1), 111–126. <https://doi.org/10.1017/S0003055402004264>
- Perrin, A. (2015). Social media usage. *Pew Research Center*, 52–68.
- Prior, M. (2009). Improving Media Effects Research through Better Measurement of News Exposure. *The Journal of Politics*, 71(3), 893–908. <https://doi.org/10.1017/S0022381609090781>
- Quattrociocchi, W., Scala, A., & Sunstein, C. (2016). *Echo Chambers on Facebook* (SSRN Scholarly Paper No. ID 2795110). Rochester, NY: Social Science Research Network. Retrieved from <https://papers.ssrn.com/abstract=2795110>
- Quintelier, E., & Vissers, S. (2008). The effect of Internet use on political participation an analysis of survey results for 16-year-olds in Belgium. *Social Science Computer Review*, 26(4), 411–427.
- Řehůřek, R., & Sojka, P. (n.d.). Software Framework for Topic Modelling.

- Scheufele, D. A., Hardy, B. W., Brossard, D., Waismel-Manor, I. S., & Nisbet, E. (2006). Democracy Based on Difference: Examining the Links Between Structural Heterogeneity, Heterogeneity of Discussion Networks, and Democratic Citizenship. *Journal of Communication*, 56(4), 728–753. <https://doi.org/10.1111/j.1460-2466.2006.00317.x>
- Schkade, D., Sunstein, C. R., & Hastie, R. (2007). What Happened on Deliberation Day Essay. *California Law Review*, 95, 915–940.
- Stevens, D., & Karp, J. A. (2012). Leadership traits and media influence in Britain. *Political Studies*, 60(4), 787–808.
- Sudulich, M. L. (2011). Can the internet reinvent democracy? *Irish Political Studies*, 26(4), 563–577.
- Sunstein, C. R. (2004). Democracy and filtering. *Communications of the ACM*, 47(12), 57–59.
- Sunstein, C. R. (2009). *Going to extremes: How like minds unite and divide*. Oxford University Press.
- Swales, K. (2016). Understanding the Leave vote. *NatCen Social Research*.
- Tewksbury, D., & Riles, J. M. (2015). Polarization as a Function of Citizen Predispositions and Exposure to News on the Internet. *Journal of Broadcasting & Electronic Media*, 59(3), 381–398.
- Tucker, J., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., ... Nyhan, B. (2018). Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature.
- Valkenburg, P. M., Peter, J., & others. (2013). Five challenges for the future of media-effects research. *International Journal of Communication*, 7, 197–215.
- Vasilopoulou, S., & Wagner, M. (2017). Fear, anger and enthusiasm about the European Union: Effects of emotional reactions on public preferences towards European integration. *European Union Politics*, 18(3), 382–405.
- Vliegenthart, R., Schuck, A. R. T., Boomgaarden, H. G., Vreese, D., & H, C. (2008). News Coverage and Support for European Integration, 1990–2006. *International Journal of Public Opinion Research*, 20(4), 415–439. <https://doi.org/10.1093/ijpor/edn044>
- Williams, M. J., Cioroianu, I., & Williams, H. T. (2016). Different News for Different Views: Political News-Sharing Communities on Social Media Through the UK General Election in 2015. In *News@ ICWSM*. Retrieved from <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/download/13232/12850>

About the Authors

Susan Banducci

Susan Banducci is a Professor in Politics (University of Exeter). Her research interests are in the areas of comparative political behaviour, media, and political communication. She is the director of the [Exeter Q-Step Centre](#), which advances quantitative methods in the social sciences.

Iulia Cioroianu

Iulia Cioroianu (PhD, NYU) is an Associate Lecturer in the Q-Step Centre and a Research Fellow. Her research focuses on electoral competition, candidate and party positions, campaign dynamics and social media, and makes use of a diverse set of methods such as econometric analysis, machine learning, quantitative text analysis, and experiments.

Travis Coan

Travis Coan (Ph.D., Claremont Graduate University, 2011) is a Lecturer in Quantitative Political Science and an academic staff member at the [University of Exeter's Q-Step Centre](#). Before joining the Politics Department at Exeter, Travis was a Statistician and Lecturer on Law at Harvard Law School, where he directed the Empirical Research Services group and taught applied quantitative analysis. While Travis' research spans a diverse set of topics in political science, his most recent work applies advances in text-

mining and natural language processing to examine the relationship between climate change scepticism, public opinion, and environmental policy.

Lorien Jasny

Lorien Jasny (PhD, UC Irvine) is Lecturer in Quantitative Political Science at the Exeter Q-Step Centre. Her research agenda focuses on two related themes: how the structure and dynamics of inter-organizational networks affect policy change, and how the structure and dynamics of belief networks affect behavioural change. Substantively, she studies how people try to bring about societal change in response to political and environmental concerns.

Dan Stevens

Dan Stevens is a Professor in Politics (University of Exeter, Cornwall). His main interests are in mass political behaviour in the United States and Britain. His work examines major influences on political attitudes and behaviour, such as the economy, political advertising, and the news media. Current projects include ongoing research into perceptions of political advertising, patterns, and effects of different forms of mobilization in elections, as well as the role of leaders in British elections.

Hywel Williams

Hywel Williams (PhD, Leeds) holds a position in Computer Science (University of Exeter), where he is a member of the Institute for Data Science & Artificial Intelligence. His research career has applied complex systems thinking and computational methods to problems in artificial intelligence, Earth system science, ecology, evolution, and more recently, social sciences.

Iain Weaver

Iain Weaver (PhD, Southampton) is a Senior Researcher (University of Exeter, Computer Science). His research interests are focussed around self-organisation of systems of many interacting components as viewed through the lens of thermodynamics and statistical physics. His research has appeared in journals such as *Social Networks*.