Illuminating the ecosystem of partisan websites

Shweta Bhatt
King's College London
London, UK
shweta.bhatt@kcl.ac.uk

Sagar Joglekar
King's College, London , Department of Informatics
London, UK
sagar.joglekar@kcl.ac.uk

Shehar Bano
University College London
London, UK
s.bano@ucl.ac.uk

Nishanth Sastry
King's College, London , Department of Informatics
London, UK
nishanth.sastry@kcl.ac.uk

ABSTRACT

Alternative news media ecosystems thrive on growing popularity of social media networking sites and thus have been successful in influencing opinions and beliefs by false and biased news reporting. The 2016 U.S. elections are believed to have suffered enormously due to hyper-partisan online journalism. Within this context, this paper aims at finding specific evidences of hyper-partisan clusters and key characteristics that mediate the traffic flow within the partisan media ecosystem. This is achieved by analyzing a data set consisting of a curated list of 668 partisan websites and 4M Facebook posts across 507 corresponding Facebook pages to understand the stake of partisan media in influencing U.S. politics. The paper successfully points out the extensive internal traffic forwarding within partisan sites, illustrates how the web and social media strengthen the political divide between the left and the right, finds temporal evidences of strong involvement of partisan sites during 2016 U.S. elections and discusses the characteristics of their target audiences.

CCS CONCEPTS

• **Social and professional topics → Technology and censorship**; *Political speech*; • **Human-centered computing →** *Social networking sites*;

KEYWORDS

Exploratory Data Analysis, Political polarization, Partisan media

ACM Reference Format:
Shweta Bhatt, Sagar Joglekar, Shehar Bano, and Nishanth Sastry. 2018. Illuminating the ecosystem of partisan websites. In *Proceedings of ACM Web conference (The Web Conference),* Jennifer B. Sartor, Theo D'Hondt, and Wolfgang De Meuter (Eds.). ACM, New York, NY, USA, Article 4, 9 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

The Web Conference, April 2018, Lyon, France

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The rise of the Internet and social media platforms has given a new dimension to the democratic process. From a political science perspective, the idea of open participation platforms like blogs and social media websites have been shown to have high impact in the democratic process [13] and may well be instrumental in shaping the elections of the future, as demonstrated by the 2016 US elections. As such, it is crucial for political and social scientists to assess the modalities and attitudes towards partisan media, and to understand their modus-operandi.

So far, research has focused on evaluating the role of *social* media and platforms like `Reddit` and `4chan` in the process of sharing mainstream and alternative news on the web and the lexical value of the language used by fake-news content [5, 9, 12, 15, 20, 24]. This research has revealed a deep divide between views of the so-called 'right' or 'conservatives', and the 'left' or 'liberals'. However, we do not have a similar understanding of the ecosystem of partisan *news* websites. For example – Who are the producers of partisan news, and what might their motivations be? How do such websites interlink with each other and rest of the web? Is there a similar divide betwen the right and the left in the supposedly professional news media as found in social media expressions?

In this work, we answer these questions by studying an ecosystem of 668 hyper-partisan[1] websites and associated Facebook pages shared with us by Buzzfeed News as part of a recent collaborative investigation connected with the 2016 US Elections[2]. Building on this dataset, we collect data from Alexa[3] about the size of the traffic to these sites, from Facebook about the structure of connections (likes, etc) between the different Facebook pages, and from URL shorteners like `bit.ly` about the clickthrough rates for short URLs associated with news stories. Using this data, we aim to understand traffic forwarding dynamics, audience engagement, ownership and demographics of hyper-partisan websites.

Taking advantage of reliable manual coding of websites into 'left' and 'right' by our Buzzfeed collaborators, we examine partisan

[1]In the rest of this paper, we interchangeably use the words *partisan* and *hyper-partisan* in connection with websites that present one side of a story. Operationally, the judgement of whether a news site is 'partisan,' and if so, whether it is right- or left-leaning, was made by journalists at Buzzfeed as part of a careful manual coding process (c.f. §3). Our analysis takes these as given. We note in passing that hyper-partisan news simply presents an unbalanced picture of reality, and is different from fake news (although it is possible that some partisan news sites manufacture 'fake' stories, this is outside the scope of our analysis).
[2]A part of this paper includes work conducted for Buzzfeed as part of the collaboration
[3]http://www.alexa.com

behaviour on both sides of the political spectrum. Our central result is that although there is more volume (i.e., larger number of sites) on the right, there is similar partisan behaviour on both the right and left. For instance, in contrast with more mainstream news sites, there is a disproportionately large amount of linking between hyper-partisan sites leading to traffic referrals. Strikingly, these referrals mostly stick to one side of the political debate – right-leaning websites refer traffic to other right-leaning sites; left-leaning sites link to left-leaning ones. In other words, among these *producers* of hyper-partisan news, we find echo chambers similar to those which have been widely discussed (e.g., [7]) among *consumers* on social media. Unsurprisingly, we find a similar right-left cliquishness when we study the corresponding Facebook pages of these news sites, and ask which Page is 'liked' by which other page. Completing the picture, the audience of the Facebook pages are similarly divided across right-left lines, which is closer to the traditional notion of echo chambers in social media.

Examining the domain ownership records, we find evidence that hyper-partisan sites were strongly influenced by the US Election cycle – a remarkable number of new sites, especially right-leaning ones, were registered in the run-up to the elections. Several are from Macedonia, providing independent confirmation of the role of Macedonian teenagers in propagating false information [21]. Similarly, a disproportionate number of these sites also "die out" and lose popularity in Alexa Traffic Rankings in the period between the Nov elections and President Trump's inauguration.

This "birth" and "death" of website domains is stronger amongst right-leaning sites, suggesting that the American Right or Conservative populace was more strongly influenced by hyper-partisan sites during the election. Complementary to this, we examine the demographics of the audience of hyper-partisan sites and find that Americans who visit these hyper partisan sites are more right-leaning: Sampled mean demographic audience sizes for some of the most populous demographic groups of USA (e.g., white caucasians, or middle income ranges) are higher for right-leaning sites.

Collectively, these results shed first light on an ecosystem of websites that affected the 2016 US Elections by presenting a one-sided view. The rest of this paper is structured as follows. §2 places our paper in the context of related work. §3 describes the dataset provided to us by Buzzfeed and our additional data collection efforts with Alexa, bit.ly, and Facebook. §4 examines the websites studying the characteristics of its owners and the evolution of traffic to these sites. §5 presents a complementary picture, studying the demographics of the *audience* of these sites. §6 examines partisan traffic flows describing the phenomenon of internal traffic referrals. §7 studies demographics and traffic flows from the perspective of Facebook pages associated with these sites, finding similar evidence of internal referrals and increased activity on the right. §8 concludes.

2 RELATED WORK

Previous studies illuminate various dimensions of the fake news ecosystem. We contextualize our work by providing a brief discussion of key themes in this area. Our work augments previous literature by presenting a data-driven analysis of the producers and consumers of partisan websites, and the traffic forwarding behavior that underpins their sustainability and supports on-line echo chambers.

Attitudes towards Partisan news. Classically political and social science heavily relies on survey-based inferences. There have been important studies on partisan news media in the US over the past few years. These methods help the community get important insights into the attitudes of audiences towards media [11] and the role of social media in news dissemination [10]. However, what they lack is an understanding of how the web is shaping these attitudes and trends, and what role social networks play in facilitating spread of partisan information. A recent work [1] provides insights into the role of fake news in US elections. The authors use interesting sources of data, such as a crowd-driven website `Snopes.com` that refutes fake rumors and conspiracy theories. They further find predictors for believability on face news. Their study is highly selective towards the fake news phenomenon, and ignores the effect of partisan but established news sources on the web and on Facebook such as *Breitbart*, *Infowars*, *ThinkProgress*, *OccupyDemocrats*. We find in our study that these sources have big enough presence to shape traffic and grab click-through rates as high as 4% on their posts. Another work[2] suggests that a population that consumes news online has a greater tendency to be polarized than those who do not read online news or consume news through television.

Filter bubbles and echo chambers. Echo chambers refer to the phenomenon when people do not subscribe to views that do not agree with their own. Coupled with recommendation system algorithms, a self-reinforcing ecosystem emerges where users get fed information that suits their opinions[6]. Analyses of recommendation systems for Facebook users have revealed interesting signals of both the presence of filter bubbles and awareness among some users about their existence [18]. Recent work [7] explores breaking the polarized echo chambers by exposing users to opposing sides of discourse.

Fake news and social media. The issue of fake news and conspiracies has received considerable attention since the 2016 US Presidential election. Various studies [5, 9, 12, 15, 24] have explored the role of social media and platforms like Reddit and 4chan in the process of sharing mainstream and alternative news on the web—but these focus on the lexical value of the language used by fake-news content. Starbird et. al [20] explore the alternative media domain networks by analyzing 58M total tweets related to mass shooting events for a period of about 10 months beginning in January 2016, and provide a qualitative analysis of how alternative news sites propagate and promote false narratives while mainstream media refrains from such behavior. They show the presence of domain clusters that control the flow of information with the alternative news ecosystem and their political polarization. As the scope of this work is limited to the Twitter activity of influential alternative news websites, it does not capture the referrer traffic patterns between social network pages or the actual web pages. Our work augments the findings of this paper using other modalities of data.

3 DATASET DESCRIPTION

Our primary dataset comprises a curated list of partisan websites from Buzzfeed [4]. We augment this data using open APIs like Alexa [4] and Facebook Graph API [5].

Partisan websites. To be classified as a partisan site, the curators at Buzzfeed considered whether the site covers American politics, is a big content-generator, publishes opinionated articles and is not a mainstream news website. Additionally, the associated Facebook pages were identified for each added website. The website's right or left alignment was decided by manually looking at its self-declaration on either its web page or on its Facebook page. The data set contains information such as-registration date, owner, political category, analytics and google adsense codes of 668 unique partisan sites including 489 conservative and 179 liberal sites. Among the 668 partisan websites, 8 websites including few popular ones such as *Infowars*, *DailyCaller* and *WashingtonJournal* were suggested by us, leveraging the Alexa upstream sites data that we collected (see below) for an initial list of partisan sites. These websites were found to be among the top-5 upstream sites of the initial list of partisan sites which prompted BuzzFeed to incorporate them as a part of their final list. Using the partisan dataset as a starting point, we collected available Alexa metrics for each site from the dataset (Table 1).

Mainstream News Sites. Mainstream media consists of the large news conglomerates that influence a vast readership and their political news coverage is not *formally* announced to be in alignment with a specific political category or ideology. For comparison purposes, we collected upstream site information for top 500 news websites ranked by Alexa such as *cnn.com*, *nytimes.com*, *theguardian.com*, *washingtonpost.com*, etc.

Alexa Metric	Size(no. of sites)
Top-5 upstream sites	484
Ranks	570
Rank charts	586
Audience Demographics	421

Table 1: Collected Alexa Dataset

Alexa Dataset. Alexa is an internet portal which provides meaningful web analytics including website traffic statistics, site comparisons, and website audience. Alexa's measurement panel is based on a diverse set of over 25,000 browser extensions and plug-ins used by millions of people [16][6] To gain further insights about the traffic flow patterns of the partisan sites, we crawled some key website metrics from Alexa between June–July 2017. Table 1 provides details about the data collected from Alexa for partisan websites. Additionally, we collected the Alexa Ranks and top-5 upstream sites for the 478 mainstream new websites. An upstream site is a website that the user visits before they visit the current website.

Although Alexa's data collection methodology may lead to some biases, it is believed to be the best available option for demographic and traffic information on the Web [22]. Alexa provides a basic notion of reliability by filtering data and only reporting a statistic when it believes sufficient data is available. Our choice of using Alexa is based on these factors.

[4]https://www.alexa.com/about
[5]https://developers.facebook.com/docs/graph-api/
[6]Alexa does not reveal the full methodology it uses to collect data, but notes that the current implementation extends beyond understanding that it is based on a popular browser toolbar: c.f.: https://blog.alexa.com/top-6-myths-about-the-alexa-traffic-rank/

Metric	Total	Average
# Fans	326M	0.67M
# Posts	4.1M	8514
# Post Reactions	0.59M	1163
# Post Comments	66924	132
# Post Shares	0.3M	605

Table 2: Facebook Dataset Overview

Facebook. Facebook is a leading online social media networking platform. Popular media entities these days maintain Facebook pages to propagate their posts on the Internet and direct people to browse more published content on their website and thus engage with a global audience. We analyze the Facebook pages owned by the partisan sites and engagement metrics related to the public posts published on those pages. Facebook dataset by Buzzfeed contains engagement information such as reactions, comments and status types of 4M posts from 507 partisan Facebook pages posted between January 1st, 2015 and March 31st, 2017. This dataset was the entry point of our analysis. In order to dig deeper into our investigation we augmented the Facebook dataset by collecting page metrics for each of the 507 pages using Facebook's Graph API including *category*, *fan_count*, *pages that follow these pages*, *start_info* of the pages, *websites* associated with the pages and their *verification status*. We also collected random samples of 1000 active *user followers* for top 100 most popular partisan pages. Table 2 provides an overview of the Facebook statuses dataset.

Bitly. Through some initial analysis of the Facebook dataset we discovered that about 6% of the 4M Facebook posts were short URLs. In order to study the relationship between their click through rates and Facebook post engagements we collected the cumulative clicks count data for a sample comprising of 10% of all the short URLs across 216 pages using the Bitly API.

4 PRODUCERS OF PARTISAN NEWS

We first look at the websites, by examining domain registrations. We look at who the owners are and when the websites were registered, relative to the US elections. We find several owners who run groups (more than one) of partisan websites, and a number of websites which appear to have been created specifically for the US elections, registered in the run up to the elections, and then losing popularity immediately after.

Who are the owners. We begin by asking who are the owners of these websites, using the ownership records of the website domain names as returned by WHOIS queries. We find that 78 (11%) of the 668 partisan sites in our list are run by Macedonian owners, further confirming the story of the macedonian teens and young men who ran successful partisan websites[21]. These are mostly conservative sites (74 conservative, 4 liberal). Additionally, we discovered that at least 12 U.S. based firms manage multiple websites and associated Facebook pages as listed in table 3.

Largely, we observe that most of the owner companies have overt political affiliations i.e. they either handle collections of conservative or liberal sites. One prominent exception to this is *Today's Growth Consultant* that runs 5 conservative sites including *My Right America* and *Red White and Right* along with *Progressive Liberal*

Owner Companies	No. of websites owned
Discount Book Distributors	22 right
Addicting Info	8 left
DB Capitol Strategies	6 right
Today's Growth Consultant	5 right & 1 left
Salem Media Group	5 right
Power Publisher	4 right
Media Research Center	4 right
Liftable Media	3 right
News Corpse	3 left
Counter Punch	2 left

Table 3: Top companies that own partisan websites

which is a left leaning site. This suggests that the motive behind running such collections of partisan sites cannot solely be attributed to propaganda but might also include financial incentives (e.g., through ad traffic).

Birth and death of partisan websites. Next, we analyse the role played by the US election timeline on the production of partisan content. We map the birth of these websites by using the date of the domain name registrations (Figure. 1). About 28% of the websites were registered in the election year of 2016, making it the year with the most number of domain name registrations for the sites in our dataset

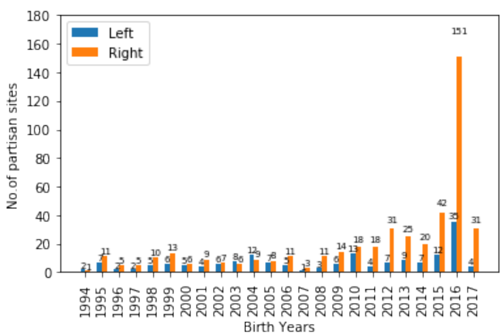


Figure 1: ***Distribution of Domain Registrations of partisan websites***

Among this lot, 81% websites support conservative views while 19% are liberal pointing to an increased activity of partisan news on the conservative side. Note that this activity is disproportionately large considering that about 73% of the sites in our overall dataset are conservative.

To study the 'death' of websites, we acquired the longitudinal Alexa rank numbers for the websites spanning across the period of one year. According to Alexa, the ranks data is derived from the traffic data provided by the users in Alexa's Global Data Panel over a rolling 3 month period and is based on the browsing pattern of people in Alexa's Global Data Panel.

Figure 2 provides the distribution of the time of peak popularity for the partisan sites, which we consider as the time of the 'death' of the website. About 31% liberal and 40% conservative sites had highest individual Alexa Ranks between Nov 2016 (when the election took place) to Feb 2017 (date when Trump moved into the White House). Again, we note that a disproportionately large number of conservative websites peaked and declined in popularity after this period.

These results indicate that the traffic to these hyper partisan sites is co-incident with the US elections. There could be several possible explanations. For instance, it may be that some websites were

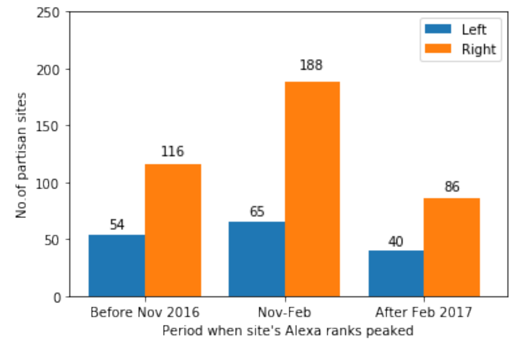


Figure 2: ***Temporal popularity of Partisan websites***

run by foreign actors or other agents with an explicit interest in promoting partisan news during the US elections, but stopped promoting their sites after the elections as it was no longer necessary. A second possibility is that the website owners were financially motivated and the sites were no longer as valuable after the elections. Alternately, the consumers of this content were more interested during the closely fought elections and lost interest afterwards. Our data can only observe the correlation of the website registrations and peak popularity with the election cycle, and cannot distinguish between these different possibilitys.

5 CONSUMERS OF PARTISAN NEWS

Having considered the producers, we next seek to understand who are the consumers. To this end, we analyzed demographics data for partisan websites collected using the Alexa API, which also provides information about how the reported statistics compare with the general Internet population. We were able to collect demographics data for 63% (421) partisan websites[7], of which 69% (290) are conservative and the remaining 31% (131) are liberal. Because of the differences in the numbers between the left and the right, we calculate the average the demographic scores for the left and the right by sampling equal numbers of sites, and reporting the mean of this sample.

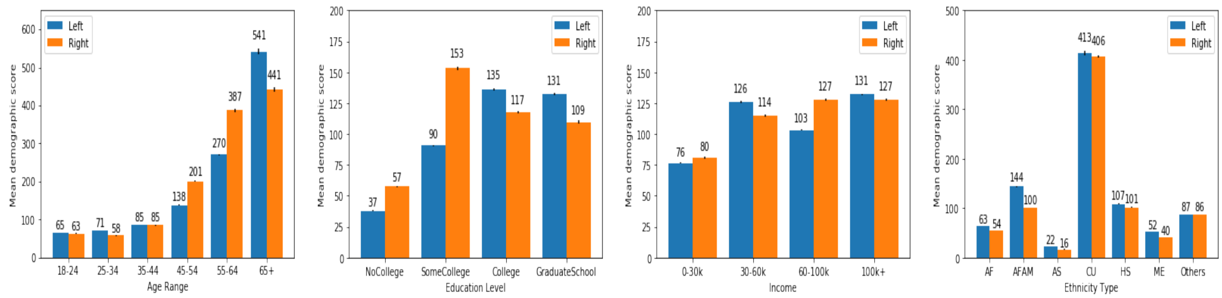
The demographics results are only as reliable as Alexa's reported demographics. Alexa states that demographics data is gathered from voluntary demographics information submitted by people in their Global Traffic Panel available as a browser extension[8]. Although this may cause biases, we rely on the guarantee that Alexa only reports a demographic score when it believes sufficient data is available. The scores reported by Alexa are normalised as shown in Table 4 into the relative scale.

Range	Audience Representation
0–50	Greatly under-represented
50–90	Under-represented
90–110	Similar to the general Internet population
110–200	Over-represented
200–Infinity	Greatly over-represented

Table 4: Audience Demographics Scale

[7] Alexa does not have enough data about the remaining 247 sites. This is partly because 114 of the 247 sites were only launched in 2016/17.

[8] https://support.alexa.com/hc/en-us/articles/200449744-How-are-Alexa-s-traffic-rankings-determined-



(a) (b) (c) (d)

Figure 3: Demographic insights into audience of partisan websites, shown as bar plots with standard error. X-axis shows demographic bracket; Y axis shows sampled mean demographic score, using the scale in Table 4. 3a: Age Fig.3b: Education ,3c: Income. 3d: ethnicity. The ethnicities are coded as AF: African, AF/AM: African-American, AS: Asian, CU: Caucasian, HS: Hispanic and finally Others.

We focused our analysis on 5 aspects of demographics namely age, education, ethnicity, income and gender. We visualize these patterns through a series of bar plots. The X axis represents the demographic bracket and the Y axis the sampled a mean audience count (normalised to 100 for the general Internet population) across all the left- or right- leaning websites. All bar plots are shown with standard error.

As expected from other surveys [19] that suggests younger people are less politically engaged, Figure 3a shows that those aged 45 and younger are under-represented relative to the rest of the Internet population, and those above 45 over-represented. Interestingly, the under-represented populations are nearly equally left- and right-leaning, but the difference between left and right also grows with age. Partisan news consumers amongst the young (25 to 34) and the oldest (65+) age groups tend to be more liberal, while the middle-aged and older consumers (45 to 64) are more conservatively inclined.

Figure 3b shows that people with no college education do not engage much with partisan news. Those with the most common educational attainment level (some college) have significantly higher affinity with right-leaning sites than left-leaning ones. On the other hand, people with a higher educational background (those who have completed a college or graduate degree) are active consumers of partisan websites, and tend to be left-inclined. We also notice that people with very high levels of education (Graduate School) are less represented than those with some college [14].

As expected from the SocioEconomic Status theory [23], Figure 3c indicates that people with economically weaker backgrounds do not tend to participate politcally and are overall less represented (i.e., have lower demographic scores) on partisan websites. We observe that people in the \$30-60k and \$100k+ income brackets are over-represented with nearly equal distributions of left-leaning and right-leaning readership. However, the upper middle-class audience (with income in the range \$60-100k) is more inclined towards conservative news. This is the most common income range in the USA and includes roughly 1 in 3 households in the USA[9].

Figure 3d shows the racial distribution of consumers of partisan websites. Caucasians represent an overwhelming majority of consumers as expected. Caucasian, Hispanic and Other audience is somewhat equally divided between left-wing and right-wing. African-Americans are found to be more left-leaning. We note that African, Asian, and Middle Eastern communities are under-represented for partisan websites in comparison with the rest of the Internet population; thus the audience for hyper partisan websites appears to be white caucasians.

We also note (no figure included) that females are under-represented amongst visitors to our list of partisan websites, and those that follow such websites are more left-leaning. Male readers are over-represented for both conservative and liberal websites, but there is a higher inclination towards conservative news.

In summary, the demographics study identifies an increased representation of conservative-leaning audiences in the most common demographic categories for education, income and ethnicity. Furthermore, many of the more engaged (i.e., demographic score > 100) age ranges also tend to be right leaning. The only highly engaged age range which is left leaning age group is the 65+ age range (15% of the US population). Finally, males also tend to be more conservative, as observed in several places (e.g., The Guardian [3] and Washington Post [8] These are reflected to some extent with the much larger numbers of conservative rather than right-leaning websites.

6 PARTISAN TRAFFIC FLOWS

Since the Random Surfer model was introduced in the PageRank [17] paper, it has been understood that sites can become important if other important sites refer to them. In this section, we use this concept to understand how the partisan websites gather traffic. Using Alexa data, we obtain and analyse the list of upstream referrers to the partisan websites in our dataset, finding strong evidence of partisan echo chambers of right- and left-leaning websites.

6.1 Drivers of traffic to partisan websites

According to Alexa, upstream sites are sites that the visitors of a website visit immediately before visiting it. These are obtained using a diverse range of browser extensions and plugins from Alexa's measurement panel. Depending on availability of statistics, Alexa

[9] http://www.npr.org/sections/money/2012/07/16/156688596/what-americans-earn

lists up to 5 top upstream referers to each website. We use Alexa's list to understand who drives traffic to the partisan sites dataset.

Category	Description	Examples
Search	Search engines	Google, Yahoo, Bing
Social	Social media sites	Facebook, Twitter
YouTube	Video-streaming	YouTube
Internal	Partisan sites	Infowars, Breitbart
External	Other sites	New York Times, Huffington Post

Table 5: **Upstream Sites Categorization**

Upstream data is available for 484 of the 668 partisan websites. Of these, 70% (339) are right leaning, and 30% (145) are left leaning. As a baseline, we compare the distribution of upstream referers to partisan websites with 478 highly popular mainstream news sites for which we collected upstream referer data.

We divide the upstream referers into five broad categories as described in Table 5. As with many other websites, search engines and social networks such as Facebook are the main referrers, apart from YouTube, which contributes a small but significant minority. The remaining two categories are 'internal' and 'external'. We define internal upstream referers as other sites within the same category. Thus, the internal referers for partisan websites are other partisan websites; internal referers for mainstream news websites are other mainstream news websites. The external category captures all sites which are not otherwise categorised (i.e., all non-internal, non-social and non search-engine and non YouTube websites).

Figure 4 shows that social networks and search engines are the top drivers of traffic to both partisan as well as mainstream news websites. However, partisan websites have a disproportionately large amount of internal referals. The Internal category consisting of 14.7% of partisan upstream sites that drive traffic among themselves together form the 3rd largest source of traffic to partisan sites as shown in figure 4. The share of traffic from the internal linkage between partisan websites is higher than many traditional drivers of internet traffic such as Twitter(9.3%) and YouTube(9%). We find this trend to be consistent for both left- and right-leaning websites (right-left split not shown in figure).

Moreover, the share of internal upstream referers is 1.6 times higher in partisan websites than mainstream news sites. In terms of individual numbers, nearly half(46%) of the unique upstream sites to partisan websites are internal, in contrast to the case of mainstream websites where only 23% of upstream sites are internal. These evidences point to an echo chamber of sorts amongst partisan websites which are driving traffic amongst each other, which we investigate further.

6.2 Echo chambers among partisan producers

We previously observed in Figure 4 that internal referers are a significant driver of traffic to partisan websites. We further break down such referring by political inclinations of the partisan websites to illuminate traffic flows among left and right websites. So, to test if we actually see any peculiar internal linking among left and right leaning websites we plotted the instances where a left leaning partisan site links to a left leaning site, right leaning links to right leaning and then sites that link across the aisles (i.e., a site on the right linking to a left-leaning website and vice versa).

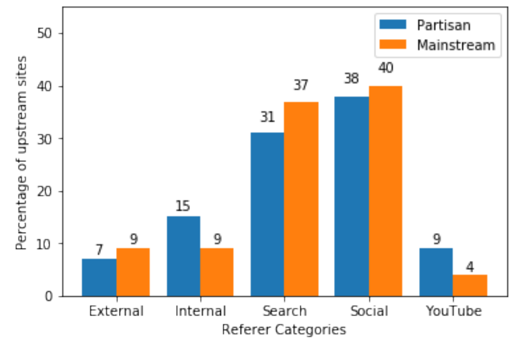


Figure 4: ***Categorical distribution of Partisan and Mainstream referrers***

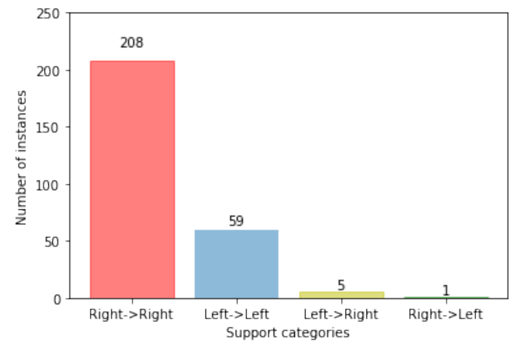


Figure 5: ***Linkages amongst and across partisan sides***

Figure 5 shows that internal referrals are clearly split along party lines—there is a very significant amount of internal referrals within the right as well as within the left but very little referrals from left to right or right to left. Indeed, there are 208 instances of a right partisan website being a top-5 referer to another right partisan website, and 59 instances of a left partisan website being a top-5 referer to another left partisan website, in comparison to the single digit number of instances where a left site is a top-5 referer for a right-leaning site or vice versa.

Internal referrals appear to be more prevalent on the right: As mentioned in §6.1, upstream traffic data is available for 339 (145) sites on the right (left). Of these, 139 (46) right (left) sites receive internal referrals according to Figure 5. In other words, 41% of sites on the right benefit from receiving internal referrals as opposed to 31.7% of sites on the left—internal linkages are 1.3 times more prevalent on the right than on the left.

This result highlights the existence of echo chambers among partisan websites in general, and conservative websites in particular. Such referral structures might serve to further reinforce and promote partisan micro-cultures.

6.3 Network of partisan links

To discover the most influential partisan domains we modelled the partisan sites as a network where an edge is drawn from one website A to another website B if A is a top-5 referer of B. Thus, we obtain a graph with a maximum in-degree of 5. We draw this graph in Figure 6, with right (left) leaning sites coloured as red (blue), and the size of the labels is proportional to the out-degree of the node, i.e., the size of a node A is proportional to the number of other sites on whose top-5 referer lists A appears. This figure

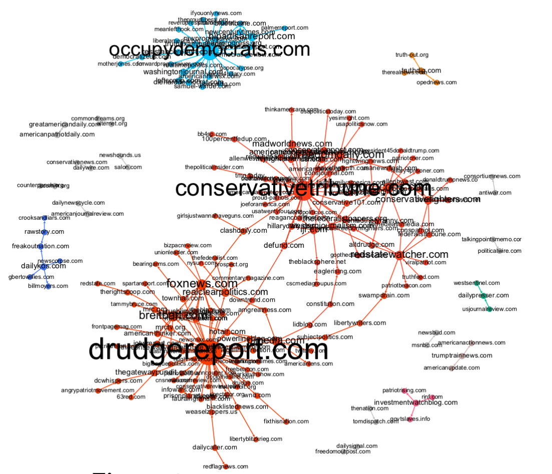


Figure 6: ***Partisan Sites Network Graph***

clearly visualises the separation between right and left leaning sites. This is confirmed by running the Louvain method for community detection, which achieves a modularity value of 0.69, indicative of an extremely strong community structure, with dense links within nodes in the same community and few links across communities.

We next move to identify the important players that send traffic to other sites, as the nodes with the largest number of outgoing edges, and the important beneficiaries as the nodes with the largest number of incoming internal edges. Table 6 lists the top 10 partisan sites with the largest outgoing and incoming links. Considering the ratio of right to left sites in our dataset (§3), we expect right-leaning sites to comprise ≈ 70% of the lists. While we see the expected number of left-leaning sites amongst the sites with the most incoming links, the preponderance of right-leaning sites in the list of sites with the most outgoing links suggests the right has a monopoly on forward traffic referals.

Top Outgoing	**Top Incoming**
Drudge Report (R)	Blue Tribune (L)
Conservative Tribune (R)	Powerline Blog (R)
Occupydemocrats (L)	My Right America (R)
Fox News (R)	Raw Progressive (L)
Young Cons (R)	Hillary Daily (R)
Freedom Daily (R)	Hot Air (R)
Breitbart (R)	Die Hard Democrat (L)
Red State Watcher (R)	Conservative101 (R)
Pj Media (R)	Red White and Right (R)
Conservative Fighters (R)	Rush Limbaugh (R)

Table 6: Highly linked websites: Top-10 websites by highest out-(in-) degree. Right (Left) websites are marked as R (L).

7 PARTISAN FACEBOOK COMMUNITIES

Many of the partisan websites in our dataset also have corresponding Facebook pages, with some of them having millions of followers. The social features of Facebook pages allows us to provide a more complete, yet different picture of the hyper partisan news ecosystem: Rather than having to infer traffic referrals from Alexa summaries, we can identify which pages 'like' which other page. Users who interact through comments and likes on Facebook pages can be identified, allowing us to directly examine the consumers of hyper partisan content. Short URLs such as bit.ly links are used on some Facebook pages. By tracking these links, we can identify engagement beyond comments and likes to measure click-through rates. Together, these studies paint a picture of the consumers of hyper partisan news as highly engaged but separated communities.

7.1 Partisan page endorsements

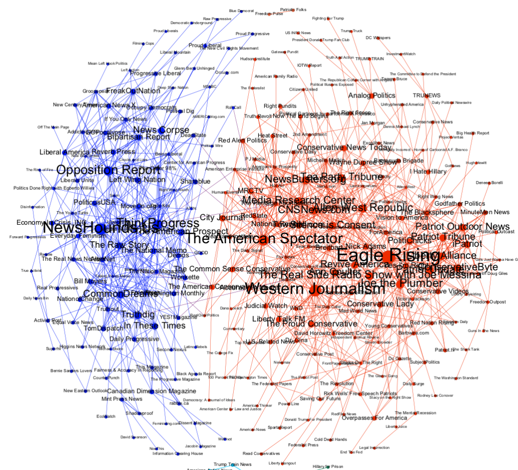


Figure 7: Graph induced from Facebook pages and pages which like these pages. Node sizes are proportional to the in-degree of that page.

Our investigation of upstream sites and their internal traffic forwarding patterns showed a highly polarized structure of the partisan news web. However, this traffic referral had to be inferred through Alexa data, and the study was limited by the fact that only the top-5 referers were available through Alexa. By contrast, Facebook pages can explicitly 'like' other pages, which allows us to infer the links between different partisan entities directly.

Since many of the partisan websites we examine also have official Facebook pages, we crawl these using the Facebook Graph API, and also crawl the pages liked by these partisan pages. We then build a graph of relationships between these pages, assigning directed links from Page A to Page B if Page A likes Page B. Nodes are coloured red or blue according to whether they are right or left leaning. The resulting graph is shown in the Fig 7. As with Fig. 5, this graph shows a clear community structure with a high modularity value of 0.46. The node labels are drawn proportional to the in-degree, i.e., the label of a page X is proportional to the number of pages which like page X. It is worth noting that the major nodes here (nodes with large labels) are mostly different from the ones in Fig. 5, suggesting that the Facebook pages are a different partisan ecosystem than the websites, although they have a similar divide across the left and right ideologies.

7.2 Echo Chambers among Facebook Users

To understand user engagement with these pages, we focus on the top 10 pages on the left and right with the most number of followers. In each page, we then identify 100 posts with the highest number of reactions. In this manner, we create a corpus of $10 * 100 = 1000$ posts on the left, and a similar number on the right. For each post, we identify up to 1000 users who have liked the post. Thus, we identify up to $1000 * 100 = 100,000$ users who have liked posts on each Facebook page.

With this data, we ask what is the overlap between the left and the right. We randomly pick $k = 2$ pages on the left and right respectively. We then take the set of users who have interacted with the top 100 posts on the $k = 2$ pages on the left and the set of users who have interacted with the top 100 posts on each of the $k = 2$ pages on the right, and compute their intersection, finding the number of users who are common. We may similarly compute the left-left and right-right overlap, finding users who have interacted with $k = 2$ pages on the left (right). This process can be repeated for $k = 2, 3, 4 \ldots 10$. Note that for any given k, the number of pages involved in a left-left or right-right interaction is k, whereas the number of pages involved in a left-right interaction is $2k$, thus offering *more* opportunity for users to interact.

Figure. 8 shows the result of computing the overlap between pages using this process. As k increases, the number of users who are common across all k sites drops, and the values become too small to report after $k = 4$. We also observe a left-left and right-right bias, wherein there is significantly more overlap of users who interact with pages all from the left or all from the right, than with pages from both the left and the right. While it is expected that the number of users who interact with k different pages drops as k increases, the significant difference between left-left and right-right interactions on the one hand and left-right interactions on the other hand again points to a hyper partisan *audience*, similar to the hyper partisan links found between *producers*.

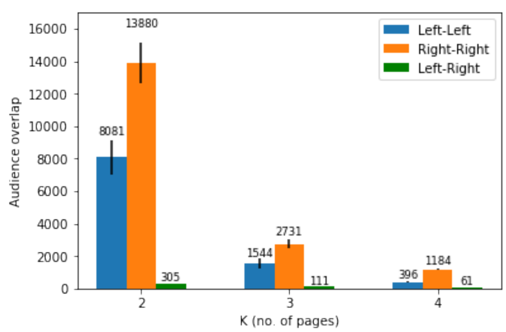


Figure 8: *Overlaps of users amongst K random pages from either sides*

7.3 Characterising Facebook Page Engagement

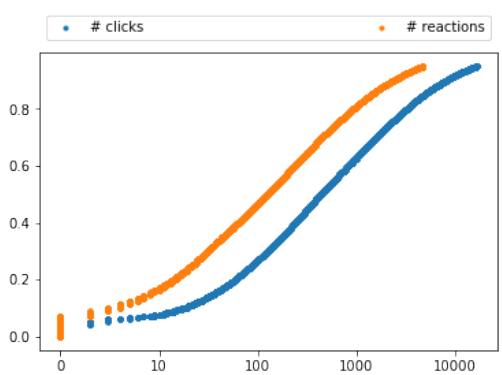


Figure 9: *CDF of reactions and Click through of bit.ly URL posts*

The 4 Million posts that we obtained by crawling across 507 partisan pages, fall into three distinct categories. They are embedded URLs (84.5%), photos(11%) and videos(3.7%). We also find that 6% of the embedded URLs are shortened URLs, shortened by services like Bit.ly or goo.gl. *Tea Party*, *ClashDaily.com with Doug Giles* and *American Thinker* are the top 3 pages with maximum number of short-urls (> 15,000 each). Finding Bit.ly shortened URLs allowed us to track the click through rates of these shortened URLs using the bit.ly API. In order to understand whether the Facebook users actually engage and if they do click through short URLs such as *Bit.ly*, we crawled the clicks data for about 25 thousand short URLs belonging to 216 pages. The Fig. 9 shows the CDF of reactions on a facebook post with an embedded URL, and the click through for these URLs. Interestingly we note that click through is much higher than reactions on facebook. Also we found that the two variables are positively correlated (R = 0.26, p « 10^{-6}). Moreover after calculating $mean(\frac{\text{Clicks for a URL}}{\text{Subscribers of the host page}})$ across all partisan pages yields a click through ratio of 4%. This rate is according to some sources[10] much higher than many marketing returns for ad campaigns. These insights show that facebook is not only promoting the echo chambers of partisanship, but are also effective in accruing clicks for these sources.

8 CONCLUSION

Hyper-partisan news media has been proven to be a major component in fake news proliferation[11]. In this study, we used a journalist-curated set of 668 hyper-partisan websites related to US News, and their associated Facebook pages, to characterise in depth the websites that produce hyper partisan content, and the consumers of this content. We also answered two questions which have not been examined until now: how does traffic get directed towards partisan websites, and how do users engage with these websites.

It was found that many of the websites, especially right-wing ones, were set up in 2016 during the run up to the US elections, and their Alexa traffic ranks declined after the Trump elections. This quick birth and death process casts some doubt on the role of these sites as purveyors of 'news' – rather they appear to be makeshift platforms for distributing content when it was important to do so, or when there was a willing audience with an appetite for partisan politics. Our data-driven approach can only uncover the correlation of these websites' lifetimes with the US election. Pinning down the cause for this would be an interesting avenue for future research.

We analysed the demographics of the audience of partisan websites, and observed an increased propensity for conservatism amongst the most populous demographic groups (e.g., Caucasians, middle income group, and average educational attainment), which points to increased interest in and potentially support for right wing causes. Analysis of traffic patterns over these partisan sites helped us uncover tightly interlinked communities of a partisan nature, which also showed evidence of echo chambers among producers of news, similar to ones observed amongst consumers previously.

Overall, this study revealed that the way traffic, information and users are forwarded amongst hyperpartisan sites might be aiding, or worse, widening the highly polar nature of news media. Our study is not without limitations: As mentioned, a data-driven study such as

[10] http://mashable.com/2009/07/07/twitter-clickthrough-rate/

[11] https://www.nytimes.com/2017/01/11/upshot/the-real-story-about-fake-news-is-partisanship.html

this cannot uncover causal relationships. The dataset itself has been curated carefully, and to our knowledge, is the largest list of partisan websites. However, our conclusions may not generalise beyond those sites active in the 2016 US Elections, or to partisan sites in other geographies. Finally, although Alexa provides several hints and reassurances about the quality of data (see text and footnotes earlier), it does not fully reveal the methodology it uses, which makes it harder to understand the true potential or limitations of some of our results which rely on this data.

REFERENCES

[1] Matthew Gentzkow Allcott, Hunt. 2017. Social media and fake news in the 2016 election. *National Bureau of Economic Research* No. w23089 (2017).

[2] Sinan Aral and Dylan Walker. 2012. Identifying influential and susceptible members of social networks. *Science* 337, 6092 (2012), 337–341.

[3] Mona Chalabi. 2016. Trump's angry white men âĂŞ and why there are more of them than you think. (2016). https://www.theguardian.com/us-news/2016/jan/08/angry-white-men-love-donald-trump

[4] Lam Thuy Vo Jeremy SingerVine Craig Silverman, Jane Lytvynenko. 2017. Inside The Partisan Fight For Your News Feed. (2017). https://www.buzzfeed.com/craigsilverman/inside-the-partisan-fight-for-your-news-feed

[5] Samantha Finn, Panagiotis Takis Metaxas, and Eni Mustafaraj. 2014. Investigating rumor propagation with twittertrails. *arXiv preprint arXiv:1411.3550* (2014).

[6] Seth Flaxman, Sharad Goel, and Justin M Rao. 2016. Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly* 80, S1 (2016), 298–320.

[7] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2017. Reducing controversy by connecting opposing views. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. ACM, 81–90.

[8] Steven M. Gillon. 2017. Why are so many white men so angry? (2017). https://www.washingtonpost.com/news/made-by-history/wp/2017/08/29/why-are-so-many-white-men-so-angry/

[9] Benjamin D Horne and Sibel Adali. 2017. This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News. *arXiv preprint arXiv:1703.09398* (2017).

[10] ELISA SHEARER JEFFREY GOTTFRIED. 2016. News Use Across Social Media Platforms 2016. (2016). http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/

[11] ELISA SHEARER JEFFREY GOTTFRIED. 2017. AmericansâĂŹ Attitudes About the News Media Deeply Divided Along Partisan Lines. (2017). http://www.journalism.org/2017/05/10/americans-attitudes-about-the-news-media-deeply-divided-along-partisan-lines/

[12] Sejeong Kwon, Meeyoung Cha, Kyomin Jung, Wei Chen, and Yajun Wang. 2013. Prominent features of rumor propagation in online social media. In *Data Mining (ICDM), 2013 IEEE 13th International Conference on*. IEEE, 1103–1108.

[13] Eric Lawrence, John Sides, and Henry Farrell. 2010. Self-segregation or deliberation? Blog readership, participation, and polarization in American politics. *Perspectives on Politics* 8, 1 (2010), 141–157.

[14] Jan E Leighley and Arnold Vedlitz. 1999. Race, ethnicity, and political participation: Competing models and contrasting explanations. *The Journal of Politics* 61, 4 (1999), 1092–1114.

[15] Dung T Nguyen, Nam P Nguyen, and My T Thai. 2012. Sources of misinformation in online social networks: Who to suspect?. In *Military Communications Conference, 2012-MILCOM 2012*. IEEE, 1–6.

[16] Greg Orelind. 2017. Top 6 Myths About The Alexa Traffic Rank. (2017). https://blog.alexa.com/top-6-myths-about-the-alexa-traffic-rank/

[17] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. *The PageRank citation ranking: Bringing order to the web.* Technical Report. Stanford InfoLab.

[18] Emilee Rader and Rebecca Gray. 2015. Understanding user beliefs about algorithmic curation in the Facebook news feed. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 173–182.

[19] SAMANTHA SMITH SHIVA MANIAM. 2017. AmericansâĂŹ Attitudes About the News Media Deeply Divided Along Partisan Lines. (2017). http://www.pewresearch.org/fact-tank/2017/03/20/a-wider-partisan-and-ideological-gap-between-younger-older-generations/

[20] Kate Starbird. 2017. Examining the Alternative Media Ecosystem Through the Production of Alternative Narratives of Mass Shooting Events on Twitter.. In *ICWSM*. 230–239.

[21] Samanth Subrahmanian. 2017. INSIDE THE MACEDONIAN FAKE-NEWS COMPLEX. (2017). https://www.wired.com/2017/02/veles-macedonia-fake-news/

[22] Liwen Vaughan and Rongbin Yang. 2013. Web traffic and organization performance measures: Relationships and data sources examined. *Journal of informetrics* 7, 3 (2013), 699–711.

[23] Sidney Verba and Norman H Nie. 1987. *Participation in America: Political democracy and social equality.* University of Chicago Press.

[24] Savvas Zannettou, Tristan Caulfield, Emiliano De Cristofaro, Nicolas Kourtellis, Ilias Leontiadis, Michael Sirivianos, Gianluca Stringhini, and Jeremy Blackburn. 2017. The Web Centipede: Understanding How Web Communities Influence Each Other Through the Lens of Mainstream and Alternative News Sources. *arXiv preprint arXiv:1705.06947* (2017).