Talking To Ourselves?

Political Debate Online and the Echo Chamber Effect

Alex Krasodomski-Jones

DEMOS

Open Access. Some rights reserved.

PARTNERS CREDITS

CONTENTS

ACKNOWLEDGMENTS

INTRODUCTION

The mainstream is shrinking. Trust in mainstream media is falling.[1] Mainstream politicians are seeing their majorities eroded by new parties on the left and right. 2016 was a year of unanticipated political decisions: the election of Donald Trump, the decision to leave the EU, even the re-election of Jeremy Corbyn as Labour Party leader. Repeatedly, long odds gave way to disbelief as the mainstream was rejected in favour of something radical and disruptive. Even the very idea of 'facts' has been shaken: by December 2016 the words 'post-truth' were on the lips of commentators around the world, eventually becoming the Oxford Dictionaries word of the year.

This has been mirrored and reflected in the landscape of political discussion. There has been, over the last decade, a dramatic change in the way political ideas, news and debates occur. When fingers have been pointed in blame, they have almost invariably been pointed at the internet.

The web is almost certainly the primary source of information for people living in the UK in 2017. Yet the idea that the breadth of information we are shown online is being technologically narrowed – filtered by algorithms and tailored by our increasing power to shape the news we see – has become a topic of keen debate in 2016. In the wake of the year's major political and cultural events, the way we use the internet to inform our news and views has been questioned. The charge levelled at the great online content providers is this: that their platforms are built to over-provide users with information that they agree with, or even to supress the content they do not. With so much of our politics now playing out online, critics have claimed that this kind of confirmation bias is causing the balkanization of political discussion, a strengthening of existing biases and political prejudices, and a narrowing of political, cultural and social awareness. This is the 'Echo Chamber'.

Much ink has been spilled on the way communities form and interact online, and on whether there is evidence for the 'echo chamber' effect. Many have pointed to fragmentation among online communities. Most recently, Jonathan Bright has shown that online party groupings that are politically divided tend to interact less with each other than internally, a phenomenon that increases towards the ideological fringes, and that centrist parties are more likely to interact with one another than centrist and fringe parties of the same ideological bent.[2] John Jost et al. found

---

[1] http://www.gallup.com/file/poll/195575/Confidence_in_Mass_Media_160914%20.pdf
[2] https://arxiv.org/ftp/arxiv/papers/1609/1609.05003.pdf

that US users were quite happy to debate across ideological lines when it came to the Superbowl, but that political discussions were balkanized.[3]

Other studies, however, have questioned this, illustrating that the echo chamber effect is counteracted by other digital trends. For instance, the *exposure* to information that the vast social media networks provide their users may also widen access to alternative viewpoints. Data from the 2015 Reuters Institute Digital News Report suggests that social media users are exposed to more diverse news sources than people in the offline world.[4] In a paper entitled *How Social Media Reduces Mass Political Polarization* (2015), Pablo Barbera has shown that being embedded in a wide and varied online network brings users into contact with diverse ideological views.[5] Research by Andy Guess supports this, with Guess arguing that the 'social' aspect of social media means that if a piece of content is popular, it is likely to be read regardless of held political prejudices.[6]

Ever since Campbell, Converse and Miller published *The American Voter* in 1960 we've known that people tend to choose partisan media outlets that they might agree with as their sources of news and views, both offline and online. A more recent question has been whether the use of algorithms to filter content by major platforms further exacerbates the polarity of this effect on the information people receive. Research by Bakshy, Messing and Adamic, data scientists at Facebook, shows that both human choice and algorithmic interference play a role in determining the structure of a user's online network, though their claim that personal preference is the main reason people click on links has been questioned, most notably by Christian Sandvig and Zeynep Tufekci.[7] [8] [9]

In this paper, we seek to add to this debate by measuring the existence of an echo chamber effect where it exists among established political groups in the UK, thereby testing commonly held assumptions around the way politics takes place online.

To do this, we collected Twitter data from 2,000 users who openly publish their support for one of four political parties in the United Kingdom. We subjected these tweets to a series of analyses, investigating whether there was evidence for an echo chamber effect in the way these users

---

[3] http://pss.sagepub.com/content/early/2015/08/21/0956797615594620.abstract

[4] http://digitalnewsreport.org/

[5] http://pablobarbera.com/static/barbera_polarization_APSA.pdf

[6] http://insights.berggruen.org/issues/issue-6/institute_posts/147

[7] http://science.sciencemag.org/content/early/2015/05/06/science.aaa1160.abstract

[8] http://socialmediacollective.org/2015/05/07/the-facebook-its-not-our-fault-study/

[9] https://medium.com/message/how-facebook-s-algorithm-suppresses-content-diversity-modestly-how-the-newsfeed-rules-the-clicks-b5f8a4bb7bab#.o9etb63co

communicated online in relation to their political party affiliations. Each piece of analysis turned on a different piece of metadata – a datapoint attached to a tweet such as the links contained within it or the user it was sent in reply to.

We found

1. Political groups in the United Kingdom are reflected in online communities of varying levels of cohesion, and at times the term 'echo chamber' is useful in describing how they engage with other users on Twitter.
2. People with the same party political affiliations tend to share news on Twitter from sites that are ideologically consistent with their party political affiliation.
3. The degree to which people share news from sites that are consistent with their party political affiliation differs by party.
4. People with more polarized political affiliations tend to be more inward-facing than people with more moderate political affiliations. In short, the echo chamber effect is more pronounced the further a group is from the centre.
5. Groups are more likely to interact with other groups who are ideologically aligned with them: our two left parties shared more similar content and interacted with each other more often than they did with the two parties on the right.
6. Breaking news, and non-party political views, have broad cross-party engagement, while news and views with strong political perspectives are disproportionately shared with those who share those perspectives.
7. Discussions of issues, when measured by words used or by hashtags used, show that certain topics are much more prevalently discussed by certain political groups than by others, and these topics are consistent with those parties' key political interests.

The paper suggests that there is a strong connection between a user's ideology and the users and news sources they interact with, and that offline beliefs play a key role in the way users behave online, a hypothesis that is often assumed but rarely measured. It also adds evidence that users with published support for political parties in the UK are more likely to share ideologically-aligned media, are more likely to keep within ideologically-aligned communities, and that this tendency increases the further the set of beliefs lies from the mainstream. It underlines the importance of mainstream news as a place where social media users with differing political viewpoints are most likely to encounter one another.

METHODOLOGY

This paper analyses four groups of Twitter users broken down by political affiliation (Conservative, Labour, UKIP and SNP) and examines the way each of these groups share media content on the platform and communicate with each other. As with much existing literature, Twitter is used here as the source of data, as it is the only major social media network which makes a good cross-section of its public data available to researchers, and is home to much political discussion in the United Kingdom.

It builds on previous work by analysing interactions between users, topics of discussion and media link sharing across a sample of different UK political groups.

The purpose is to determine the extent to which 'echo-chambers' exist within these groups, based on the links and stories they have publicly shared, or the users they have publicly messaged.

The data used in this study paints a picture of how these networks look on Twitter. What it cannot account for, however, is how these networks came to look the way they did. Whether the networks were formed by user prejudice, informing the people they follow and interact with, or by the result of algorithmic influence, is not possible to judge at this stage and would require data beyond the reach of this study. The likely answer is a bit of both. Users expressing a political affiliation on Twitter are a rarity – just a few percentage points of Twitter's total users. As Tufekci points out, publicly self-identifying as belonging to a party is likely the preserve of a certain minority type of Twitter user.[10]

This analysis is limited to a small cross-section of Twitter users, and may not be representative of all supporters of those parties on Twitter, or offline supporters of those parties. The inclusion of a control group as a point of comparison is therefore useful, but the findings are thus limited to Twitter's minority 'political classes' only.

Data collection

Between 9th May – 18th August 2016, we used Twitter's public API to collect all Tweets sent to a UK Member of Parliament. This dataset contained 644,000 unique accounts. Each user has a description which gives some biographical detail into the Twitter account. Although this field can be left blank, or can be deliberately misleading, this is rare, and

---

[10] https://medium.com/message/how-facebook-s-algorithm-suppresses-content-diversity-modestly-how-the-newsfeed-rules-the-clicks-b5f8a4bb7bab#.w19xrkhau

we judged the description field to contain useful data about a Twitter user.

A detailed methodological note about how data was collected and how users were selected is contained in the appendix below.

Building a sample

In order to identify political party supporters, we trained algorithms to spot users based on whether they mentioned a party in their biographical field. These excluded users mentioning a party in a negative way, for example 'ex-Labour' or 'Hates UKIP'. These algorithms were, on average, 90% accurate in making this distinction. A full breakdown is shown in the technical annex.

Using this approach, we randomly selected 500 users who supported each of the parties and a further 500 as a control group who had not mentioned any explicit support for a party. Due to limits on data storage and analysis capabilities, the study was deliberately limited to four parties deemed spanning the left and the right, and therefore excluded other UK parties such as the Liberal Democrats and Green Party.

We then collected 1.34 million tweets sent by those 2,500 users between 6th October – 16th November 2016.

Activity by party group

Not all accounts selected tweeted during the six week period. Of the 2,500 possible accounts, 2,295 (92%) sent at least one Tweet.

Table 1:

	Users (All)	Users (Labour)	Users (Conservative)	Users (SNP)	Users (UKIP)	Users (Control)
Users active over the period	2295	471	439	474	446	436
Users active over the period (% of list)	92%	94%	88%	95%	89%	87%

Ten accounts sent tweets more than 10,000 times during the period. 225 users Tweeted fewer than ten times. For this reason, the analysis is focused on the number of unique users sharing information or communicating with one another, rather than the number of times they

are tweeting, as focusing on the tweets risks skewing the analysis in favour of those user groups most active on Twitter.[11]

Removing automated accounts

Occasionally a Twitter account is programmed to send tweets automatically, often sending the same message multiple times or retweeting users automatically. These were deemed irrelevant to this study. Identifying and removing automated accounts required an analyst to judge whether an account was being operated by a computer programme, following a qualitative analysis of an account's output. For instance, if an account sent thousands of identical tweets, it was judged to be an automated account. In total, four accounts were removed from the dataset: two from the control group and one each from the SNP and UKIP groups.

The final dataset contained 1.25 million tweets from 2,263 users. The data was stored on a secure server in JSON format.

Analysis

Using our final data set (1.25 million tweets from 2,263 users) we analysed six types of behaviour seen publicly on Twitter.

- Sharing links to external websites
- Tweeting using a hashtag
- Sending a tweet mentioning another user
- Replying to another user
- Retweeting another user
- Tweeting, and the text of a tweet

Sharing links to external websites

This analysis turns on the links people have shared on Twitter linking to an external site, such a news site. Link sharing was analysed to identify the similarities and differences among our user groups in the news and other websites they were sharing.

330,000 links were shared by users during the six week period.

Many of the links shared were reduced in length using link shortening 'middle-man' services like bit.ly. To understand which websites were the eventual targets, links to external sites were then passed through Steno, a piece of software developed by the team at Kings College to turn these

---

[11] Studies show that insurgent parties are more active users of social media than their establishment counterparts in the UK. In this study, we found that on average, a user from our UKIP user group tweeted 713 times during the six week period, while the number for Labour and Tory users was 384 and 399 respectively. SNP users fell between the two, tweeting on average 644 times over the six week period.

shortened links into the original URLs they link to. More information on Steno is contained in the technical annex.

Shortened Link	Original URL
http://bit.ly/2heNL3U	http://quarterly.demos.co.uk/article/issue-10/businesses-behaving-badly/

Once the links had been collected and tabulated, the domain was isolated to identify the source website. Examples of this are shown below.

URL	Source Website
http://www.bbc.co.uk/news/uk-politics-38315259	bbc.co.uk
http://www.businessinsider.com/leaked-uber-email-minimum-wage-ruling-2016-10	businessinsider.com
https://www.change.org/o/voices_for_pets_2	change.org

Using this method, we identified 12,800 different websites that had been linked to during six week period. Websites containing multiple possible domain names (youtu.be, m.youtu.be and youtube.com, for instance) were treated as separate entities.

This allowed researchers to see how many users from each group had shared a link to a particular website, and compare proportional popularity in some sites among the political user groups.

28 media websites were linked to by at least 100 users in the sample. For these sites, a label was assigned based on where the outlet fell on a five point ideological scale. The possible categories were 'left, centre-left, centre, centre-right and right'. This allowed researchers to identify whether certain user groups more frequently shared links to websites from one end of the scale or the other by grouping links by category. For instance, if a user only shared articles from the BBC and Bloomberg, these would be aggregated to a 'centre' category. The categories were agreed by a group of researchers, and are shown in the analysis below. We accept there may be disagreement about an outlet's ideological position on the scale.

Tweeting using a hashtag

The hashtags contained in a tweet were also collected and analysed. Use of one or more hashtag on Twitter tends to signify an intent to comment on an issue or join a debate and is a good indicator of the subject of the Tweet. During the six weeks, users used 58,900 different hashtags.

Researchers then carried out two types of analysis.

First, the 200 hashtags used by the most number of users were extracted and qualitatively coded by topic, such as Healthcare, the EU Referendum or American Politics. Ten topics were identified. hashtags judged irrelevant (e.g. #Halloween) were categorised as 'Other'. This allowed researchers to identify whether certain user groups tweeted more frequently on one topic over another, and estimate whether certain topics were dominated by one of our user groups and ignored by another.

Secondly, the hashtags used to engage with five UK political television shows were analysed in the same way to estimate whether certain shows were more popular among one user group or another.

Mentioning, replying to or retweeting another user

Mentions, replies and retweets are pieces of Twitter data that indicate an interaction between two users on Twitter. A mention, for instance, is a tweet containing the screenname of another user on Twitter. A reply indicates a tweet by a user has been replied to directly.

When a tweet contained multiple mentions, these were treated separately.

Over the six week period, users mentioned 190,000 users, replied to 59,000 users and retweeted 122,000 users.

For each data point, researchers compared how many users within each user group had interacted with a user from the same user group or outside their user group. These metrics were used in two ways. First, they were cross-tabulated to see whether users tended to interact more frequently with one group or another, and which type of behaviour was most likely to be internally-facing. Second, they were used as the basis for a network analysis to visually represent the four communities and the interactions between them.

Twitter text

Finally, we analysed the words used by each of our user groups over the six week period. This aimed to identify key topics or themes that were disproportionately used by each user group, estimate how far each group were overly represented in using those words, and identify whether those words were ideologically linked to the parties each user group represented. This analysis did not turn on the use of NLP classifiers.

Displaying results

Most of the analysis below is shown in tables produced in Excel or through the analytics package QlikView.[12] Network analyses were performed and exported through the open-source network visualization package Gephi.[13]

Ethics

Conducting research using Twitter data presents ethical challenges in respect of how researchers should collect, store, analyse and present publicly posted tweets. Because it is a new field of research, there are no widely accepted protocols and approaches for how to do this ethically. Some useful guidance has been issued by the New Social Media New Social Science organisation, which recognises that there remain a number of outstanding ethical questions for research of this kind.

However, the Economic and Social Research Council principles of ethical research is an excellent guide for conducting research of all kinds – and can be usefully applied to online research as well as offline.

After reviewing these principles, we considered that the most important and relevant principles for this research paper were whether informed consent is necessary to collect, store, analyse and present their public tweets; whether there are any possible harms to participants in including and possibly re-publishing their tweets, as part of a research project; and whether directly publishing personal information about an individual that might make them identifiable was important for the research purpose (including where material might identify an individual via a search engine).

The question of whether informed consent was necessary is the most complex. Informed consent is widely understood to be required in any occasion of 'personal data' use when research subjects have an expectation of privacy. Determining the reasonable expectation of privacy someone might have is important in both offline and online research contexts.

Within this frame, an important determinant of an individual's expectation of privacy on social media is by reference to whether the individual has made any explicit effort or decision in order to ensure that third parties cannot access this information. In the UK, there are a number of polls and surveys that have gauged public attitudes on this subject, including a small number of representative, national level surveys. Taken together, they similarly find that citizens are increasingly

---

[12] http://www.qlik.com/en-gb
[13] https://gephi.org/

worried about losing control over what happens to their personal information, and the potential for misuse, by both governments and commercial companies. These surveys also show, however, that it is less clear what people actually understand online privacy to entail. They found that there is no clear agreement on what constitutes personal or public data on the internet.

Applying these two tests to Twitter in respect of our work we believe that there is, in general, a low level of expectation of privacy for those who tweet publicly available messages. (This is not true of all social networks). Twitter's Terms of Service and Privacy Policy state: "What you say on Twitter may be viewed all around the world instantly. We encourage and permit broad re-use of Content. The Twitter API exists to enable this". Societal expectation of privacy on Twitter, we believe, is also relatively low given recent court cases that have determined tweets are closely analogous to acts of publishing, and can thus also be prosecuted under laws governing public communications, including libel.

That said, it is possible that different users have quite different views about reasonable expectations of privacy in respect to Twitter. For example, a user posting from an official account of an organisation might have a different expectation from someone posting in a personal capacity with a small number of close followers.

In this study, we considered that although there is a generally low expectation of privacy for those who post publicly on Twitter, this could vary across users and is not always very easy to determine.

With regards to both republishing tweets and using usernames, we resolved to avoid this where possible. All measures are aggregated at a group level. Usernames are removed from the data with the exception of prominent, highly-publicised accounts who feature in the research due to their popularity on Twitter (e.g. garylineker, skynews, jihadwatch). We avoided republishing specific posts.

We are careful not to pass judgement on user groups of any ideological leaning, nor on any of the media sources shared by our user groups. The use of the term 'alternative' to describe media outlets like Breitbart or Truthfeed reflect their position outside of the mainstream, and is a term frequently used by these sites to describe their own position in opposition or contrast to the mainstream.

FINDINGS

Link sharing

Twitter users often use the platform to share links to other websites and this data can begin to paint a picture of the types of information flowing through the network. Across the six week period, 330,000 links to 12,800 websites were shared by our user groups and 28 outlets were interacted with by at least 100 of our users. These were categorised by orientation, and the number of users linking to their websites is shown in table 2.

Table 2: Number of users per user group linking to an external site

Site	Control	Labour	SNP	Tory	UKIP	Total
bbc.co.uk	80	246	268	198	221	1013
theguardian.com	80	286	242	142	158	908
independent.co.uk	59	178	222	89	148	696
telegraph.co.uk	35	107	138	168	229	677
dailymail.co.uk	63	58	47	115	247	530
mirror.co.uk	25	158	97	58	115	453
express.co.uk	22	24	51	64	254	415
news.sky.com	22	53	80	68	156	379
huffingtonpost.co.uk	9	117	103	45	72	346
thetimes.co.uk	14	66	84	76	62	302
itv.com	26	50	64	50	108	298
ft.com	17	68	113	48	41	287
thesun.co.uk	21	19	35	54	151	280
standard.co.uk	9	63	39	55	101	267
breitbart.com	24	5	4	29	190	252
order-order.com	6	12	11	83	139	251
bloomberg.com	39	42	73	42	40	236
buzzfeed.com	35	54	83	38	23	233
bbc.com	40	39	65	31	54	229
huffingtonpost.com	37	64	68	20	19	208
newstatesman.com	6	79	68	19	18	190
blogs.spectator.co.uk	8	21	29	62	66	186
thecanary.co	5	46	103	11	7	172
metro.co.uk	12	26	64	19	45	166
economist.com	11	34	37	21	10	113
infowars.com	13	1	1	16	78	109
labourlist.org	2	84	7	9	3	105
wingsoverscotland.com	0	1	100	1	0	102

Fewer members of the control group linked to these media sites than the other four political user groups. This is a pattern that is repeated throughout the research, and is not surprising: the control group is less political than the four groups representing Twitter's political classes. In total, these pages were shared 720 times by the control group, around a third as frequently as the political user groups. The top three sites shared most often by the control group were YouTube (190 users), Vine (145 users) and Instagram (117 users), three of the most widely shared websites on the internet, and further indication that our control group shows much less interest in UK politics than our political user groups.

Table 3: External websites linked to by 100+ users, percentage of sharing users by site

Site	Orientation	Labour	SNP	Tory	UKIP
bbc.co.uk	Centre	26%	29%	21%	24%
theguardian.com	Centre-Left	35%	29%	17%	19%
independent.co.uk	Centre	28%	35%	14%	23%
telegraph.co.uk	Centre-Right	17%	21%	26%	36%
dailymail.co.uk	Right	12%	10%	25%	53%
mirror.co.uk	Left	37%	23%	14%	27%
express.co.uk	Right	6%	13%	16%	65%
news.sky.com	Centre-Right	15%	22%	19%	44%
huffingtonpost.co.uk	Centre-Left	35%	31%	13%	21%
thetimes.co.uk	Centre-Right	23%	29%	26%	22%
itv.com	Centre	18%	24%	18%	40%
ft.com	Centre	25%	42%	18%	15%
thesun.co.uk	Right	7%	14%	21%	58%
standard.co.uk	Centre	24%	15%	21%	39%
breitbart.com	Right	2%	2%	13%	83%
order-order.com	Right	5%	4%	34%	57%
bloomberg.com	Centre	21%	37%	21%	20%
buzzfeed.com	Centre-Left	27%	42%	19%	12%
bbc.com	Centre	21%	34%	16%	29%
huffingtonpost.com	Centre-Left	37%	40%	12%	11%
newstatesman.com	Left	43%	37%	10%	10%
blogs.spectator.co.uk	Right	12%	16%	35%	37%
thecanary.co	Left	28%	62%	7%	4%
metro.co.uk	Centre	17%	42%	12%	29%
economist.com	Centre	33%	36%	21%	10%
infowars.com	Right	1%	1%	17%	81%
labourlist.org	Left	82%	7%	9%	3%
wingsoverscotland.com	Left	1%	98%	1%	0%

A number of the websites are almost exclusively the preserve of one user group. WingsOverScotland, a Scottish political blog, unsurprisingly saw 98% of its shares coming from the SNP user group. LabourList, a pro-Labour news site, had 82% of its shares from the Labour user group. The UKIP user group was over-represented on a number of right-wing websites, including Breitbart (83%), Infowars (81%) and the Express (65%). Among Conservative users, the stand-out links are the Guido Fawkes blog and the Spectator, though both actually received more shares from the UKIP user group by percentage. We cannot be certain whether a user shared a link in agreement or to hold it up for scrutiny, though the data does show some consistency in the ideological positions of the groups and the media they share. This suggests that link-sharing is usually done with interest in or support for the link shared.

Some websites had a much broader range of users sharing their content: the five most shared websites received a significant share of each user group. A look at each website's shares as a percentage by each user group shows how some user groups lean more heavily on some sources, but also the cross-user group appeal of some centrist-sites.

Table 4: External websites linked to by 100+ users, percentage of sharing users by user group

Site	Orientation	Labour	SNP	Tory	UKIP
bbc.co.uk	Centre	12%	12%	12%	8%
theguardian.com	Centre-Left	14%	11%	9%	6%
independent.co.uk	Centre	9%	10%	5%	5%
telegraph.co.uk	Centre-Right	5%	6%	10%	8%
dailymail.co.uk	Right	3%	2%	7%	9%
mirror.co.uk	Left	8%	4%	4%	4%
express.co.uk	Right	1%	2%	4%	9%
news.sky.com	Centre-Right	3%	3%	4%	6%
huffingtonpost.co.uk	Centre-Left	6%	4%	3%	3%
thetimes.co.uk	Centre-Right	3%	4%	5%	2%
itv.com	Centre	2%	3%	3%	4%
ft.com	Centre	3%	5%	3%	1%
thesun.co.uk	Right	1%	2%	3%	5%
standard.co.uk	Centre	3%	2%	3%	4%
breitbart.com	Right	0%	0%	2%	7%
order-order.com	Right	1%	0%	5%	5%
bloomberg.com	Centre	2%	3%	3%	1%
buzzfeed.com	Centre-Left	3%	4%	2%	1%
bbc.com	Centre	2%	3%	2%	2%
huffingtonpost.com	Centre-Left	3%	3%	1%	1%
newstatesman.com	Left	4%	3%	1%	1%

blogs.spectator.co.uk	Right	1%	1%	4%	2%
thecanary.co	Left	2%	4%	1%	0%
metro.co.uk	Centre	1%	3%	1%	2%
economist.com	Centre	2%	2%	1%	0%
infowars.com	Right	0%	0%	1%	3%
labourlist.org	Left	4%	0%	1%	0%
wingsoverscotland.com	Left	0%	4%	0%	0%

The top websites by number of users sharing them – in this case the BBC, the Guardian and the Independent – are often linked to by users across the four political user groups. In three of the four groups the BBC is the most widely shared external news site, suggesting that, regardless of political opinion, it remains a pillar of online news sharing in the United Kingdom. At best, these sites host or provoke debate from across the political spectrum. At worst, they may be held up or criticised by one group or another while being praised by another, but they nevertheless transcend the 'echo chamber'.

The UKIP user group diverges from the other three user groups in the news links they share, with sites that the other three user groups broadly ignore (the Express and Breitbart, for instance) forming a comparatively significant part of their user group's external news sources. This may be explained by a greater scepticism of UKIP voters to mainstream or establishment media outlets and a greater willingness to engage with alternative media outside of traditional outlets.

These findings show a clear selectiveness among user groups about the media they are interacting with, and is consistent with the theory that political orientation plays a role in determining the types of outlets Twitter users share.

Grouping media outlets and categorising them along a 'left-right' spectrum offers a different picture, shown in the table below.

Table 5: Percentage split of user group by the orientation of media they linked to.

	Left	Centre-Left	Centre	Centre-Right	Right
Labour	18%	26%	37%	11%	7%
SNP	16%	22%	41%	13%	8%
Tory	6%	15%	34%	19%	26%
UKIP	5%	10%	28%	16%	41%

For the majority of the user groups (control, Labour, SNP and Tory) the tendency is to share news from the centre, and there is a lot of diversity across all user groups. In one sense, this paints a rosier picture of

information flows: user groups are sharing links to media from across the left-right spectrum.

However, we should not ignore the groupings: the Labour and SNP user groups are twice as likely to share from left and centre-left sites as right and centre-right. This is reversed for the Conservative user group, and even more pronounced for the UKIP user group.

The relationships between the user groups and the external websites they are linking to can be visualized as a network diagram (generated in Gephi, an open-source network analysis platform). In the diagram below, users and external sites are represented by nodes, sized by the number of user sharing links from those sites and positioned using the force atlas algorithm.[14] Users are 'pulled' towards the nodes belonging to the sites to which they have linked.

Figure 1: Network structure of users sharing links and the external sites they link to. Users are coloured according to the user group that they belong to, with top external sites labelled and coloured grey
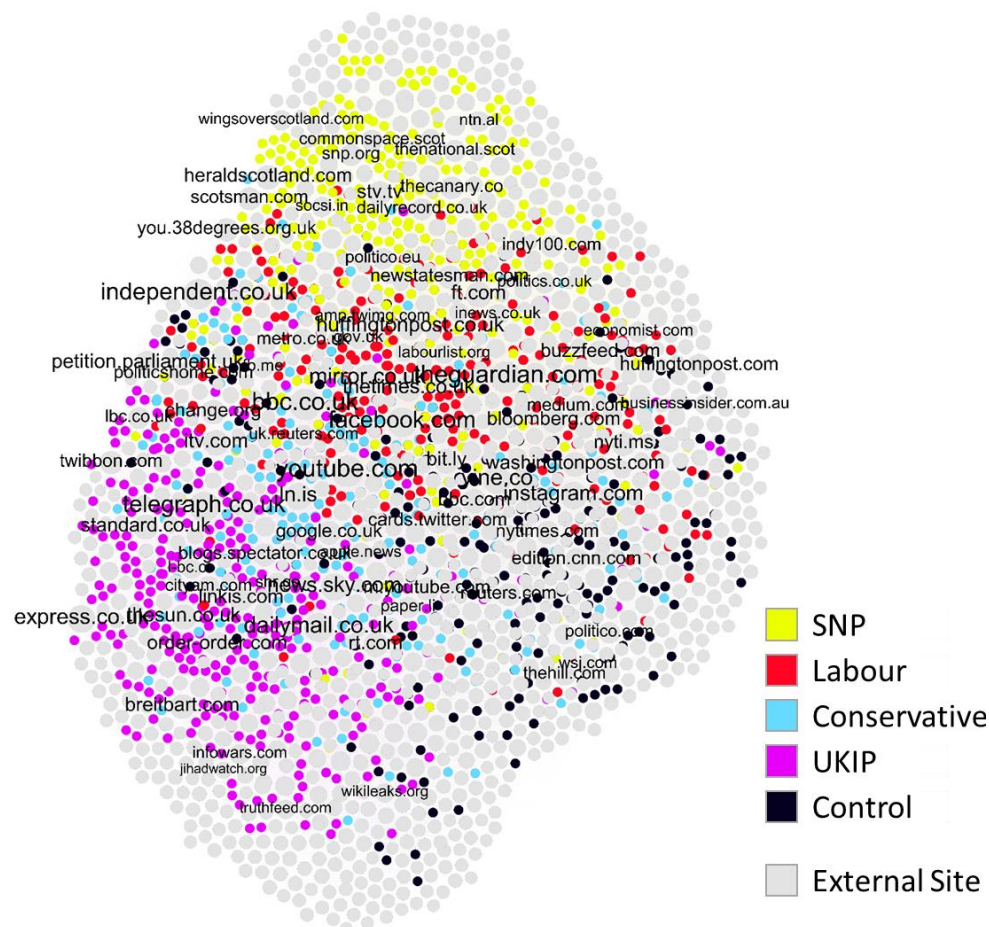
This kind of network visualization helps illustrate the data above. The impression given by the figure is that the SNP and UKIP user groups have formed stronger clusters than the other more politically mainstream user groups, and their positioning illustrates the fact there is little overlap in the sites they share. We find mainstream media and social media sites (bbc.co.uk, youtube.com, facebook.com) across the centre of the chart, explained by their use by the full spectrum of user groups, while extreme or special-interest sites like truthfeed.com, jihadwatch.org and wingsoverscotland.com are relegated to their respective user groups.

Table 6: Top media accounts (100+ unique users retweeting) by user group (%)

Screen Name	# Users	Control	Labour	SNP	Tory	UKIP
skynews	389	6%	15%	26%	19%	34%
independent	317	9%	27%	40%	11%	13%
guardian	283	8%	35%	36%	11%	10%
channel4news	246	5%	27%	57%	5%	6%
lbc	224	3%	15%	23%	16%	44%
guidofawkes	221	3%	5%	5%	34%	52%
telegraph	201	6%	16%	10%	36%	32%
bbcbreaking	197	13%	19%	23%	22%	23%
itvnews	195	5%	18%	26%	21%	30%
bbcnews	184	6%	24%	21%	21%	28%
skynewsbreak	182	4%	14%	26%	23%	34%
scotnational	182	1%	1%	96%	1%	1%
uk__news	175	3%	12%	13%	27%	45%
stvnews	172	0%	4%	87%	4%	5%
bbcnewsnight	148	3%	25%	33%	16%	22%
bbcscotlandnews	141	2%	6%	88%	1%	3%
mirrorpolitics	137	1%	53%	23%	11%	12%
pestononsunday	135	1%	10%	63%	7%	19%
daily_express	134	1%	1%	1%	17%	79%
politicshome	133	1%	32%	23%	28%	17%
huffpostuk	130	5%	29%	45%	12%	9%
rt_com	125	11%	6%	39%	6%	38%
onlinemagazin	123	6%	1%	2%	13%	79%
socialistvoice	120	2%	38%	41%	9%	10%
euroguido	113	2%	0%	2%	29%	67%
breitbartlondon	112	4%	0%	0%	12%	85%
telegraphnews	110	3%	12%	11%	32%	43%
dailymirror	105	12%	33%	20%	11%	23%
ft	105	10%	16%	46%	20%	9%
dailymailuk	100	10%	4%	3%	24%	59%

This analysis is supported by an analysis of a second piece of metadata, retweets. Twitter accounts belonging to media organisations who received retweets from at least 100 of our users are contained in the table above.

The pattern is similar to that of the external site link sharing. Centrist and breaking media has a broad appeal and is retweeted by users from across the four political groups, including the control group, while those with a distinct ideological position on the left or right or whose focus is geographically or topically narrower are more frequently retweeted by their respective user groups.

The control group retweets these accounts much less frequently than the user groups with express political positions.

We can again get an impression of how these patterns play out through a network analytic.

Figure 2:
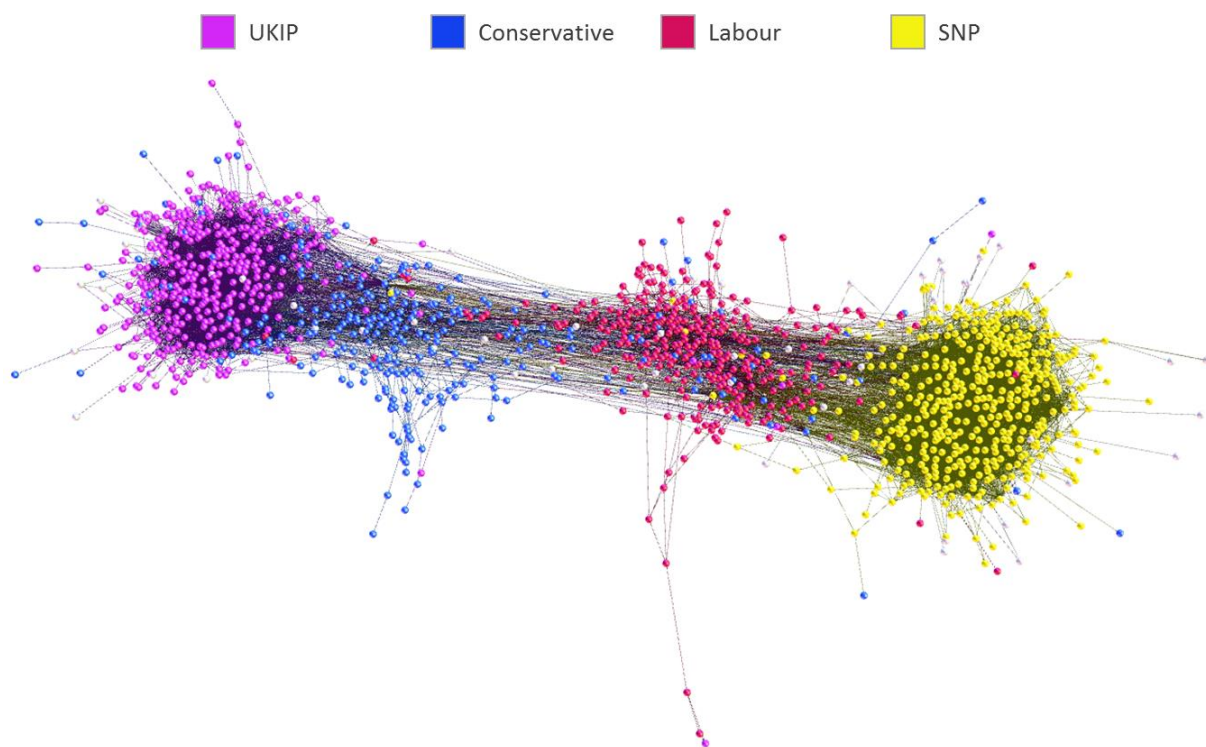
Here again, we see the SNP and UKIP user groups forming strongly defined clusters at opposite ends of the network, while the more centrist Labour and Conservative user groups are drawn to the centre and less clearly defined.

Retweet analysis

Retweeting another user – sharing their content with your own network – presents further data that can be used to measure how fragmented the user groups are. A retweet can be an indicator of endorsement: a user retweeting an account believes their network would be interested or supportive of the message. However, there are some exceptions – such as using the retweet to hold a message up to scrutiny or criticism. The data below suggests that a retweet is more likely to occur when two users are ideologically aligned.

This can be illustrated by identifying how both the control group and those with express political affiliations retweeted each other.

Table 7: Number of users retweeted by user group

		Users Retweeted				
		Control	Labour	SNP	Tory	UKIP
Users Retweeting	Control	77	15	25	24	63
	Labour	7	182	50	33	16
	SNP	9	53	339	23	19
	Tory	10	52	29	171	117
	UKIP	9	24	13	68	279

Again, the control group is unsurprisingly less concentrated: only 112 of their users retweeted someone from within the sample, compared to 319 for Conservative, 326 for Labour, 456 for SNP and 494 for UKIP user group.

Table 8: Percentage of users retweeted by user group

		Users Retweeted			
		Labour	SNP	Tory	UKIP
Users Retweeting	Labour	65%	18%	12%	6%
	SNP	12%	78%	5%	4%
	Tory	14%	8%	46%	32%
	UKIP	6%	3%	18%	73%

Unsurprisingly political user groups consistently retweeted users belonging to their own group more frequently than those belonging to others. Within the sample, over 75% of retweets by SNP users were of other SNP users. The two centrist groups are the most willing to retweet users

beyond their own group, with the Conservative user group the most fragmented: nearly a third of their retweets were of the UKIP user group. By contrast, the UKIP and SNP user groups focused on their own user groups most consistently.

Ideological division is also evidenced in the table. Labour and SNP users retweeted each other more often than the right-wing user groups, with the reverse true of the Tory and UKIP groups. On both sides, the division is weaker for the centrist parties. Just 9% of UKIP retweets were of the left user groups (compared to 22% for the Tory group). 9% of SNP retweets were of right-wing user groups, compared with 18% of Labour retweets.

A closer look at the Twitter users retweeted by our user groups shows how widely they differed. The figures below show the Twitter accounts retweeted by the greatest number of users across the user groups.

Table 9: Top ten accounts retweeted by users across political user groups

Screen Name	Labour	SNP	Tory	UKIP	Total
skynews	59	101	73	132	365
garylineker	158	155	30	10	353
britainelects	88	102	91	70	351
hillaryclinton	121	85	61	8	275
realdonaldtrump	31	27	58	148	264
nicolasturgeon	15	289	9	3	316
nigel_farage	4	7	36	257	304
independent	86	126	34	42	288
davidjo52951945	3	11	34	242	290
davidschneider	109	132	30	6	277

Table 10: Top 10 accounts retweeted by users across user groups (% of user group)

Screen Name	Labour	SNP	Tory	UKIP
skynews	9%	10%	16%	14%
garylineker	23%	15%	7%	1%
britainelects	13%	10%	20%	8%
hillaryclinton	18%	8%	13%	1%
realdonaldtrump	5%	3%	13%	16%
nicolasturgeon	2%	28%	2%	0%
nigel_farage	1%	1%	8%	28%
independent	13%	12%	7%	5%
davidjo52951945	0%	1%	7%	26%
davidschneider	16%	13%	7%	1%

The table above shows how the user groups retweets were distributed across the top accounts. A few observations can be made:

- Our two party leaders (at the time of writing) that appear in the top ten accounts are central to discussions by users who belong to their party. Nicola Sturgeon for SNP users (28%) and Nigel Farage by UKIP users (28%).
- Users who have publicly taken a stance on political issues are consistently retweeted by user groups who are ideologically aligned with them. Anti-establishment right-wingers (Donald Trump, David Jones and Nigel Farage, above) are popular with UKIP supporters. Accounts taking liberal or left-wing stances (Hillary Clinton, Gary Lineker and David Schneider) are popular with Labour and SNP users.

Table 11: Top ten accounts proportionally retweeted by unique users across user groups (minimum 100 retweets)

Labour			Conservative		
Account	# Users	% Labour	Account	# Users	% Cons
uklabour	112	90%	jamin2g	111	52%
jeremycorbyn	164	74%	mrharrycole	142	47%
mrbrendancox	104	71%	iainmartin1	115	46%
ed_miliband	108	70%	danieljhannan	217	46%
jessphillips	107	69%	dpjhodges	107	42%
wesstreeting	115	64%	montie	185	41%
angelarayner	172	63%	1jamiefoster	101	39%
jk_rowling	137	58%	skiplicker	105	37%
kevin_maguire	123	55%	afneil	254	36%
owenjones84	201	55%	johnrentoul	145	36%
SNP			UKIP		
Account	# Users	% SNP	Account	# Users	% UKIP
petermurrell	117	100%	Ukippoole	105	94%
alisonthewliss	107	100%	paulnuttallukip	131	92%
glasgowcathcart	101	100%	Busybuk	112	92%
zarkwan	100	100%	Ukip	118	92%
bjcruickshank	135	99%	2tweetaboutit	117	89%
markmcdsnp	126	99%	Prwhittle	108	88%
christinasnp	119	99%	Oflynnmep	112	88%
mhairihunter	116	99%	michael_heaver	118	87%
berthanpete	100	99%	euvoteleave23rd	116	87%
joannaccherry	164	99%	fight4uk	128	87%

The control user group showed little interest in retweeting UK political accounts, focusing instead on the extremely newsworthy US election. BritainElects, the third most widely retweeted account across our user groups, received just 1% of its retweets from our control group despite popularity across all four politically identified groups, again suggesting they are operating in a different online universe to the political user groups.

These observations are reinforced by an analysis in the table above of the accounts proportionally most retweeted by each of our five user groups.

Party political accounts dominate each user group. Five of the top ten Labour accounts are Labour MPs, six of the top SNP accounts are SNP MPs or MSPs. A similar pattern is present in the UKIP table. It is notable that there are no Conservative MPs in the Conservative table, which is dominated by journalist and blogger accounts (Harry Cole, Dan Hodges, Tim Montgomerie, Andrew Neil and John Rentoul).

Equally noteworthy are the average percentages across the top ten. As with the word analysis above, the top accounts are dramatically less contested in the two user groups furthest from the political centre. Among the SNP and UKIP user groups the accounts receiving the most unique retweets were those retweets almost entirely to their affiliated user group. This is particularly stark with the SNP, where all ten accounts are either 99 or 100 percent retweeted by the SNP user group. By contrast, the averages for the more centrally-aligned groups, Labour and Conservative, are 67 and 42 percent respectively.

This suggests that the accounts most retweeted by Labour and Conservative users are also being retweeted by a much broader audience, while within the UKIP and the SNP user groups the messaging from their key accounts is almost exclusively being shared within their own networks.

Tweet analysis

The text of a tweet is the most fundamental data point available to researchers analysing Twitter data. For this study, we looked to compare and contrast the words our user groups used to identify whether certain themes or topics were predominantly used by one political user group, and whether that reflected the political issues they are known for.

Each tweet was split into the words that made it up. We calculated the number of users who had used a word from our user groups, before selecting those words used by at least twenty users from across the sample. This represented 26,000 unique words.

The data was normalised for the number of active users in each user group before each word was assigned a value for its proportional use by one user group in comparison to the others. For instance, the word 'remainers', referring to those who voted to remain it the EU in the 2016 referendum, was six percent more likely to be used by the Conservative user group and 30 percent more likely to be used by the UKIP user group. The findings for the four political user groups are shown in the table below.

Table 12: Words most proportionally, frequently used by user group

Labour		SNP		Conservative		UKIP	
jo	+16%	#snp16	+52%	conservative	+15%	#ukip	+33%
corbyn	+16%	#wearescotland	+48%	robertcourts	+12%	nigel	+32%
jeremy	+16%	scotland's	+46%	#heathrow	+9%	migrants	+31%
shadow	+16%	glasgow	+44%	abbott	+8%	remainers	+30%
batley	+16%	#indyref2	+42%	#witneybyelection	+7%	establishment	+29%
nhs	+15%	scots	+39%	heathrow	+7%	calais	+28%
spen	+15%	wee	+39%	q3	+7%	farage	+27%
services	+15%	independence	+36%	witney	+7%	clegg	+26%
cuts	+14%	edinburgh	+35%	residents	+6%	biased	+25%
labour's	+14%	scotgov	+34%	icm	+6%	migrant	+25%
#labourdoorstep	+14%	nicola	+34%	remainers	+6%	remoaners	+24%
homelessness	+13%	scotnational	+33%	diane	+6%	islam	+24%

Proportional use of language by each user group suggests that users are taking to Twitter to discuss subjects related to their political user group. The percentages are particularly high for the two parties further from the centre: the SNP and the UKIP user groups. The words associated with these accounts are much more highly concentrated within these user groups than the words topping the Labour and Conservative lists (+16% and +15% respectively). The less a word is contested across the political user groups, the more relevant the label 'echo chamber' becomes. This supports the argument that non-centrist parties are more liable to communicate in echo chambers than those in the centre.

It is also clear that an explicit party affiliation plays a role in the things you tweet about on Twitter. Aside from mentions of parties or their leaders, discussions of the establishment, of migration, of the EU referendum and of Islam are recognised as central to UKIP discussions, and the pattern is mirrored on social media. Similarly, discussions of a second independence referendum for Scotland are largely limited to the SNP.

The control group proportionally overused non-political, everyday language and slang. By proportion, the top three words for the control group were 'wanna' (+16%), 'ur' (+15%) and 'ya', (+14%). The top ten are shown below.

Table 13: Words most proportionally, frequently used by the control user group

Word	Control
wanna	+16%
ur	+15%
ya	+14%
favourite	+12%
gonna	+12%
y'all	+12%
girl	+11%
song	+11%
cute	+10%
lmao	+10%

Hashtag analysis

Use of a hashtag on Twitter tends to signify an intent to comment on an issue and is a good indicator of the subject of the tweet. During the six weeks, our user groups used 58,900 different hashtags. Researchers took the 200 most popular hashtags and manually coded them by subject.

1. English Politics (e.g. #Labour, #Conservative)
2. Scottish Politics (e.g. #IndyRef2, #SNP)
3. US Politics (e.g. #Trump, #USElection2016)
4. EU & EU Referendum (e.g. #Brexit, #Article50)
5. Television and Media (e.g. #BBCQT, #Marr)
6. Healthcare (e.g. #NHS, #MentalHealth)
7. Foreign Affairs & Policy (e.g. #Syria, #Aleppo)
8. Immigration, Religion and Multiculturalism (e.g. #RefugeesWelcome, #Islam)
9. Environment (e.g. #Fracking, #ClimateChange)
10. Remembrance Sunday (e.g. #WeWillRememberThem, #RemembranceDay)

A final category, 'Other', contained hashtags that were not deemed relevant to the study. 50 of the top 200 hashtags (25%) were classified as 'Other', including #Halloween, #EnglandVScotland and #Breaking. The number of unique hashtags per category is shown in the table below.

Table 14: Number of unique hashtags per category

Category	# Hashtags
US Politics	47
Television and Media	29
UK Politics	23
EU & EU Referendum	11
Foreign Policy	10
Scottish Politics	10

Remembrance Sunday	8
Immigration & Multiculturalism	6
Healthcare	4
Environment	2
Other	50

The user groups differed sharply in their use of different categories of hashtag. Their respective percentages are shown in the table below.

Table 15: Hashtag category usage by user group (%)

	Control	Labour	SNP	Tory	UKIP
Environment	13%	25%	44%	9%	8%
EU & EU Referendum	4%	12%	24%	16%	44%
Foreign Policy	17%	13%	21%	15%	34%
Healthcare	6%	35%	34%	15%	10%
Immigration, Religion & Multiculturalism	7%	13%	17%	13%	50%
Remembrance Sunday	7%	19%	21%	24%	28%
Scottish Politics	1%	3%	82%	7%	6%
Television and Media	6%	22%	28%	18%	27%
English Politics	2%	29%	25%	20%	23%
US Politics	17%	11%	14%	16%	41%
Other	12%	17%	31%	14%	26%

Hashtags about Immigration, Religion and Multiculturalism, the EU and the EU Referendum, and US Politics, were most commonly used by members of the UKIP user group. Environment (although with only two hashtags more research would be necessary here) and Scottish Politics were dominated by the SNP. The global focus on the hugely newsworthy US election is reflected in its uptake by the control group.

Discussions on these partisan hashtags will therefore be heavily influenced by the user group that dominates it, and anyone following them will have the information they receive skewed by the position taken by one side.

Where the hashtags are more contested (English Politics and Television and Media in particular), the lines of argument and opinion will likely be more varied. Political television in particular provokes comment from across the four political user groups. The percentages of each group who tweeted to the hashtags of the five most popular television programmes is shown below.

Table 16: Political television hashtag usage by user group (%)

	Labour	SNP	Tory	UKIP
BBC Question Time	43%	38%	36%	33%
The Andrew Marr Show	26%	27%	24%	28%
BBC Daily Politics	10%	11%	14%	15%
Peston on Sunday	11%	12%	14%	13%
BBC Sunday Politics	10%	12%	11%	11%

Replies analysis

Twitter users can reply directly to a tweet. Of the metadata analysed, replies represent perhaps the strongest indicator of a user-to-user interaction, and the closest proxy available in the data for discussion or debate.

The number of times accounts across the five user groups replied to one another is shown in the tables below.

Table 17: Number of users replied to by user group

		Users Replied To				
		Control	Labour	SNP	Tory	UKIP
Users Replying	Control	184	10	11	8	23
	Labour	3	208	27	29	17
	SNP	5	34	305	31	21
	Tory	5	38	30	206	63
	UKIP	9	32	28	47	263

The control group replied to other users in the group surprisingly often (206 users), but was not as interconnected as our political groups, where on average 357 users replied to someone else in the sample.

Table 18: Number of users retweeted by user group (%)

		Users Replied To			
		Labour	SNP	Tory	UKIP
Users Replying	Labour	74%	10%	10%	6%
	SNP	9%	78%	8%	5%
	Tory	11%	9%	61%	19%
	UKIP	9%	8%	13%	71%

User groups overwhelmingly replied to accounts within their user group. The highest exception - Conservative accounts replying to UKIP accounts – still represented just 19% of the Conservative user group account replies. This lends further weight to the argument that discussions held on social media tend to be between ideologically aligned users, although the percentages show that it is by no means completely inward facing.

Again, an ideological divide is evident. Tory users replied to UKIP users more frequently than to any other user group, and vice versa, though the pattern is less pronounced on the left. Replies find the Labour and SNP user groups (74 and 78 percent respectively) more likely to be internally-looking than both the user groups on the right.

Mentions analysis

A mention represents a user deliberately notifying another user to a tweet by adding their screenname to the message. Although this can be part of a wider discussion, a mention can also be used to identify the subject of a tweet or even hold the user up for rebuke. It is therefore best used as another neutral signifier of engagement. Over the six week period, our user groups mentioned 190 thousand unique accounts. The number of users who mentioned another user within our sample is shown in the table below.

Table 19: Number of users mentioned by user group

		Users Being Mentioned				
		Control	Labour	SNP	Tory	UKIP
Users Mentioning	Control	234	24	41	35	78
	Labour	8	326	73	56	36
	SNP	10	78	391	58	39
	Tory	11	76	61	269	144
	UKIP	14	56	46	98	330

Table 20: Number of users mentioned by user group (%)

		Users Being Mentioned			
		Labour	SNP	Tory	UKIP
Users Mentioning	Labour	66%	15%	11%	7%
	SNP	14%	69%	10%	7%
	Tory	14%	11%	49%	26%
	UKIP	11%	9%	18%	62%

As with other pieces of metadata, the groups predominately mentioned users from within their user groups, though it is worth noting the percentages are lower than for other measured behaviours, perhaps owing to the more frequent use of a mention negatively towards opposing ideologies.

The left-right split we identified in reply and retweet data is replicated here: left-wing party groups mentioning each other's accounts more often than those of right-wing party groups and vice versa.

CONCLUSIONS

This paper argues that there is evidence that an echo chamber effect does exist on social media, and that its effect may be more pronounced the further a user sits from the mainstream.

Political groups in the United Kingdom are reflected in online communities that reinforce ideological positions. This differed by party. Those in the centre, represented by our Labour and Tory user groups, were less balkanized than the UKIP and SNP user groups. By contrast, political parties that are further removed from the centre on policy or simply geographic focuses are more likely to be interacting with those of a similar mindset.

It suggests supporters of UK political parties tend to talk to themselves online, to read and share news that is ideologically in tune with their party and discuss issues on which they hold strong ideological views. Alignment across ideological lines is evident. Across media sharing and the measured interactions, the Tory and UKIP user groups interacted more with each other than they did with the Labour and SNP user groups. With the exception of replies, the same was true for the left user groups: they favoured interactions with each other over those on the right.

Although the two smaller political groups analysed are minority parties at a UK level, they represent a significant body of opinion. UKIP received 12.6% of the vote in the 2015 General Election, while the SNP – 5% of the national vote - tore through their Labour opposition to a landslide north of the border.[15] The analysis shows that these groups are the most at risk of communicating in an echo chamber.

The evidence presented here supports the idea that the position of a political party within the political system changes the way they operate online. Mainstream and established parties seem more likely to embrace contested news sources, and to discuss issues and to communicate with users across a wider political spectrum. The smaller parties analysed, the SNP and UKIP, were on the whole more likely to communicate within the group than without.

Across the whole sample, however, every measure taken indicated that a political position was a factor in how our users consumed and shared news, who they spoke to and what they spoke about. All groups showed ideological cohesiveness in their behaviour and a preference to interact with users within their groups. When they did communicate outside the

---

[15] http://www.bbc.co.uk/news/election/2015/results

group, it was most likely to be with someone from the same end of the political spectrum.

Unfortunately, it is beyond the scope of this research to compare the online and the offline. We cannot say whether, in spite of a tendency to look inwards, users of social media are still exposed to more diverse political opinions or news than they would be offline. The sample used is necessarily small, and further analysis on the purpose or motivations behind the behaviours investigated would require more study.

What we can say is that partisan information consumption is an online phenomenon that closely resembles findings showing the same in the offline world. A rational approach to learning where a held belief is repeatedly challenged before being accepted is not a common model. Numerous studies have shown that we seek confirmation before we seek a challenge[16]; that we are less sceptical of evidence that supports our held suspicions[17]; and that we are heavily influenced by those around us.[18]

This poses a greater problem online than it does offline. The huge plurality of possible media sources online combined with the economy of attention and clicks drives outlets towards 'likeable', shareable, ideologically-driven content, often at the expense of nuance or balance.

These concerns are not new. Demos' 2011 report *Truth, Lies and the Internet* identified the problems facing internet users trying to learn about the world.[19] A lack of gate-keepers and editorial control, anonymity, pseudo-sites and, of course, echo chambers.

The long-term effects of the echo chamber effect aren't known, but there are early indications that they threaten the health of our democracies.

Polarisation of political opinion is on the rise. Some have placed this at the doorstep of echo chambers. "The more that the members of an individual's conversation networks speak with one political voice", write Pattie and Johnston, "… the more that person's opinions on major issues tend to move away from where they place the other parties…"[20] The number and resilience of conspiracy theories are on the rise.[21] Studies

---

[16] For a study suggesting that humans tend to seek confirmation of hypotheses they hold, see PC Watson, 'Reasoning' in B Foss (ed) New Horizons In Psychology (Harmondsworth: Penguin, 1966)

[17] S Sutherland, Irrationality (Pinter & Martin, 2007), 104-112

[18] M McPherson, L Smith-Lovin and JM Cook, 'Birds of a Feather: Homophily in Social Networks', Annual Review of Sociology 27 p 415-444 (Volume publication date August 2001).

[19] https://www.demos.co.uk/files/Truth_-_web.pdf?1317312220

[20] http://journals.sagepub.com/doi/full/10.1177/1369148115620989

[21] http://www.tandfonline.com/doi/full/10.1080/08913811.2016.1167404?scroll=top&needAccess=true

show users are still struggling to distinguish truth and lies on the web, and anecdotal evidence suggests the proliferation of fabricated news stories is also accelerating.[22]

If the trend for using digital channels as tools to form and share beliefs continues, the echo chamber effect may represent a significant challenge to democracy. Compromise, the ability to process a diverse range of opinion and, above all, an acceptance of some kind of shared reality and truth are central to a functioning democracy. All are threatened by the echo chamber effect.

---

[22] http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0150989

TECHNICAL ANNEX

Method52

Data drawn from social media are often too large to fully analyse manually, and also not amenable to the conventional research methods of social science. The research team used a technology platform called Method52, developed by CASM technologists based at the Text Analytics Group at the University of Sussex. [23] It is designed to allow non-technical researchers to analyse very large datasets like Twitter.

Sharing links to external websites

Method52 allows researchers to train algorithms to split apart ('to classify') tweets into categories, according to the meaning of the tweet, and on the basis of the text they contain. To do this, it uses a technology called natural language processing. Natural language processing is a branch of artificial intelligence research, and combines approaches developed in the fields of computer science, applied mathematics, and linguistics.

An analyst 'marks up' which category he or she considers a tweet to fall into, and this 'teaches' the algorithm to spot patterns in the language use associated with each category chosen. The algorithm looks for statistical correlations between the language used and the categories assigned to determine the extent to which words and bigrams are indicative of the pre-defined categories.

The accuracy of algorithms

To measure the accuracy of algorithms, we used a 'gold standard' approach. For each, around 100 user descriptions were randomly selected from the relevant dataset to form a gold standard test set for each classifier. These were manually coded into the categories defined above. These tweets were then removed from the main dataset and so were not used to train the classifier.

As the analyst trained the classifier, the software reported back on how accurate the classifier was at categorising the gold standard, as compared to the analyst's decisions. On the basis of this comparison, classifier performance statistics – 'recall', 'precision', and 'F-score' are created and appraised by a human analyst. Each measures the ability of the classifier to make the same decisions as a human in a different way:

---

[23] This group is led by Professor David Weir and Dr Jeremy Reffin. More information is available about their work at: www.taglaboratory.org

Overall accuracy:

This represents the percentage likelihood of any randomly selected description within the dataset being placed into the appropriate category by the algorithm. It is based on three other measures (below).

Recall:

The number of correct selections that the classifier makes as a proportion of the total correct selections it could have made. If there were 10 relevant descriptions in a dataset, and a relevancy classifier successfully picks 8 of them, it has a recall score of 80 per cent.

Precision:

This is the number of correct selections the classifiers make as a proportion of all the selections it has made. If a relevancy classifier selects 10 descriptions as relevant, and 8 of them actually are indeed relevant, it has a precision score of 80 per cent.

F-Score:

All classifiers are a trade-off between recall and precision. Classifiers with a high recall score tend to be less precise, and vice versa. The 'overall' score reconciles precision and recall to create one, overall measurement of performance for each decision branch of the classifier.

The F-score ranges between 0 and 1, with a higher number indicating better performance.

Caveats:

The research of large social media datasets is a reasonably new undertaking. It is important to set out a series of caveats related to the research methodology that the results must be understood in the light of:

- The algorithms used are very good, but not perfect: throughout the report, some of the data will be misclassified. The technology used to analyse tweets is inherently probabilistic, and none of the algorithms trained and used to produce the findings for this paper were 100% accurate. The accuracy of all algorithms used in the report are clearly set out in this report.
- Twitter, and especially political Twitter, is not a representative window into British society: Twitter is not evenly used by all parts of British society. It tends to be used by groups that are younger, more socio-economically privileged and more urban. Additionally, the poorest, most marginalised and most vulnerable groups of society are least represented on Twitter.

A full description of the algorithms used and their accuracies is shown below.

Party classifiers

For each user, the description field was analysed, and each user labelled according to mentions of one of four political parties in the UK contained within their biographical field: Conservative, Labour, Scottish National Party (SNP) and the United Kingdom Independence Party (UKIP). The totals for each were as follows:

UKIP/UK Independence Party: 1,055

Labour/UKLabour: 6,111

Conservative/Conservatives/Tory/Conservativeparty: 3,574

SNP/theSNP/Scottish National Party: 814

To differentiate between those users who were supporters and those who weren't, a classifier was built for each group to identify those who supported the party mentioned in their description and those who used the field to reject the party. These classifiers were built in Method52, CASM's technology for understanding large unstructured datasets developed by the Text Analytics Group at the University of Sussex. Further information about how Method52 classifies text, and the accuracies of the classifiers built, are contained in the Technical Annex.

Examples of users who rejected the parties named in their description fields are shown below.

"Ex Labour.. Ex Trade Union.. Ex Unionist.."

"Dislikes Ukippers and right wing"

"…Libertarian & Interdimensional time traveler sent through time to troll the far-left & SNP cult members."

Of those deemed supportive of each party, a random sample of 500 was taken.

Labour

Label	Precision	Recall	F-Score	Accuracy	Coded	Prior Multiplier
Supporter (sample)	0.979	0.990	0.984		339	1
Other (sample)	0.750	0.600	0.667		69	1
Unlabelled	1685	**Features**	0	0.970		Sent out:10

UKIP

Label	Precision	Recall	F-Score	Accuracy	Coded	Prior Multiplier
Supporter (sample)	0.966	0.915	0.940		298	1
Other (sample)	0.333	0.571	0.421		129	1
Unlabelled	316	**Features**	0	0.891		Sent out:10

SNP

Label	Precision	Recall	F-Score	Accuracy	Coded	Prior Multiplier
Supporter (sample)	0.976	0.922	0.949		199	1
Other (sample)	0.563	0.818	0.667		68	1
Unlabelled	179	**Features**	3	0.911		Sent out:10

Conservative

Label	Precision	Recall	F-Score	Accuracy	Coded	Prior Multiplier
Supporter (sample)	0.921	0.829	0.872		355	1
Other (sample)	0.676	0.833	0.746		237	1
Unlabelled	1028	**Features**	6	0.830		Sent out:10

Steno

Steno is software we developed by Campbell, Moore and Ramsay at Kings College London that allows for the collection of very large amounts of news articles or tweets, and the subsequent analysis of those articles or tweets using relatively straightforward digital content tools.

Steno is written in 'Go', an open source programming language developed by Google. It consists of a server-side set of programmes that collect the textual content and metadata from each URL or tweet, and a client-side graphical user interface (GUI) desktop application for performing analysis. The server side runs continuously to collect news articles or tweets from a set of target sites or profiles. These articles or tweets are stored in a structured database, where they can be cleaned and tagged.

Using the GUI application, Steno can then pull in articles and tweets from one or more servers. Once downloaded, we can analyse them using a set of tools that allow us to sort, tag, and export them based on content, publisher, author and date/time published.

Steno also allows us to lengthen – or 'embiggen' – short links (i.e. URLs that are a shortened version of the full link), even if the short link has to bounce through multiple other short links before finding the original URL. The whole system is modular – different servers can be configured to collect different data, and the server Application Programme Interface (API) for extracting articles can be used by other tools, not just the GUI application.

Demos – Licence to Publish

The work (as defined below) is provided under the terms of this licence ('licence'). The work is protected by copyright and/or other applicable law. Any use of the work other than as authorized under this licence is prohibited. By exercising any rights to the work provided here, you accept and agree to be bound by the terms of this licence. Demos grants you the rights contained here in consideration of your acceptance of such terms and conditions.

1   Definitions

a   'Collective Work' means a work, such as a periodical issue, anthology or encyclopedia, in which the Work in its entirety in unmodified form, along with a number of other contributions, constituting separate and independent works in themselves, are assembled into a collective whole. A work that constitutes a Collective Work will not be considered a Derivative Work (as defined below) for the purposes of this Licence.

b   'Derivative Work' means a work based upon the Work or upon the Work and other pre-existing works, such as a musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgment, condensation, or any other form in which the Work may be recast, transformed, or adapted, except that a work that constitutes a Collective Work or a translation from English into another language will not be considered a Derivative Work for the purpose of this Licence.

c   'Licensor' means the individual or entity that offers the Work under the terms of this Licence.

d   'Original Author' means the individual or entity who created the Work.

e   'Work' means the copyrightable work of authorship offered under the terms of this Licence.

f   'You' means an individual or entity exercising rights under this Licence who has not previously violated the terms of this Licence with respect to the Work, or who has received express permission from Demos to exercise rights under this Licence despite a previous violation.

2   Fair Use Rights

Nothing in this licence is intended to reduce, limit, or restrict any rights arising from fair use, first sale or other limitations on the exclusive rights of the copyright owner under copyright law or other applicable laws.

3   Licence Grant

Subject to the terms and conditions of this Licence, Licensor hereby grants You a worldwide, royalty-free, non-exclusive,perpetual (for the duration of the applicable copyright) licence to exercise the rights in the Work as stated below:

a   to reproduce the Work, to incorporate the Work into one or more Collective Works, and to reproduce the Work as incorporated in the Collective Works;
b   to distribute copies or phonorecords of, display publicly, perform publicly, and perform publicly by means of a digital audio transmission the Work including as incorporated in Collective Works; The above rights may be exercised in all media and formats whether now known or hereafter devised. The above rights include the right to make such modifications as are technically necessary to exercise the rights in other media and formats. All rights not expressly granted by Licensor are hereby reserved.

4   Restrictions

The licence granted in Section 3 above is expressly made subject to and limited by the following restrictions:

a   You may distribute, publicly display, publicly perform, or publicly digitally perform the Work only under the terms of this Licence, and You must include a copy of, or the Uniform Resource Identifier for, this Licence with every copy or phonorecord of the Work You distribute, publicly display, publicly perform, or publicly digitally perform. You may not offer or impose any terms on the Work that alter or restrict the terms of this Licence or the recipients' exercise of the rights granted hereunder. You may not sublicence the Work. You must keep intact all notices that refer to this Licence and to the disclaimer of warranties. You may not distribute, publicly display, publicly perform, or publicly digitally perform the Work with any technological measures that control access or use of the Work in a manner inconsistent with the terms of this Licence Agreement. The above applies to the Work as incorporated in a Collective Work, but this does not require the Collective Work apart from the Work itself to be made subject to the terms of this Licence. If You create a Collective Work, upon notice from any Licencor You must, to the extent practicable, remove from the Collective Work any reference to such Licensor or the Original Author, as requested.

b   You may not exercise any of the rights granted to You in Section 3 above in any manner that is primarily intended for or directed toward commercial advantage or private monetary compensation. The exchange of the Work for other copyrighted works by means of digital filesharing or otherwise shall not be considered to be intended for or directed toward commercial advantage or private monetary compensation, provided there is no payment of any monetary compensation in connection with the exchange of copyrighted works.

C   If you distribute, publicly display, publicly perform, or publicly digitally perform the Work or any Collective Works, you must keep intact all copyright notices for the Work and give the Original Author credit reasonable to the medium or means You are utilizing by conveying the name (or pseudonym if applicable) of the Original Author if supplied; the title of the Work if supplied. Such credit may be implemented in any reasonable manner; provided, however, that in the case of a Collective Work, at a minimum such credit will appear where any other comparable authorship credit appears and in a manner at least as prominent as such other comparable authorship credit.

5   Representations, Warranties and Disclaimer

A   By offering the Work for public release under this Licence, Licensor represents and warrants that, to the best of Licensor's knowledge after reasonable inquiry:

i   Licensor has secured all rights in the Work necessary to grant the licence rights hereunder and to permit the lawful exercise of the rights granted hereunder without You having any obligation to pay any royalties, compulsory licence fees, residuals or any other payments;

ii   The Work does not infringe the copyright, trademark, publicity rights, common law rights or any other right of any third party or constitute defamation, invasion of privacy or other tortious injury to any third party.

B   except as expressly stated in this licence or otherwise agreed in writing or required by applicable law, the work is licenced on an 'as is' basis, without warranties of any kind, either express or implied including, without limitation, any warranties regarding the contents or accuracy of the work.

6   Limitation on Liability

Except to the extent required by applicable law, and except for damages arising from liability to a third party resulting from breach of the warranties in section 5, in no event will licensor be liable to you on any legal theory for any special, incidental, consequential, punitive or exemplary damages arising out of this licence or the use of the work, even if licensor has been advised of the possibility of such damages.

7   Termination

A   This Licence and the rights granted hereunder will terminate automatically upon any breach by You of the terms of this Licence. Individuals or entities who have received Collective Works from You under this Licence, however, will not have their licences terminated provided such individuals or entities remain in full compliance with those licences. Sections 1, 2, 5, 6, 7, and 8 will survive any termination of this Licence.

B   Subject to the above terms and conditions, the licence granted here is perpetual (for the duration of the applicable copyright in the Work). Notwithstanding the above, Licensor reserves the right to release the Work under

different licence terms or to stop distributing the Work at any time; provided, however that any such election will not serve to withdraw this Licence (or any other licence that has been, or is required to be, granted under the terms of this Licence), and this Licence will continue in full force and effect unless terminated as stated above.

8   Miscellaneous

A   Each time You distribute or publicly digitally perform the Work or a Collective Work, Demos offers to the recipient a licence to the Work on the same terms and conditions as the licence granted to You under this Licence.

B   If any provision of this Licence is invalid or unenforceable under applicable law, it shall not affect the validity or enforceability of the remainder of the terms of this Licence, and without further action by the parties to this agreement, such provision shall be reformed to the minimum extent necessary to make such provision valid and enforceable.

C   No term or provision of this Licence shall be deemed waived and no breach consented to unless such waiver or consent shall be in writing and signed by the party to be charged with such waiver or consent.

D   This Licence constitutes the entire agreement between the parties with respect to the Work licensed here. There are no understandings, agreements or representations with respect to the Work not specified here. Licensor shall not be bound by any additional provisions that may appear in any communication from You. This Licence may not be modified without the mutual written agreement of Demos and You.

Alex Krasodomski-Jones is a Researcher at in the Centre for the Analysis of
Social Media at Demos. His research interests include political extremism and its
reportage on social media. He manages CASM's analytics capability,
including data collection, analytics and visualisation.