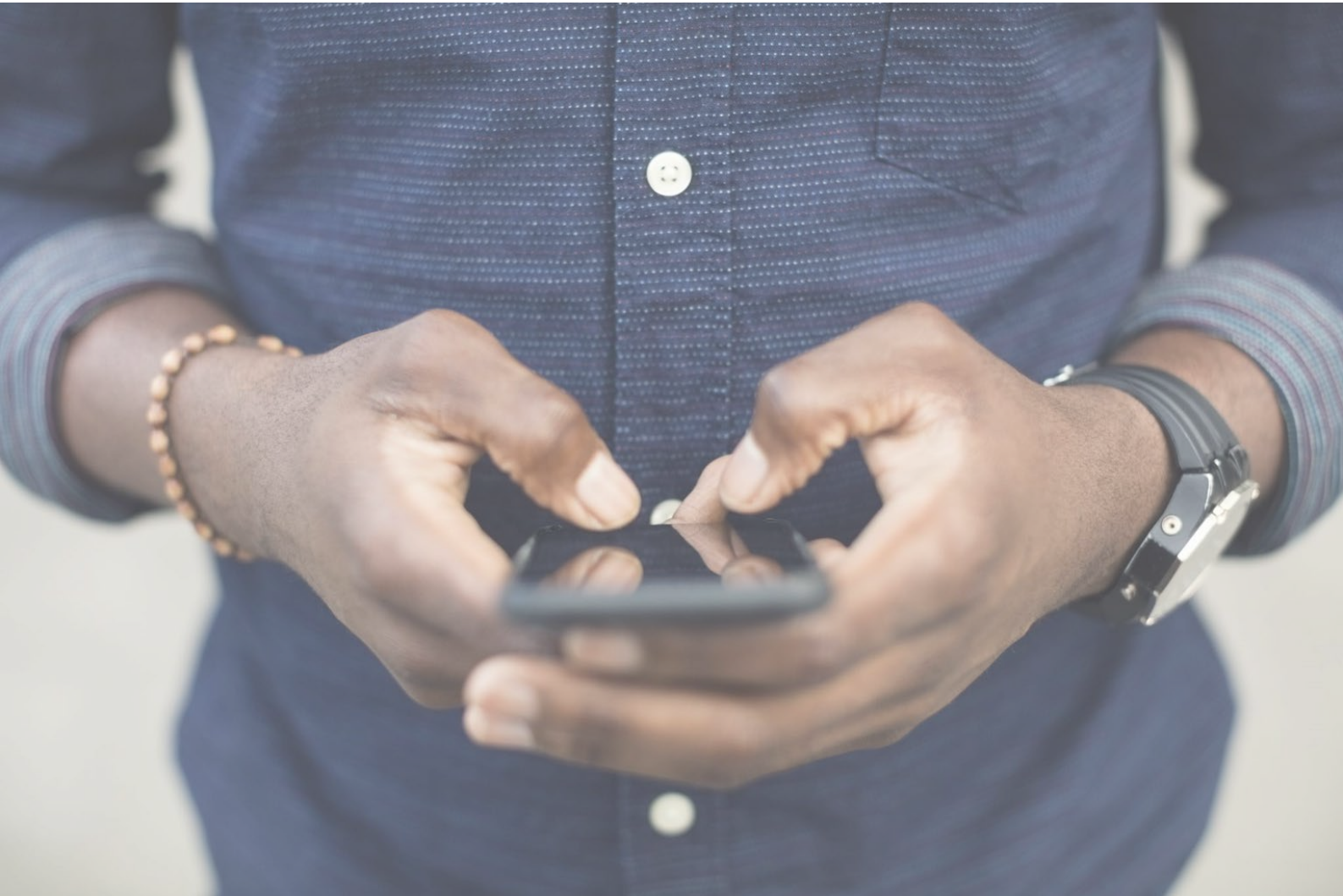


News consumption and media plurality on Twitter in the UK

Economics discussion paper series, issue number 5



Ofcom economics discussion paper series in communications regulation

The discussion paper series

Ofcom is committed to encouraging debate on all aspects of media and communications regulation and to creating rigorous evidence to support its decision-making. One of the ways we do this is through publishing a series of Discussion Papers, extending across economics and other disciplines. The research aims to make substantial contributions to our knowledge and to generate a wider debate on the themes covered.

Disclaimer

Discussion Papers contribute to the work of Ofcom by providing rigorous research and encouraging debate in areas of Ofcom's remit. Discussion Papers are one source that Ofcom may refer to, and use to inform its views, in discharging its statutory functions. However, they do not necessarily represent the concluded position of Ofcom on particular matters.

Contents

Section

1. Overview	4
2. News consumption and engagement with news on Twitter	9
3. Twitter API data	12
4. Descriptive news consumption and media plurality metrics	16
5. Homophily as an echo-chamber metric	25
6. Conclusion	35

Annex

A1. Definition of the homophily metric	37
--	----

1. Overview

- 1.1 Media Plurality is a cornerstone of a well-functioning democratic society, in which citizens are exposed to a wide range of viewpoints from a variety of different sources, and where any one media owner, or voice, does not have too much influence over public opinion and the political process. Ofcom's statutory duties include securing and maintaining a sufficient plurality of providers of different TV and radio services.¹ Ofcom also has a duty to conduct periodic reviews of the operation of the media ownership rules listed under section 391 of the Communications Act 2003.
- 1.2 Since Ofcom took on these duties in 2003, however, news consumption in the UK has changed considerably. In particular, online news sources have grown in importance and online intermediaries, including social media, search engines, and news aggregators, have become key channels for consumers to access news, especially among younger people. Ofcom's News Consumption Survey from 2022, for example, reports that social media is the second largest medium for news consumption, used by 46% of UK adults (after 74% for TV) and by 79% of 16-24 year olds.²
- 1.3 In June 2021, Ofcom therefore launched a programme of work to understand the impact of these changes on media plurality.³ Following an initial call for evidence, Ofcom's statement published in November 2021 highlighted several risks to media plurality associated with online intermediaries and we subsequently embarked on an assessment of these risks and potential remedies.⁴ The findings from this assessment are being published alongside this Economics Discussion Paper in Ofcom's discussion document titled "Media Plurality and Online News" (henceforth the "Media Plurality Discussion Document").
- 1.4 The research presented in this Economics Discussion Paper complements this broader review of media plurality in the UK by focusing on one prominent online intermediary used to access news, Twitter. We focus on Twitter because it makes a rich dataset publicly available through its Application Programming Interface (API), and the focus of our analysis is to explore how we can measure news consumption and elements of media plurality on Twitter using this API. Our work is therefore primarily methodological and exploratory in nature.
- 1.5 Our approach is guided by Ofcom's existing media plurality measurement framework.⁵ Specifically, we set out to answer the following research questions:

¹ See Section 3(2)(d) of the Communications Act 2003.

² See Ofcom (2022), "News Consumption in the UK: 2022", accessed on 27 August 2022 at https://www.ofcom.org.uk/data/assets/pdf_file/0027/241947/News-Consumption-in-the-UK-2022-report.pdf.

³ See Ofcom (June 2021) "The Future of Media Plurality in the UK", accessed on 26 August 2022 at https://www.ofcom.org.uk/data/assets/pdf_file/0012/220710/media-plurality-in-the-uk-condoc.pdf

⁴ See Ofcom (November 2021) "The Future of Media Plurality in the UK", accessed on 13 October 2022 at https://www.ofcom.org.uk/data/assets/pdf_file/0019/228124/statement-future-of-media-plurality.pdf

⁵ See Ofcom (2015), "Measurement framework for media plurality", accessed on 19 August 2022 at https://www.ofcom.org.uk/data/assets/pdf_file/0024/84174/measurement-framework-for-media-plurality-statement.pdf.

- a) Which metrics relating to news consumption and media plurality can be computed using public data from Twitter's API and what are the key limitations?
- b) According to these metrics, what is the presence of news outlets and what patterns of news consumption exist on Twitter?
- c) Is there evidence of homophily (similarities within communities of users) in relation to news consumption on Twitter?⁶
- d) What additional data would be required to explore media consumption and media plurality more fully, both on Twitter and other social media platforms?

What we have found – in brief

The data available from the Twitter API can be used to compute meaningful metrics that relate to news consumption and elements of media plurality as captured by Ofcom's existing media plurality measurement framework. We measure the activity of news outlets on Twitter according to the number of published tweets. Users' subscription to news outlets (i.e., following) and interactions with news content (liking) provide proxies for consumption. "Following" a news outlet on Twitter appears to offer a more robust basis for these metrics given that user interactions with news content tend to be very rare.

While the Twitter API is a rich source of valuable information, key aspects of news consumption on Twitter remain unobservable using this API. In particular, there is no publicly available data on the content viewed by users. Proxies for this that are available through the API are imperfect, with some measures capturing content that is unlikely to have been viewed and others not capturing content that is likely to have been viewed.

The data from the Twitter API can be used to measure homophily – i.e., the tendency to associate with other people on a network who are similar – as one possible indicator for the presence of echo chambers. We compute the homophily metric based on subscriptions (i.e., following) and interactions (liking). Although preliminary, the analysis shows only limited evidence of homophily among most Twitter users. There are, however, some users whose networks exhibit relatively high levels of homophily according to our metric and these tend to follow/like the content of only a few news outlets.

Given the limited scope of this exploratory paper, we highlight several areas for potential further research. These include further research on Twitter using additional information that is available via the API (e.g., the content of tweets) and data that is not publicly available (e.g., tweets viewed, which would provide a more direct measure of consumption). In addition, there is also scope for conducting research across a broader range of social media platforms, such as Facebook and TikTok, to obtain a more holistic picture of news consumption and media plurality within social networks.

⁶ Homophily here refers to the behaviour of users seeking out and engaging with users that are like them. We provide an explanation of homophily in the context of media plurality in Section 5.

- 1.6 We concentrate on Twitter for the following two reasons.
- a) Twitter represents one of the most important sources of news in the UK. According to Ofcom’s News Consumption Survey, 17% of adults aged 16+ years used Twitter to access news in 2022, a higher share than that attributed to any traditional print newspaper and several TV and radio channels.⁷ In addition, among those using social media to access news, Twitter ranks second (37% using the app) after only Facebook (69%).
 - b) Twitter offers detailed publicly available data, relative to other popular social media platforms used to access news, via the company’s API.⁸ The API allows us to retrieve detailed data on news outlets and other Twitter users that are relevant in the context of news consumption.⁹
- 1.7 In an effort to limit the data required and corresponding computational requirements, we restricted our study to the activity of a small subset of news producers on Twitter and users who followed them or liked their content during the month of December 2021. Despite this restricted focus, the data that needed to be retrieved from the Twitter API and processed still comprised close to 100 million rows of user information, 6.5 million follower relationships among UK users alone, and over 1.3 million likes.
- a) We focus on the Twitter accounts of 12 major news outlets that have a strong presence in traditional media: BBC News, Channel 4 News, CNN, The Economist, the Financial Times, the Guardian, the Independent, Mail Online, the New York Times, Sky News, the Telegraph, and the Times.¹⁰
 - b) In terms of news consumers, we consider Twitter users that followed at least one of these 12 outlets or interacted with their content in December 2021. We focus on users that are located in the UK.¹¹
- 1.8 We approach the above questions concerning measurement of news consumption on Twitter through the lens of Ofcom’s existing media plurality measurement framework, which consists of four categories: availability, consumption, impact, and contextual factors.¹² These measures are well established in the context of the traditional news media of broadcasting, and print news.
- 1.9 The availability, consumption, and impact of news are typically measured using a range of quantitative metrics, such as number of providers, reach of audience, share of reference (a

⁷ See *supra* note 2.

⁸ See Twitter API documentation available at <https://developer.twitter.com/en/docs/twitter-api>.

⁹ Twitter has the only publicly available API that has the information required to do this analysis. While the Twitter API is a rich source of information, it has some limitations for our analysis. We discuss those limitation in more detail in the next sub-section.

¹⁰ We selected the news outlets taking into account factors such as their importance in the Reuters Digital News Report and their followership on Twitter. More detail is given in paragraph 3.5.

¹¹ The location of Twitter users can only be retrieved from an optional account field. While this field is filled in for the majority of followers of the 12 news outlets, it is not guaranteed to contain the actual location of the user. We also did not capture the location information for users that interacted with tweets by the 12 news outlets but did not follow any of the outlets.

¹² See *supra* note 5.

metric developed by Ofcom as a proxy for consumption), multi-sourcing, and personal importance. These measures are typically captured through industry measurement systems and consumer surveys. In this paper, we explore similar metrics to measure media plurality on Twitter regarding the 12 above-mentioned news outlets.

- 1.10 Contextual factors, which are qualitative descriptions of the market and news organisations, include internal plurality, governance, editorial policy, impartiality requirements, and future market developments, among others. In this paper, we do not discuss contextual factors regarding specific news outlets. We do, however, describe key features of Twitter and online intermediaries more generally that are important for the interpretation of the quantitative metrics for plurality (e.g., the role of algorithms as an editorial mechanism).

Limitations of the analysis

- 1.11 It is important to note that the metrics presented in this paper cannot be interpreted as findings about the extent of media plurality on social media. Instead, our findings provide insights about the extent to which measurement of aspects of media plurality on Twitter is possible using the publicly available data from Twitter's API.¹³
- 1.12 There are two key reasons for this limited interpretation:
- a) One reason is our focus on Twitter, which is valuable as a case study but does not provide information about media plurality in the online space more generally let alone media plurality across traditional and online news media.¹⁴ As consumers tend to use multiple news sources online and offline, assessments of media plurality need to ultimately incorporate these different sources to arrive at an aggregate view.
 - b) Another reason is our focus on the Twitter accounts of 12 major news outlets, which represent a small subset of a potentially very large number of news creators more generally. For the metrics to provide insights for media plurality (even just on Twitter), this list of news producers would first have to be expanded to capture the accounts of further editorial outlets. In addition, the list could be further expanded to capture other types of news creators, such as individual journalists, politicians, celebrities, or other organisations.
- 1.13 There are also key limitations for measuring media plurality that are rooted in the information that is available via the Twitter API. Most fundamentally, we do not have access via the Twitter API to any measure of "impressions" – i.e., the views that news

¹³ Some avenues for future research as discussed in Section 6. See also Ofcom's November 2022 Media Plurality Discussion Document on accompanying this paper for a broader discussion of research we propose to undertake in the next phase of our work.

¹⁴ Twitter is the only platform that has a publicly available API allowing us to undertake this analysis.

content received – and who viewed particular content.¹⁵ This precludes us from analysing news consumption directly.¹⁶

- 1.14 Finally, we do not attempt to present an exhaustive list of media plurality metrics in this paper. Due to the exploratory nature of the work, the metrics presented are necessarily a subset of those that might be useful in a comprehensive assessment of media plurality. Specifically, we focus on metrics that closely relate to the existing media plurality measurement framework, but we do not include all possible metrics or variations thereof for all available data from the Twitter API. In addition, when we present metrics that relate to the potential presence of echo chambers, we focus on “homophily” – i.e., the tendency to associate with others who are similar – as one aspect of echo chambers, but do not explore metrics that relate to other aspects of echo chambers such as filter bubbles or high levels of segregation.¹⁷

¹⁵ While we have information on likes, quotes and retweet, which indicate interaction with news content on Twitter by users, this behaviour is very rare in our data.

¹⁶ Even information on the number of views received by tweets, of course, would not capture more nuanced notions of news consumption, such as the time spent viewing a tweet and the corresponding share of attention given to the tweets by some outlets over others.

¹⁷ For a general definition of the term “homophily” and related discussion see Currarini et al. (2009), “An Economic Model of Friendship: Homophily, Minorities, and Segregation”, *Econometrica* 77(4):1003-1045.

2. News consumption and engagement with news on Twitter

- 2.1 Users can access news on Twitter in a variety of ways. According to Ofcom’s News Consumption Survey from 2022, browsing trending topics or following other users are the most prominent ways Twitter users access news on the platform.¹⁸ We describe the different ways of accessing news in this subsection in order to clarify which aspects of news consumption and engagement with news on Twitter are observable via the Twitter API, and to explain some of the limitations of the present analysis and potential extensions in the future.
- 2.2 Twitter is a social media platform that allows users to post and interact with short messages called “tweets”.¹⁹ On the platform, users can follow other users, which could be individuals, companies, or other entities. “Following” in this context refers to one Twitter account subscribing to the tweets of another.²⁰ The other Twitter accounts that a given user follows are called that user’s “friends”.²¹ Following allows users to see the tweets of the followed accounts – i.e., their friends – on their Home timeline as soon as the latter post something new. Content is accessed by users on Twitter through their timeline, which serves tweets from other users or topics that a user follows, or other screens, like the Explore tab, which allows the user to browse tweets according to topics or search for particular users or content.
- 2.3 Users can interact with original tweets by liking, retweeting, and quoting. “Liking” is an action that allows Twitter users to show that they have an appreciation for a particular tweet. “Retweeting” is an action that allows Twitter users to share another account’s tweet with their followers; “quoting” is the same action with the addition of the user’s own content (e.g., comments on the original tweet).²² Retweeting and quoting are important, because they are one of the major ways to propagate tweets to the timeline of users who do not follow the account that originated the tweet.
- 2.4 Figure 1 shows how two different users may access the same news content from a single original news outlet on Twitter. Because user 1 follows the news outlet, they receive the outlet’s tweet on their timeline. Although user 2 has no direct connection through following with the news outlet, they may still be able to access the content in other ways, but this will also depend on the popularity of the tweet and the actions of other users.

¹⁸ See supra note 2.

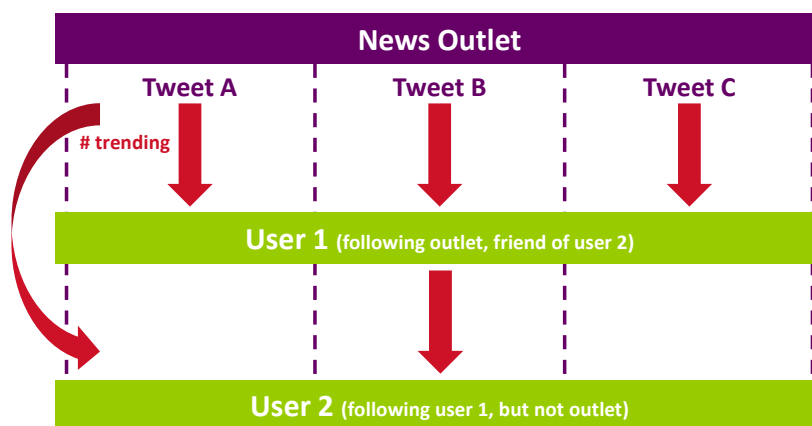
¹⁹ Each tweet has a maximal length of 280 characters, but multiple tweets can be linked to each other as a longer “thread”.

²⁰ See Twitter glossary available at <https://help.twitter.com/en/resources/glossary>.

²¹ Unlike some other social media platforms, connections between users on Twitter are asymmetrical in the sense that one user might follow another, but not the other way around. This means that there is a fundamental distinction between a follower and a friend on Twitter, the latter representing the followed Twitter account. “Follower” and “friend” are terms used by Twitter to make this distinction. See, for example, Twitter documentation on the API endpoint GET friends/ids available at <https://developer.twitter.com/en/docs/twitter-api/v1/accounts-and-users/follow-search-get-users/api-reference/get-friends-ids>.

²² See supra note 20.

Figure 1: Accessing news on Twitter



Notes: Figure illustrates major ways for users to access news on Twitter. This is not a comprehensive illustration of all such ways. Source: Ofcom.

2.5 Figure 1 shows three different kinds of tweets that reflect these factors: tweet A belongs to a trending topic, tweet B does not belong to a trending topic but is retweeted/quoted by user 1, and tweet C does not belong to a trending topic and is not retweeted/quoted by user 1. “Trending” topics are those that are especially popular on Twitter at a given point in time and show high activity in terms of tweets and views. Tweets that belong to trending topics tend to be easier to find than other tweets.

2.6 Now we consider each of the main ways of accessing news on Twitter in turn and describe how these relate to Figure 1.²³

- a) Tweets that belong to trending topics are represented by tweet A in Figure 1. This tweet will appear on the timeline of user 1, because they follow the news outlet that originated tweet A. In addition, it will tend to appear higher up on the timeline of user 1 in the default “Home” view, which is algorithmically curated, favouring tweets that are more popular in general or among similar (and/or possibly connected) accounts. Tweet A will usually not appear on the timeline of user 2, because they do not follow the news outlet that originated the tweet.²⁴ User 2 may, however, find the tweet under trending topics if they browse tweets or search via Twitter’s Explore tab.²⁵ The fundamental insight for news tweets that belong to trending topics is that they are generally accessible to both types of users, those who follow the originating account (news outlet) and those who do not.

²³ While Figure 1 illustrates key channels through which consumers may access news from the Twitter account of a traditional news outlet, similar channels would also apply to other news producers, such as journalists or public figures. In this case, the account shown in purple that originates the tweets would be different (e.g., journalist instead of news outlet), but the relevant types of tweets and consumers would still apply.

²⁴ There are tweets that appear on a user’s timeline even if they do not follow the account originating the tweet. These include tweets that belong to topics the user follows and recommended tweets, based on their popularity and interactions with the tweet within the user’s Twitter network. See Twitter Help Center article “About your Home timeline on Twitter”, accessed on 14 October 2022 at <https://help.twitter.com/en/using-twitter/twitter-timeline>.

²⁵ Twitter’s Explore tab shows general trending topics and trending news specifically. See Twitter glossary referenced in supra note 20 for a description of the Explore tab.

- b) Tweets received from other users followed – i.e., “friends” – are represented by tweet B in Figure 1. This tweet will appear on the timeline of user 1, because they follow the news outlet that originated the tweet. It will also appear on the timeline of user 2, because they follow user 1 who, in turn, retweeted/quoted the tweet. The fundamental insight here is that news tweets that do not belong to trending topics may still appear on the timeline of users that do not follow the originating account (news outlet); this happens when other accounts they follow retweet or quote the original tweet.
 - c) Tweets received by actively following the originating account (news outlet) are represented by tweet C in Figure 1. This tweet will appear on the timeline of user 1, because they follow the news outlet that originated the tweet. Tweet C will, however, usually not appear on the timeline of user 2, because they do not follow the news outlet that originated the tweet. While user 2 could, in principle, find tweet C, this would require rather specific action on their part in the absence of that tweet trending (e.g., searching for the news outlet’s account and viewing its posts). The fundamental insight here is that news tweets that do not belong to trending topics tend to be available mostly to users who follow the originating account (news outlet).
- 2.7 Most of these ways in which users access news on Twitter involve some type of algorithmic curation of the content a given user sees. This applies to the trending topics and underlying tweets that the user is shown when they explore content outside their timeline. It also typically applies to the order of tweets by any Twitter accounts they follow – including those by original news producers and retweets/quotes by other users.²⁶
- 2.8 Beyond the presentation of tweets, algorithms also play a role in suggesting other Twitter accounts to follow and therefore in the establishment of followership links that may bring about exposure to news tweets in the first place. This has implications for the interpretation of our analysis because algorithms take user characteristics and actions as inputs, but also affect these. These topics are discussed in further detail in the following sections.

²⁶ Users have the option of viewing tweets in chronological order on their timeline rather than in the default top-tweets-first view. See Twitter Help Center article “About your Home timeline on Twitter”, accessed on 14 October 2022 at <https://help.twitter.com/en/using-twitter/twitter-timeline>.

3. Twitter API data

- 3.1 Twitter provides public access to a wide range of data about users and content on the platform via an API.²⁷ Specifically, for the analysis in this paper, it provides information on the activity of news outlets, the followership relationships between Twitter accounts (i.e., between users and news outlets and among users), and users' interactions with news tweets (i.e., likes, retweets, quotes, and replies) that relate to engagement with and propagation of news content.
- 3.2 In addition to the information used in the present analysis, the Twitter API also offers other data that could be used for future research. These include, notably, the content of tweets, retweets, quotes, and replies. We discuss possible further avenues of research in the concluding Section 6.
- 3.3 Despite the rich data available through the Twitter API, especially relative to other social media platforms, some important information for the analysis of news consumption and media plurality is not available. Most importantly, there is no data on users' viewing of individual tweets – i.e., “impressions” in Twitter terminology – such that news consumption is not directly observable. In this paper, we therefore rely on following and liking as proxies of news consumption.

Data retrieved from the Twitter API for the analysis

- 3.4 The analysis presented in this paper relies on data retrieved from the Twitter API that relates to the Twitter accounts of 12 major news outlets that also have a strong presence in the traditional media: BBC News, Channel 4 News, CNN, The Economist, the Financial Times, the Guardian, the Independent, Mail Online, the New York Times, Sky News, the Telegraph, and the Times.²⁸
- 3.5 These outlets were chosen to broadly reflect major news outlets used by UK consumers that also have a strong presence on Twitter. To construct this list, we started with the top online news outlets in the UK according to the Reuters “Digital News Report 2021” and complemented this with additional outlets listed in other sources, such as Ofcom’s “News Consumption in the UK: 2021” report.²⁹ Among these, we selected the news outlets with the highest number of Twitter followers. The twelve outlets chosen, provide an

²⁷ There are different Twitter API subscriptions that vary in scope of the data that can be retrieved and in the specificity of possible search queries. For example, the Premium APIs, which we are using, allow for some full-archive searches of all tweets but searches focused on related tweets that are part of a “conversation” (e.g., all replies, quotes, retweets relative to some original tweet) are only possible on the last seven days. An alternative subscription reserved for academic research purposes, by contrast, allows for such a focused search on the full archive of tweets.

²⁸ For the remainder of this paper, we refer to the Twitter account of these news outlets by the outlet name, unless otherwise specified. For example, when referring to BBC News, we mean the Twitter account of BBC News, not the BBC News organisation more generally.

²⁹ See Reuters Institute (2021) “Digital News Report 2021”, accessed on 11 November 2022 at https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2021-06/Digital_News_Report_2021_FINAL.pdf and Ofcom (2021) “News Consumption in the UK: 2022”, accessed on 11 November 2022 at https://www.ofcom.org.uk/___data/assets/powerpoint_doc/0026/222479/news-consumption-in-the-uk-2021-report.pptx.

appropriate basis for our exploratory analysis but this list is not intended to be fully representative of news outlets used by UK consumers in general.

- 3.6 For each of these 12 news outlets, we retrieved all tweets published during the month of December 2021.³⁰ These data provide information on the tweet content, the originating account (i.e., the news outlet), and the publication date. They also include aggregate metrics on the number of likes, retweets, quotes, and replies that the tweet had received by the time of retrieval.
- 3.7 Furthermore, we retrieved data on the likes that these news tweets received to measure user engagement with the news outlets.³¹ These data specify the liked tweet and the user liking the tweet. Importantly, in the context of retweets, such as tweet B in Figure 1 above, likes are attributed to the original tweet, so that these data capture the likes irrespective of the way by which the user accessed the tweet.
- 3.8 In relation to consumers of news content, we retrieved data on the followers of the 12 news outlets in December 2021.³² This is the core group of users that we consider for the descriptive media plurality metrics involving followership in Section 4. Beyond the information on their followership regarding the news outlets, these data capture detailed user characteristics, including their (self-reported) location and aggregate metrics on their Twitter activity at the time of retrieval (e.g., the number of friends, followers, and tweets). We rely on the user characteristics to narrow down the full list of followers to those that are most meaningful for our analysis.
- a) We use the location to identify Twitter users from the UK, although this classification is imperfect due to its being self-reported and not required for every account.
 - b) We use the metrics on Twitter activity to exclude accounts that appear to be inactive (i.e., have low activity) or that appear to belong to professional Twitter users or organisations for parts of the analysis.
- 3.9 Finally, we also retrieved the friends of some of the news-outlet followers – i.e., Twitter accounts that the followers of news outlets follow.³³ In combination with the other data, this allows us to compare how users that are connected on Twitter relate to news outlets in terms of following and liking of their content. This is necessary for the analysis of homophily, which is one potentially underlying feature of echo chambers in news consumption. We collect a random sample of the followers of news outlets and retrieve their social network of friends to reduce the time to collect the data and the computational burden.³⁴ We did this separately for the following two user groups.
- a) The first group comprises followers of any of the 12 news outlets that are active and not professional Twitter users or organisations. This group represents users that we

³⁰ We retrieved tweets using the GET /2/users/:id/tweets API endpoint via the Python tweepy library.

³¹ We retrieved likes using the GET /2/tweets/:id/liking_users API endpoint via the Python tweepy library.

³² We retrieved followers using the GET followers/ids API endpoint via the R rtweet library. Also, we retrieved user characteristics using the GET users/lookup API endpoint via the R rtweet library.

³³ We retrieved friends using the GET friends/ids API endpoint via the R rtweet library.

³⁴ Random sampling is used to ensure representativeness of the news-outlet followers.

consider for the homophily measure in Section 5 based on followership. Friends of Twitter users in this group were retrieved in March 2022.

- b) The second group comprises news-outlet followers from the first group that also liked at least one tweet by one of the 12 news outlets from December 2021. This group represents users that we consider for the homophily measure in Section 5 based on liking. Friends of Twitter users in this group were retrieved in May 2022.

Suitability of Twitter API data for news consumption and media plurality metrics

- 3.10 The data retrieved via the Twitter API allow us to compute several descriptive measures of news consumption and media plurality that relate to the three quantitative categories of Ofcom's measurement framework.
- 3.11 The *availability* of news sources on Twitter can be captured by information on the activeness of news outlets, in terms of the number of original tweets. The number of tweets notably captures the degree of activity of news organisations, going beyond simply counting the number of outlets. In addition, the data from the Twitter API could, in principle, also be used to measure the propagation of news content (e.g., via retweets and quotes), although this extends beyond the scope of the analysis in the present paper.
- 3.12 The *consumption* of news on Twitter cannot be captured well with the publicly available data. Specifically, there is no direct information on whether a consumer viewed a given tweet and we therefore rely on following news outlets and liking their content as proxies for consumption.³⁵ Despite their limitations, these proxies still shed some light on important topics, such as the reach of news outlets and evidence of consumers' multi-sourcing.
 - a) Following a news outlet captures one important way in which consumers can access news content on Twitter (corresponding to user 1 in Figure 1 above). It does, however, not guarantee that the user actually views that outlet's content as the sequence of tweets on the timeline is typically curated by algorithms. It also does not capture consumption of content from news outlets that is accessed by non-followers (corresponding to user 2 in Figure 1 above).
 - b) Liking of content by news outlets, on the other hand, implies that the content has been viewed and captures followers and non-followers. It does, however, not apply to all viewed content given that users are expected to only like a small fraction of viewed content. In addition, viewed and liked tweets are also systematically different than viewed and not-liked tweets in that the former often reflect some level of agreement

³⁵ While liking of content is a direct measure of news consumption – a liking user is highly likely to have read the Tweet the user is liking – it is not a comprehensive measure because more users may have read a Tweet, but not have liked it.

with the content. Consumption of news that challenges a consumer's opinions may therefore not be well reflected in liking data.³⁶

3.13 One dimension of the *impact* of news outlets on Twitter can be captured relatively well by the data on liking. The number of tweets of a news outlet liked by a given consumer provides a measure of that outlet's personal importance to the consumer. The liking data, however, does not capture more fundamental aspects of the impact of news content on Twitter on the user's views. We note that users' original content included with quotes and replies could, in principle, be used to measure their sentiment vis-à-vis particular news outlets or topics. This, in turn, could allow us to infer the consumers' perception of the impartiality, reliability, and trust of news outlets and shed further light on potential areas of concern, such as polarization. Such analysis, however, extends beyond the scope of the present paper.

3.14 Finally, the data available from the Twitter API also allow for important metrics that extend beyond the evidence covered by Ofcom when applying the media plurality measurement framework. Specifically, the information on the social network between Twitter users allows us to compute metrics of homophily, which is one potential underlying feature of echo chambers.³⁷ If connected news consumers have a similar followership/liking profile vis-à-vis news outlets, then the social network and algorithms that use the network as input for recommendations may reinforce exposure to a narrower set of news and views. Conversely, if connected consumers' preferences over news tend to be dissimilar, then the social network may enhance plurality by expanding exposure to a diverse set of news.

³⁶ Some users may also like content in a way that does not represent agreement, e.g., as a means to "bookmark" tweets they would like to find again. In these cases, liking would be a broader indicator of tweets that are of interest to the user.

³⁷ See Section 5 and Ofcom's concurrent Media Plurality Discussion Document for further details.

4. Descriptive news consumption and media plurality metrics

- 4.1 We have used the publicly available data from the Twitter API to compute metrics that relate to Ofcom's existing media plurality measurement framework. These capture the activeness of the 12 news outlets on Twitter, their UK followership, and the liking of their content. As discussed in the previous section, these metrics do not address all aspects of the measurement framework – notably, news consumption cannot be observed directly – but they nonetheless provide valuable insights into the measurement of news consumption and elements of media plurality on Twitter.

Activity of news outlets on Twitter

- 4.2 We have measured the activity of news outlets on Twitter according to the number of tweets they published and their share out of the total such news tweets. In relation to Ofcom's existing media plurality measurement framework, these metrics reflect the availability of news sources on Twitter. Indeed, they provide more detail than standard metrics of availability that involve merely counting the number of outlets. This is particularly important in the context of social media, where setting up an account is essentially costless and where the delineation of what constitutes relevant news sources is potentially ambiguous (e.g., among journalists and public figures).
- 4.3 Table 1 below shows the number of tweets by news outlet and the outlet share of the total tweets across all the outlets in our analysis during December 2021. According to these metrics, there is great variation in the level of activity between outlets. The Independent stands out as the most prolific, tweeting over 11,000 times and almost twice as much as the second-most tweeting outlet, the Guardian. Most of the news outlets published between 2,000 and 6,000 tweets during the month. The Financial Times and Channel 4 News were the least active on Twitter with only around 860 and 400 tweets during the month respectively.

Table 1: Number of Tweets by news outlet

News Outlet	Number of Tweets (count)	Outlet share (%)
Independent	11,309	24.3
Guardian	5,694	12.2
Mail Online	5,445	11.7
BBC News	4,738	10.2
Sky News	4,223	9.1
Telegraph	4,154	8.9
The Economist	3,329	7.1
Times	2,444	5.2
CNN	2,244	4.8
New York Times	1,721	3.7
Financial Times	859	1.8
Channel 4 News	406	0.9
Total	46,566	100.0

Notes: Table reports the number of tweets of the listed news outlets in December 2021 and the outlets' share out of the total number of tweets of these outlets. **Source:** Ofcom analysis of Twitter data.

4.4 Although the number of tweets of news outlets provides more insights into their activity than metrics that merely capture their presence, this measure still provide only limited insight into the availability of news and news consumption on Twitter. In particular, in the absence of analysis of the content of tweets, it is not clear how much information a single tweet contains. Tweets that include links to entire articles on the news outlets' websites, for instance, may be short, but point towards very detailed news content. Individual tweets without external references, by contrast, may provide very limited information or must be considered in the context of other tweets that are connected via a "thread". These underlying contextual features obscure the precise meaning of a given number of tweets. Analysis of the content of tweets is, in principle, possible based on the publicly available data from the Twitter API, but it extends beyond the scope of the present paper.

Following of news outlets

4.5 The followership of news outlets is useful for two aspects of measuring media plurality on Twitter. First, it represents one of the key ways by which users may access news content, since tweets by the news outlets that a user follows appear on that user's timeline. As such, it can be interpreted as a proxy for consumption, albeit an imperfect one, because not all tweets on a user's timeline are necessarily viewed and because there are other important ways for accessing news content on Twitter (see Section 2 above for details). Second, following a news outlet reflects a general interest in that outlet's content even if it does not directly correspond to consumption of all its content. In this sub-section, we explore several metrics of followership, including the number of followers, outlet reach, and statistics on multi-sourcing.

- 4.6 Table 2 below shows the number of UK followers by news outlet and the outlet's reach out of the total UK followers across the 12 news outlets. Given that the location is self-reported by Twitter users and not all users in the UK will choose to report their location (accurately), this table likely understates the number of UK followers. The absolute number of followers of each outlet is therefore less meaningful than the number of followers relative to one another that is reflected in their ranking and the outlet reach metrics.³⁸

Table 2: Number of UK Followers by News Outlet

News Outlet	Number of Followers (count)	Outlet Reach (%)
BBC News	1,268,795	44.8
Sky News	840,779	29.7
Guardian	837,723	29.6
New York Times	735,638	26.0
The Economist	730,644	25.8
CNN	566,126	20.0
Channel 4 News	429,354	15.2
Telegraph	329,459	11.6
Financial Times	275,523	9.7
Mail Online	202,778	7.2
Times	193,313	6.8
Independent	81,148	2.9
Total	2,832,132	100.0

Notes: Table reports the number of UK followers of the listed news outlets in December 2021 and the outlets' reach out of the total number of followers of these outlets. Source: Ofcom analysis of Twitter data.

- 4.7 The figures for outlet reach present a picture of the relative size of outlets that is broadly consistent with known statistics of news consumption in the UK. Similar to Ofcom's "News Consumption Report: 2022" and Reuters "Digital News Report 2022", for instance, BBC News, Sky News, and the Guardian are amongst the top outlets; CNN and the Telegraph are in the middle; and the Times and the Independent are at the bottom. There are, however, also some notable differences: Mail Online ranks lower in terms of UK Twitter followership reach than in terms of consumption in the other two reports and Channel 4 News ranks higher.
- 4.8 In addition to the reach of any one news outlet, we also compute the overlap in followership between different news outlets to analyse multi-sourcing. We do this in two ways. First, Table 3 below shows, for followers of any given outlet, the number of outlets followed. Second, Figure 2 below illustrates, for followers of any given outlet (reported in the row), the share of these followers that also follow another specific outlet (named in the column).

³⁸ In particular, if reporting of the location is unrelated to news outlet followership, then the ranking of a given outlet among these 12 outlets will be unaffected by missing location data.

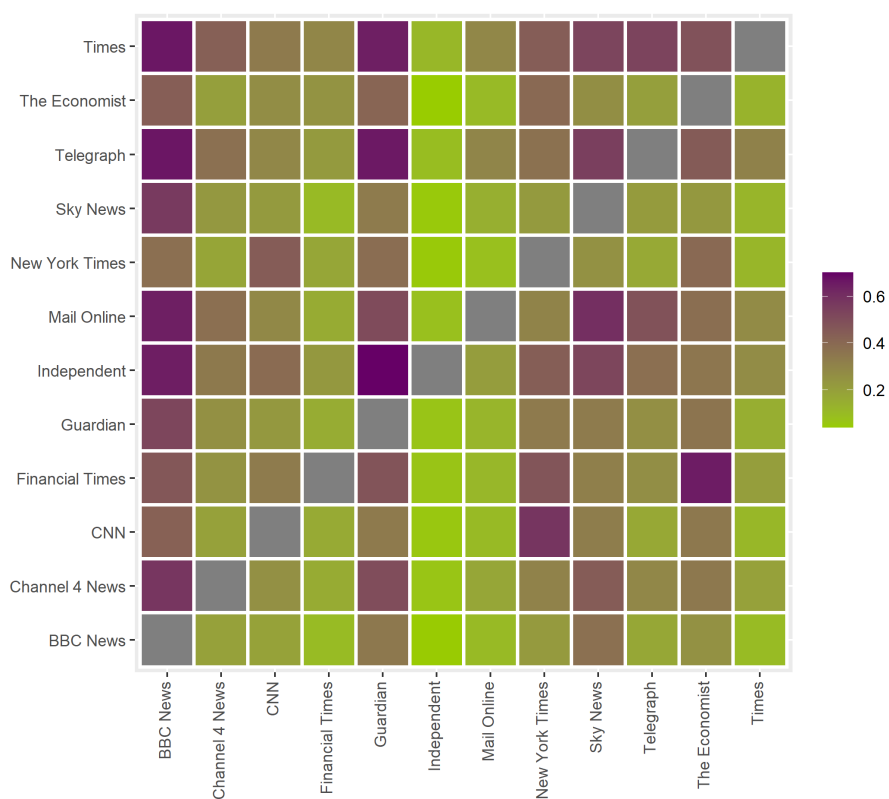
Table 3:3: Share of Followers by Number of Outlets Followed

News Outlet	Number of Outlets Followed											
	1 (%)	2 (%)	3 (%)	4 (%)	5 (%)	6 (%)	7 (%)	8 (%)	9 (%)	10 (%)	11 (%)	12 (%)
BBC News	30.7	23.1	14.1	9.7	6.9	5.2	3.6	2.8	1.8	1.3	0.8	0.2
Sky News	27.6	21.9	13.0	9.7	7.3	6.1	4.6	4.0	2.5	1.9	1.1	0.2
Guardian	19.7	18.7	16.2	12.6	9.6	7.6	5.3	4.3	2.6	1.9	1.1	0.2
New York Times	19.8	23.0	16.4	11.9	8.5	6.0	4.5	3.5	2.7	2.1	1.3	0.3
The Economist	20.3	20.1	16.8	12.2	8.6	6.8	4.8	4.1	2.6	2.1	1.3	0.3
CNN	18.7	21.9	16.1	11.8	8.6	6.3	4.9	3.9	3.2	2.7	1.7	0.4
Channel 4 News	17.8	14.9	14.6	12.0	9.6	8.1	6.7	6.3	4.1	3.4	2.1	0.5
Telegraph	8.2	10.5	11.9	12.1	12.0	12.0	9.7	9.2	6.1	4.8	2.9	0.6
Financial Times	13.2	16.3	13.6	12.9	10.4	8.2	6.7	5.5	4.8	4.3	3.2	0.7
Mail Online	12.7	12.1	11.9	11.1	9.4	10.2	8.1	8.8	5.6	5.0	4.3	1.0
Times	7.9	9.7	10.3	10.1	10.0	9.7	9.5	11.5	8.1	7.3	4.8	1.0
Independent	6.3	10.4	12.2	12.8	12.3	11.0	9.5	8.1	6.4	4.9	3.5	2.5
All Outlets	48.4	22.4	11.2	6.5	4.0	2.7	1.7	1.3	0.8	0.6	0.3	0.1

Notes: Table shows the shares of followers of the news outlet given in rows that follow the number of news outlets given in the columns. The corresponding shares of followers of all listed news outlets is given in the bottom. Source: Ofcom analysis of Twitter data.

4.9 Table 3 shows that just over 50% of UK followers of the 12 news outlets follow multiple outlets – i.e., 48.4% follow only one outlet and 51.6% at least two outlets. It also shows, however, that there is considerable variation between the followers of different news outlets. Roughly 31% of the followers of BBC News and about 28% of those of Sky News, for instance, follow only those outlets, whereas this figure is below 10% for followers of the Telegraph, the Times and the Independent.

Figure 2: Share of Followers of One Outlet (Row) Following Another (Column)



Notes: Figure shows share of followers of the news outlets given in rows that also follow the news outlets given in the columns. The share is indicated for each tile by its shading along the colour spectrum given in the legend, purple indicating a high share and green a low share. Source: Ofcom analysis of Twitter data.

4.10 Figure 2 illustrates the overlap of followership between any two outlets. The colour of the tile at the intersection of a given row and column indicates the row-outlet’s share of followers that also follow the column-outlet. Tiles that are shaded towards the purple end of the spectrum have a higher follower overlap and those shaded towards the green end of the spectrum have a lower overlap.

4.11 Figure 2 shows that there is considerable heterogeneity in terms of the follower-overlap between any pair of news outlets. Some of this reflects the overall variation in followership between (column) outlets. For example, a larger share of the followers of most outlets also follow BBC News as compared to the Independent. This can be seen from their respective columns in mostly purple and green. There is, however, also variation between the followers of different (row) outlets in their overlap with a given other (column) outlet. Followers of the Times, the Telegraph, or the Independent, for instance, are more likely to also follow the Guardian than the followers of BBC News or Sky News. This can be seen by the mostly purple tiles for the Times, the Telegraph, and the Independent in the Guardian column and the more mixed tiles for BBC News and Sky News in that column.

Liking of news outlet content

- 4.12 Information on the liking of content by news outlets allows us to compute metrics that broadly relate to two categories of Ofcom’s media plurality measurement framework. First, liking represents a proxy for consumption, because any tweet that has been liked by a given user was necessarily also viewed. Similar to followership, however, this is an imperfect proxy for consumption as not all viewed tweets are liked (see further discussion in Section 3 above). Secondly, liking often represents a degree of agreement with a given tweet, so that the number of likes received by a news outlet provide a measure of the outlet’s personal importance to the consumer, which relates more closely to the impact category of the measurement framework.³⁹
- 4.13 Importantly, while liking represents the most common form of user engagement with news tweets, it is still very rare in our dataset.⁴⁰ Indeed, of the 2.8 million UK followers of the 12 news outlets, less than 1% liked any of the outlets’ tweets from December 2021. For this reason, we do not study liking among the followers of the news outlets, but rather focus our analysis on the broader Twitter users that liked at least one tweet by the 12 news outlets during that month (henceforth “liking users”). Importantly, this broader set of Twitter users is not restricted to UK users as the location information has not been retrieved for all liking users.
- 4.14 Even among this set of Twitter users, liking of content by the 12 news outlets in December 2021 remains rare. On average, these users liked only 3 of the news outlets’ tweets and over 75% of the users liked 2 or fewer tweets. Furthermore, the likes also tend to accrue to few outlets, with about 84% of the users having liked tweets by only 1 outlet and about 95% of the users having liked tweets by up to 2 outlets. This suggests that metrics based on liking are unlikely to provide a robust basis to measure media plurality, at least for a relatively short period such as the single month investigated in this paper.
- 4.15 We present metrics based on the likes of news tweets in December 2021 in this subsection, but do so as a means of exploring an approach to measuring news consumption and elements of media plurality; these metrics should not be interpreted as strong evidence on media plurality itself. We do not include metrics that relate to the overlap of likes between outlets as these appear even more rarely during the single month.
- 4.16 Table 4 below shows liking metrics that are similar to those presented in the previous subsection for followership.⁴¹ The first two columns report the number of users who liked at least one tweet from the outlet (i.e., “liking users”) and the number of liking users for the outlet divided by the total number of liking users for all outlets in the analysis (i.e., “outlet reach”) for December 2021. The last two columns report the number of likes by outlet and the share of total likes for the same period.

³⁹ We recognise that liking may be an imperfect approximation for users agreeing with Tweets. For example, users may like Tweets in order to bookmark them.

⁴⁰ The other forms of user engagement with tweets are retweeting, quoting, and replying.

⁴¹ Unlike the followers in the previous subsection, the likes included in the statistics of Table 4 are not restricted to those by users from the UK. This is the case because we did not have the location information for all liking users.

Table 4: Number of Liking Users and Number of Likes (December 2021)

News outlet	Liking Users		Likes	
	Number (count)	Reach (%)	Number (count)	Reach (%)
CNN	89,280	18.7	185,409	13.5
Guardian	85,646	18.0	201,404	14.7
Sky News	76,147	16.0	188,001	13.7
New York Times	65,453	13.7	120,729	8.8
BBC News	62,808	13.2	173,466	12.7
The Economist	56,442	11.8	126,963	9.3
Independent	54,397	11.4	146,368	10.7
Mail Online	34,451	7.2	81,482	5.9
Telegraph	29,623	6.2	62,273	4.5
Times	23,603	5.0	42,419	3.1
Financial Times	11,597	2.4	24,287	1.8
Channel 4 News	10,453	2.2	16,715	1.2
Total	476,663	100.0	1,369,516	100.0

Notes: Table shows (i) the number and share of liking users of content by the listed news outlets and (ii) the number and share of likes of content by the listed news outlets in December 2021. Source: Ofcom analysis of Twitter data.

- 4.17 One interesting insight from a measurement perspective from Table 4 comes from the comparison of the number of liking users and the number of likes. In principle, these present potentially different measures of the relative sizes of news outlets because some outlets may have more likes per user than others. Comparing the reach according to liking users and likes in Table 4 reveals that there are some differences between these measures, but that they are broadly consistent in relative terms.⁴² Two outlets that stand out as ranking higher according to liking users than likes are the New York Times and The Economist.⁴³ Barring these exceptions, there is little difference in the relative ranking of news outlets between these two measures.
- 4.18 In addition, despite the broader set of users considered in the context of liking versus followership, it is useful to compare the relative ranking of news outlets between the two, especially for the UK outlets. Relative to the followership Table 2, the Guardian and Sky News appear in a similar position at the top in Table 4 while BBC News appears somewhat lower in the ranking. The Economist is in a similar position in the middle. Mail Online and the Independent are ranked higher according to the liking metrics; the Financial Times and the Independent are ranked lower. These differences between Table 4 and Table 2 suggest

⁴² The figures for outlet reach according to likes notably add up to 100% while those according to liking users do not. This is the case, because a given user can like the content of multiple outlets but a given like can only apply to one outlet. The outlet reach according to likes therefore tends to be smaller than outlet reach according to liking users and comparison of these measures for a given outlet is not particularly meaningful.

⁴³ These have their likes spread over a larger number of distinct users relative to other outlets.

that it is important to consider both followership and liking as proxies for news consumption on Twitter, especially in the absence of more direct measures.

- 4.19 Table 5 below reports several additional metrics based on the likes for December 2021 that relate to the engagement of Twitter users in response to the tweets by news outlets. The first column shows the number of tweets, the second and third columns the number of likes and distinct liking users (replicated from Table 4 above), the fourth column shows the average number of likes per tweet, and the fifth column shows the average number of liking users per tweet.

Table 5: Tweets, Likes, and Liking Users (December 2021)

News outlet	Number of			Average Number of	
	Tweets (count)	Likes (Count)	Liking Users (count)	Likes per Tweet (Ratio)	Liking Users per Tweet (Ratio)
Independent	11,309	146,368	54,397	12.9	4.8
Guardian	5,694	201,404	85,646	35.4	15.0
Mail Online	5,445	81,482	34,451	15.0	6.3
Sky News	4,223	188,001	76,147	44.5	18.0
BBC News	4,738	173,466	62,808	36.6	13.3
Telegraph	4,154	62,273	29,623	15.0	7.1
The Economist	3,329	126,963	56,442	38.1	17.0
New York Times	1,721	120,729	65,453	70.2	38.0
Times	2,444	42,419	23,603	17.4	9.7
CNN	2,244	185,409	89,280	82.6	39.8
Financial Times	859	24,287	11,597	28.3	13.5
Channel 4 News	406	16,715	10,453	41.2	25.7
Total	46,566	1,369,516	476,663	29.4	10.2

Notes: Table shows the number of tweets by the listed outlets in December 2021, the number of likes of those tweets, and the number of the corresponding liking users; and the average number of likes and liking users per tweet. **Source:** Ofcom analysis of Twitter data.

- 4.20 The metrics in Table 5 indicate a wide range of engagement across the different outlets and provide insights whether this level of engagement is achieved with low or high levels of tweeting activity. The American outlets – i.e., CNN and the New York Times – stand out as achieving many likes with relatively few tweets, followed by UK outlets that were high-ranking among followership – i.e., BBC News, Sky News, and the Guardian.⁴⁴
- 4.21 One of the key advantages of liking as a proxy for consumption is that, in principle, it captures additional ways of accessing news that are not captured by following (e.g., searching for trending topics or receiving content from other users). A comparison of liking

⁴⁴ CNN and the New York Times achieving many likes with relatively few tweets is also understood to be related to their large global followership relative to the other outlets.

by followers of news outlets to liking by non-followers may therefore shed some light onto the relevance of these other ways of accessing news.

- 4.22 Table 6 below reports precisely this split of the liking users from December 2021. The first column shows the number of liking users that are also followers of the given outlet, the second column shows those that are non-followers, and the third column shows the follower share out of all liking users. The statistics show that a considerable share of liking users are non-followers, ranging from roughly 30% for CNN and the New York Times to over 70% for the Independent and the Times. Across all outlets, only about 58% of the liking users of tweets by the 12 news outlets in December 2021 follow any of these outlets. This confirms that there are important routes to accessing news content on Twitter other than through following outlets.

Table 6: Liking Users by Followership (

News Outlet	Number of		
	Followers (count)	Non Followers (Count)	Follower Shares (%)
New York Times	46,137	19,316	70.5
CNN	62,224	27,056	69.7
The Economist	39,084	17,358	69.2
Guardian	53,933	31,713	63.0
BBC News	37,021	25,787	58.9
Financial Times	6,645	4,952	57.3
Sky News	41,197	34,950	54.1
Channel 4 News	5,479	4,974	52.4
Mail Online	13,442	21,009	39.0
Telegraph	11,037	18,586	37.3
Independent	15,405	38,992	28.3
Times	6,395	17,208	27.1
All Outlets	277,940	198,723	58.3

Notes: Table shows the number of users liking tweets by the listed outlets in December 2021), separately for followers and non-followers of those outlets, and the share of these users accounted for by followers. **Source:** Ofcom analysis of Twitter data.

5. Homophily as an echo-chamber metric

- 5.1 An echo chamber, broadly defined, is as a place where the ideas seen by a user reflect the ideas that they already hold. Ofcom’s Media Plurality Discussion Document, published alongside this paper, finds evidence that social media users may be at risk of experiencing some echo chamber effects. It distinguishes the following three underlying reasons for the emergence of echo chambers that relate to potential “echo-chamber metrics”.
- i) Filter bubbles: algorithms may filter the news users receive based on their previous online behaviour or the behaviour of people like them in order to maintain their level of engagement.
 - ii) Homophily: a tendency for people to associate more often with people who are similar.
 - iii) High segregation: in a highly segregated news environment people are less likely to see news with the same political viewpoint (i.e., there will be a greater difference in the political slant of news seen by two randomly chosen individuals).
- 5.2 In this paper, we focus on one of these: homophily in news consumption on Twitter. We chose this focus, because there is some evidence of homophily between connected users on social media from the academic literature. With respect to Twitter, for instance, Halberstam and Knight (2016) find that politically engaged users disproportionately follow users with similar political ideology and are disproportionately exposed to like-minded information.⁴⁵ Similarly, Cinelli et al. (2021) find that users that engage in discussions on selected controversial topics on Twitter (e.g., abortion) tend to follow others who exhibit a similar political leaning.⁴⁶ This suggests that homophily could also be relevant in the context of news consumption on Twitter. The other two potential reasons for echo chambers – i.e., filter bubbles and high segregation – are important and metrics for them could be investigated in the future.⁴⁷
- 5.3 Homophily is an important indicator for the potential emergence of echo chambers in news consumption on Twitter for two separate reasons.
- a) First, it captures the notion that the network of followed users (i.e., friends) may provide indirect exposure to news content by retweeting or quoting. This indirect exposure may be similar to a given user’s own exposure to news content (exhibiting homophily) or dissimilar (exhibiting heterophily).

⁴⁵ See Halberstam and Knight (2016), “Homophily, group size, and the diffusion of political information in social networks: Evidence from Twitter”, *Journal of Public Economics* 143: 73-88.

⁴⁶ See Cinelli et al. (2021), “The echo chamber effect on social media”, *Proceedings of the National Academy of Sciences* 118(9): 1-8.

⁴⁷ The Twitter API, in principle, offers data to analyse filter bubbles and segregation. An analysis of filter bubbles, for instance, would require data on the interaction of users with tweets over time. An analysis of segregation is feasible even with followership data as retrieved for the analysis presented in this paper; indeed, the metrics on multi-sourcing in Section 4 are closely related to segregation.

- b) Secondly, homophily captures the notion that algorithms recommend content partly based on tweets viewed within a user's network. If Twitter users and their friends are broadly similar, the algorithm may therefore recommend relatively uniform content; within a more diverse network, the recommended content may also be more varied.
- 5.4 Therefore, homophily in news consumption could in principle be influenced either by the way in which news content is algorithmically curated on Twitter, or by the way in which Twitter facilitates the creation of social networks of users who interact with and share news content, or by a combination of these factors. In our analysis, we do not distinguish between these potential mechanisms that relate homophily to the presence of an echo chamber.⁴⁸ We are simply presenting metrics that can document homophily, which may arise through several alternative mechanisms.
- 5.5 We measure homophily as the similarity between connected Twitter users with respect to the news outlets they follow or whose content they like relative to the corresponding similarity with the general population of followers/liking users.⁴⁹ Our measure of homophily ranges between -1 and 1, with the lower bound representing complete dissimilarity and the upper bound representing complete similarity.
- 5.6 The homophily metric we explore in this paper has the following two key advantages over some standard measures from the academic literature.
- a) First, it does not require a classification of news outlets along a single dimension to make them comparable (e.g., left and right of the political spectrum). Classifications like this are frequently implemented in the academic literature, e.g., research by Cinelli et al. (2021) on echo chambers on social media in terms of users' leaning (i.e., pro/cons) on certain contentious topics.⁵⁰ This is not feasible in our context without relying on an a-priori classification of the news outlets (e.g., along the political left-right spectrum).
- b) Second, our homophily metric allows us to capture multi-sourcing, which is an important feature of news consumption in the UK and integral to understanding plurality. It measures the similarity between the followership/liking profiles of Twitter users by considering the overlap for each news outlet in turn. As such, it does not require an aggregation into two mutually exclusive groups as in the canonical homophily index (Currarini et al., 2009).⁵¹

⁴⁸ Indeed, an echo chamber may also lead to homophily, e.g., homophily in terms of liking when the similar content is presented to users based on a filter bubble according to past online behaviour.

⁴⁹ See Appendix A1 for the mathematical definition.

⁵⁰ See supra note 46.

⁵¹ Currarini et al. (2009) define homophily precisely as an individual's share of ties with others of the same type out of all ties. This definition requires splitting the population of individuals into two mutually exclusive and exhaustive sets – e.g., followers and non-followers of the BBC News – but this is not particularly meaningful for the context of news consumption. See supra note 17 for the full Currarini et al. (2009) reference.

- 5.7 Importantly, since liking of news content is very rare for a given user, we stress that homophily based on liking is potentially not very robust and place more emphasis on homophily based on followership.⁵²

Echo chambers and diversity

- 5.8 While social media may foster echo chambers, they also have the potential to broaden users' exposure to diverse news sources. Indeed, as also mentioned in Ofcom's Media Plurality Discussion Document, there is evidence that the latter holds across online intermediates, including social media platforms. It is therefore important not to consider echo-chamber metrics in isolation, but in conjunction with measures of diversity.
- 5.9 In this paper, we consider our homophily metrics jointly with simple measures of diversity: the number of news outlets followed or the number of outlets whose content a consumer liked. The match of the underlying measure between the homophily and diversity metrics – i.e., following or liking – is useful, because it allows for a more fundamental link between homophily and diversity.⁵³ Homophily combined with a diverse news outlet followership/liking profile means that a given user and their friends are exposed to a relatively wide range of news sources. Homophily combined with following fewer or linking behaviour focused on fewer news outlets, by contrast, is evidence suggestive of an echo chamber that involves access to few news sources. The latter scenario is more of a concern from a media plurality standpoint.
- 5.10 Finally, it is important to remember that the homophily and diversity metrics presented in this paper are affected by the limitations discussed in Section 1. In particular, the focus on only 12 prominent news outlets impacts the weights assigned to each news outlet in the homophily metric and also the number of news outlets followed/liked.⁵⁴ Therefore, the homophily and diversity metrics below should be understood as a part of the exploratory analysis of measuring news consumption on Twitter rather than firm evidence of echo chambers and diversity of news sources on the platform.

Followership of news outlets

- 5.11 Figure 3 below illustrates the distribution of our homophily measure based on following news outlets in the form of box-and-whisker plots. The left plot is for the similarity with connected Twitter accounts (i.e., with a user's friends); the middle plot is for the similarity

⁵² In addition, given the rare instances of liking, we define liking at the outlet, not at the tweet level. This may further lead to unrelated likes being associated through the outlet, even though the actual content may be very different. This is particularly important for outlets that cover a broad range of topics and viewpoints, such as the BBC.

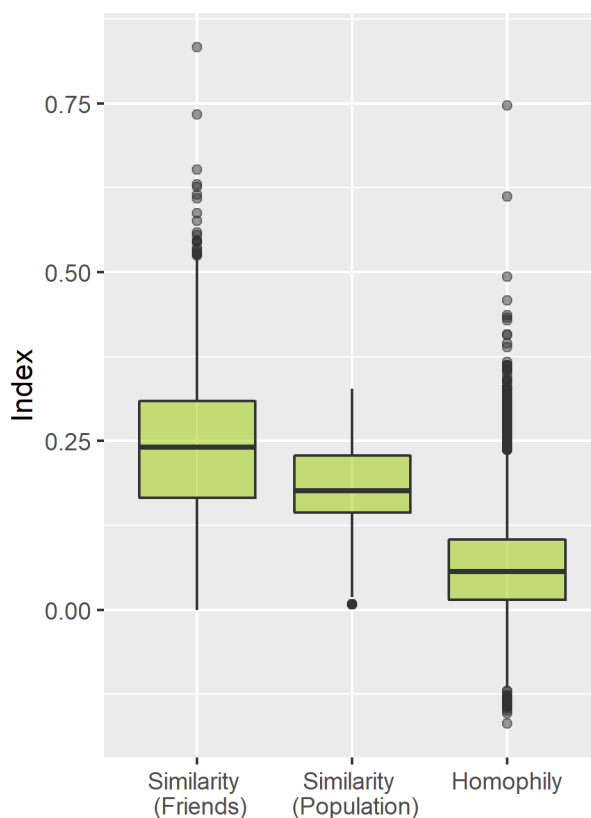
⁵³ This link between the homophily and diversity metrics with a uniform underlying measure (i.e., following or liking) is closer than without it. In the latter case, users on a social media platform may be in an echo chamber that involves exposure to relatively narrow set of news sources through their network (e.g., as measured by followership of news outlets) while being exposed to a relatively diverse set of news sources themselves (e.g., as measured by liking of news content). The different measures underlying the homophily and diversity metrics, of course, remain important, but their alignment facilitates the interpretation of whether homophily is potential cause for concern from a media plurality standpoint.

⁵⁴ See Appendix A1 for the mathematical definition of the homophily metric.

with the population of news-outlet followers as a whole; and the right plot is for the difference between the two, i.e., relative similarity as our measure of homophily.⁵⁵

- 5.12 The bottom and the top of the box represent the 25th and 75th percentile of the distribution, respectively, and the solid horizontal line inside the box represents the median. The ends of the “whiskers” – i.e., the vertical lines – show the upper and lower end of 1.5 times the range between the 25th and 75th percentile. The individual points outside this range represent extreme values (i.e., users with very high or low levels of similarity).

Figure 3: Homophily in Following



Notes: Figure illustrates the distribution of the similarity in news outlet following between a given user and her friends (left), the similarity in news outlet following between a given user and the population (middle), and homophily as measured by the difference between these two (right). Precise definitions are available in Appendix A1. **Source:** Ofcom analysis of Twitter data.

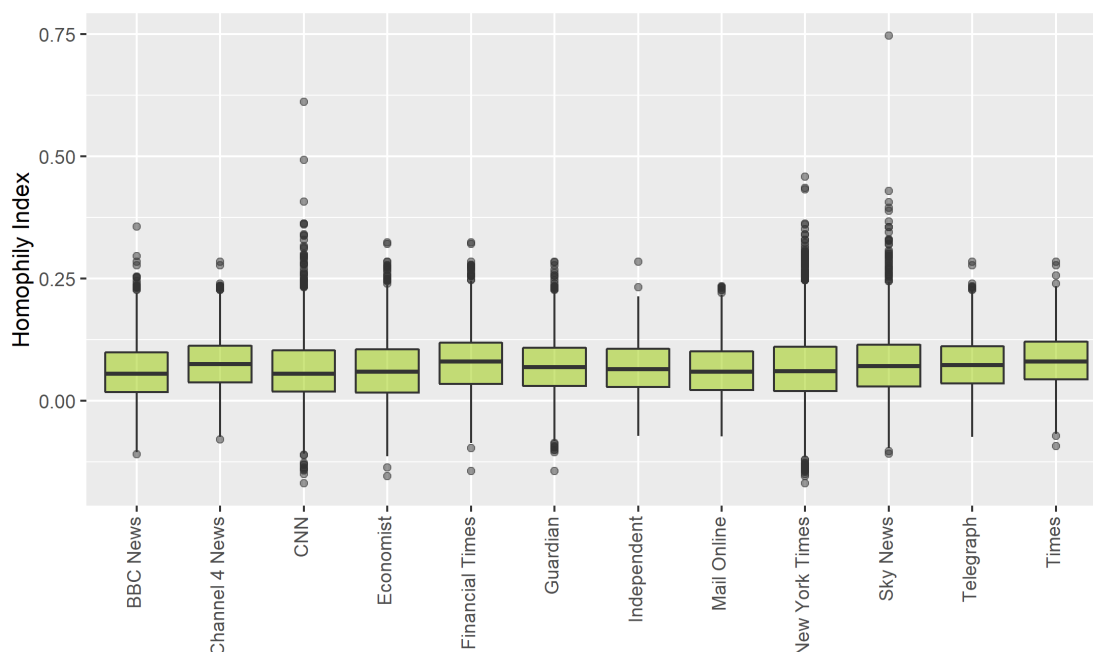
- 5.13 Figure 3 shows that most followers of news outlets on Twitter exhibit relatively low levels of homophily according to this measure. The median homophily is roughly 0.06 and the 25th and 75th percentile of the distribution are approximately 0.02 and 0.12, respectively. This means that a given Twitter users’ friends are somewhat more likely than the

⁵⁵ See Appendix A1 for the precise definition of “similarity” and how this relates to the homophily metric.

population of news-outlet followers as a whole to follow the same news outlets as that user.

- 5.14 There are, however, some users with more extreme values of homophily, with a higher likelihood of friends' following the same outlets as the population (more than 0.25). Users' exhibiting high homophily values may be a concern from a plurality standpoint if they also tend to access few news outlets. The wide range of homophily across Twitter users suggests that it is important to consider the distribution as we have here rather than simply focusing on the average level of homophily.
- 5.15 The distribution of homophily could, in principle, differ between followers of different news outlets and this is a hypothesis we explore in Figure 4. The figure shows box-and-whisker plots separately for followers of a given news outlet.⁵⁶ Overall there is relatively little variation in the distribution between followers of different outlets. There are slightly more users with extreme values of homophily among followers of the CNN, the New York Times, and Sky News than for those of the remaining outlets. Beyond that, the distributions do not appear to exhibit any other meaningful differences.

Figure 4: Homophily in Following by News Outlet



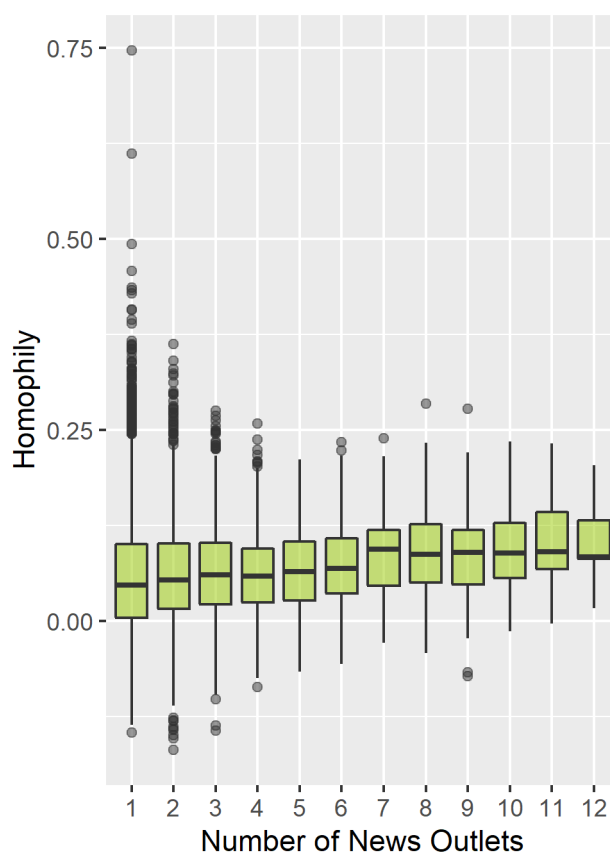
Notes: Figure illustrates the distribution of homophily as measured by the difference between the similarity in news outlet following of a given user with her friends and the population, separately for followers of a given news outlet. The precise definition is available in Appendix A1. **Source:** Ofcom analysis of Twitter data.

- 5.16 As discussed above, the homophily metric should be considered in conjunction with information on the diversity of the underlying news sources for a meaningful

⁵⁶ These plots are for the homophily measure only and therefore correspond to the right plot in Figure 3.

interpretation. In Figure 5, we therefore present box-and-whisker plots of homophily based on following news outlets separately by the number of outlets followed by a given user. These plots show that users with relatively high levels of homophily (i.e., those indicated by individual points) tend to follow relatively few news outlets. Interestingly, users in the central part of the homophily distribution (i.e., those represented by the box) seem to exhibit a slight positive relationship between homophily and the number of outlets followed. This suggests that Twitter users that follow a higher number of the 12 news outlets also tend to have friends who follow a higher number of outlets.⁵⁷

Figure 5: Homophily in Following by Number of News Outlets Followed



Notes: Figure illustrates the distribution of homophily as measured by the difference between the similarity in news outlet following of a given user with her friends and the population, separately for user groups according to the number of news outlets followed. The precise definition is available in Appendix A1. **Source:** Ofcom analysis of Twitter data.

5.17 Overall, the analysis in this sub-section provides three insights about homophily as an indicator for echo chambers on Twitter:

⁵⁷ Although a higher number of outlets followed raises the chance of overlap with another Twitter user, this is equally true for the user's friends and followers of the news outlets more generally. Given that the homophily metric is a relative measure of a user's similarity with these two groups, this observed pattern does not reflect this higher chance of overlap.

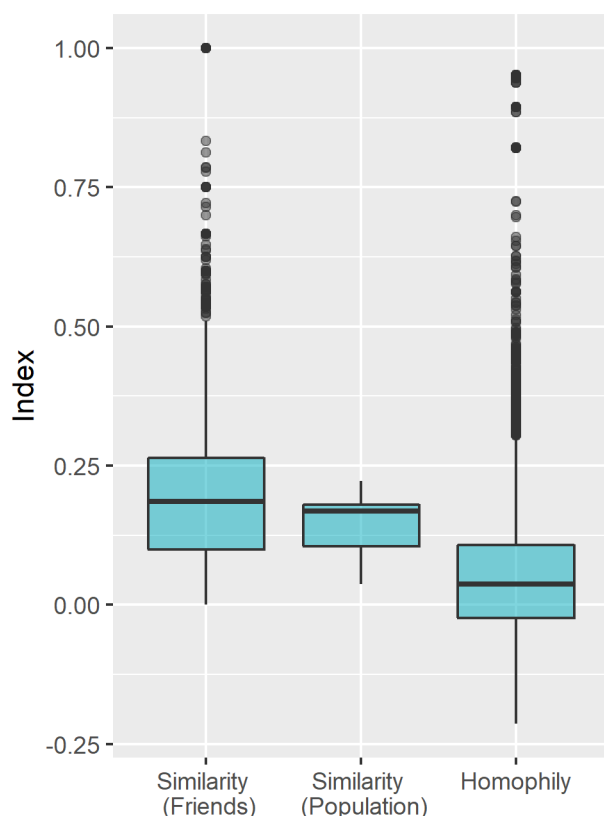
- a) First, it captures variation in the similarity between connected Twitter users in terms of their news-outlet following. Most Twitter users have a slightly greater similarity in news outlet followership with other users that they follow than with the general population of news-outlet followers. There are, however, some users whose followership profile is much more similar to that of their friends than to the general population of news-outlet followers.
- b) Second, we can view the distribution of homophily for different user groups to assess whether different levels of homophily are specific to certain groups. We have demonstrated this by defining user groups according to whether they follow a given news outlet. While we do not observe meaningful variation in the homophily distribution by news-outlet followership, there may well be other user groups that reveal meaningful differences.
- c) Third, we have presented information jointly for the distribution of homophily and a measure of diversity, namely the number of news outlets a given user follows. This revealed that users with relatively high levels of homophily tend to follow relatively few outlets. It also showed that the average level of homophily in following the 12 outlets increases slightly with the number of outlets followed.

Liking of news outlet content

- 5.18 In addition to measuring homophily based on following news outlets, we also compute similar metrics based on liking of news outlet content during December 2021. As discussed in detail in Section 4, liking potentially captures aspects of news consumption that extend beyond those captured by followership.
- 5.19 As also mentioned above, however, liking of content by the 12 news outlets in this analysis is quite rare and overlap in liking between users even more so. Therefore, it should be emphasised that the homophily metrics computed on the basis of liking are less reliable than those based on followership and we are cautious to draw firm implications from them.⁵⁸
- 5.20 Figure 6 and Figure 7 present box-and-whisker plots of the distribution of homophily based on liking of news outlet content similar to those based on followership above. Figure 6 shows the distribution across all outlets (including the similarity with friends' and the population's liking of news outlet content) and Figure 7 shows separate distributions for consumers that liked a given outlet's content. Homophily based on liking appears to be lower for the central 50% of the distribution than homophily based on followership. At least 25% of liking users exhibit heterophily (i.e., negative homophily) across most outlets. At the same time, there also remain a notable number of consumers with relatively high levels of homophily based on liking.

⁵⁸ Future research could consider a longer period to measure liking of news outlet content and could also complement information on liking with additional measures of consumption and/or user engagement.

Figure 6: Homophily in Liking

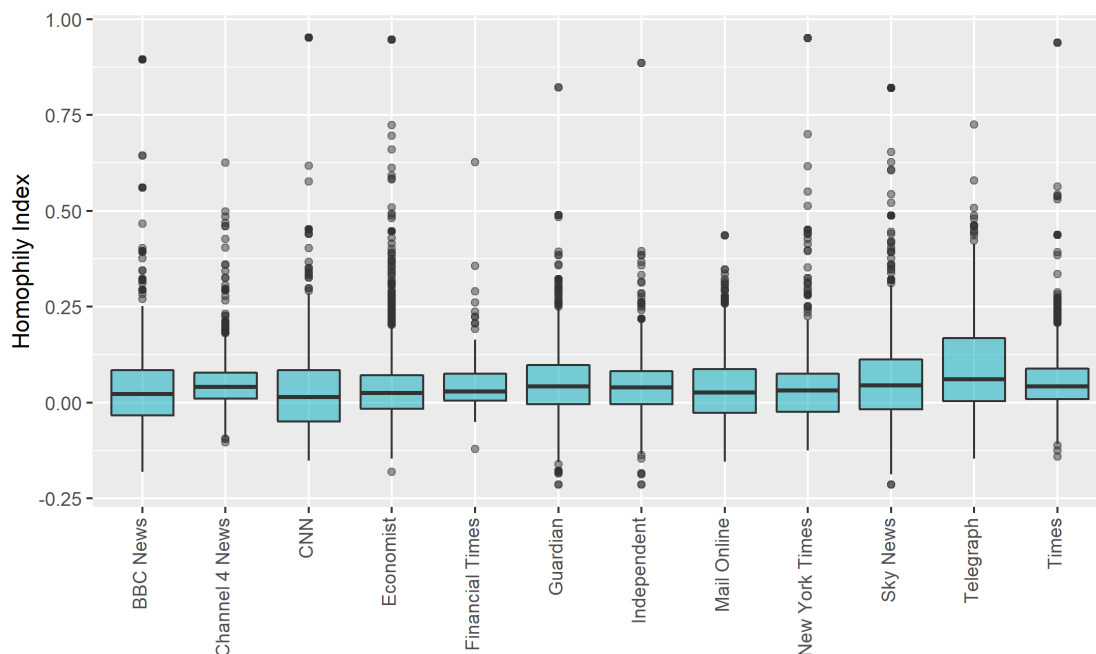


Notes: Figure illustrates the distribution of the similarity in liking of news outlet content between a given user and her friends (left), the similarity in liking of news outlet content between a given user and the population (middle), and homophily as measured by the difference between these two (right). Precise definitions are available in Appendix A1. **Source:** Ofcom analysis of Twitter data.

5.21 The difference in patterns of homophily defined by likes relative to that defined by followership is likely to be related to the rare incidence of liking. Infrequent likes reduce the number of overlapping outlets between users on average but result in high values of homophily when there is disproportionate overlap in liking between friends.⁵⁹ This suggests that the central part of the distribution of users (i.e., the box) is closer to zero and that extreme values of homophily (high or low) are further spread out (for users that have liking similar to their friends or similar to most others in the population, respectively).

⁵⁹ The high values of homophily with overlap are due to the normalization embedded in the homophily metric. With few likes overall, any overlapping news outlet receives a relatively high weight. Essentially, for most pairs of users there is either full overlap (of one news outlet) or none.

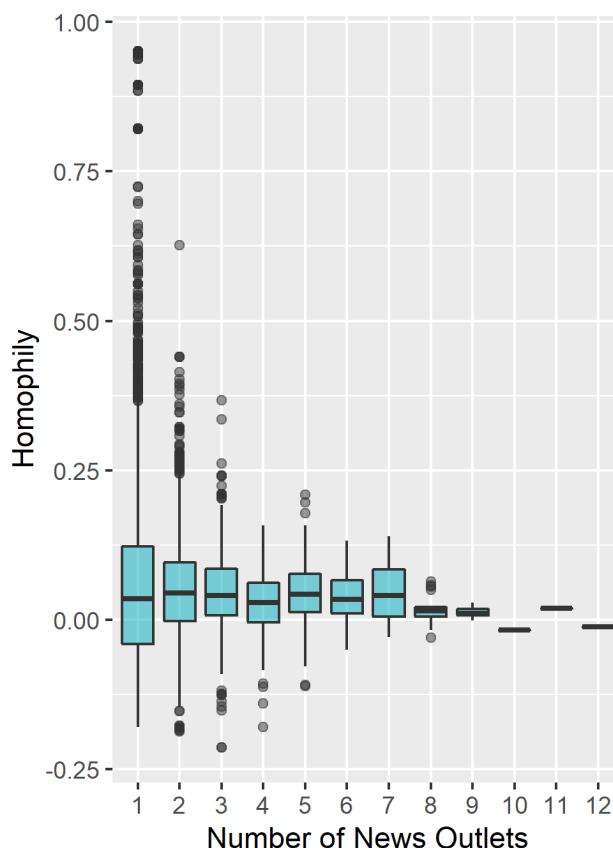
Figure 7: Homophily in Liking by News Outlet



Notes: Figure illustrates the distribution of homophily as measured by the difference between the similarity in liking of news outlet content of a given user with her friends and the population, separately for users that liked content of a given news outlet. The precise definition is available in Appendix A1. **Source:** Ofcom analysis of Twitter data.

Furthermore, as potential evidence regarding the presence of echo chambers, we also consider homophily based on liking jointly with the number of outlets whose content a given consumer liked. We do this again in the form of box-and-whisker plots as shown in Figure 8.

Figure 8: Homophily in Liking by Number of News Outlets Liked



Notes: Figure illustrates the distribution of homophily as measured by the difference between the similarity in liking of news outlet content of a given user with her friends and the population, separately for user groups according to the number of news outlets whose content was liked. The precise definition is available in Appendix A1. **Source:** Ofcom analysis of Twitter data.

- 5.22 Similarly, to the metrics based on followership, these plots show that users with relatively high levels of homophily (i.e., those indicated by individual points) tend to follow relatively few news outlets. Unlike the followership metrics, however, homophily based on liking and the number of liked news outlets do not seem to exhibit a positive relationship for users in the central part of the homophily distribution (i.e., those represented by the box).⁶⁰
- 5.23 Overall, the analysis in this sub-section provides one additional key insight about the use of homophily as an indicator for the presence of echo chambers on Twitter: the frequency of the measure underlying the homophily (i.e., likes in this case) seems to have implications for the both the central and extreme parts of the distribution that need to be further investigated.

⁶⁰ This may also be due to the rare incidence of liking relative to following multiple news outlets.

6. Conclusion

- 6.1 Motivated by the increasing importance of social media for news consumption in the UK, this discussion paper presents an exploratory analysis to understand how we can measure news consumption and elements of media plurality on social media using the example of Twitter. Following Ofcom's existing media plurality measurement framework, we put forward metrics that can be computed based on publicly available data from the Twitter API, but we also highlight important data limitations.
- 6.2 We compute descriptive metrics covering the three quantitative categories of the measurement framework: availability, consumption, and impact. We measure availability via the activity of news outlets on Twitter in terms of number of tweets. We measure impact via consumers' engagement in terms of liking content by news outlets.
- 6.3 Key aspects of news consumption on Twitter are unobservable based on public data, so we use followership and liking as proxies for consumption. Consumers frequently view news content through channels that are not captured by the Twitter API. In addition, the content presented to and viewed by consumers is typically curated by algorithms and affected by other consumers in the social network. As a result, following news outlets is likely a poor proxy for the consumption of their content. Information on the likes of news tweets circumvents some of these shortcomings but applies only to a subset of viewed tweets and occurs too rarely to provide a robust proxy for consumption.
- 6.4 In addition to the descriptive media plurality metrics, we explore metrics that could shed some light on the presence of echo chambers among communities of news consumers on Twitter. Overall, we find limited such evidence for most users. Most users' overlap in news outlet followership is slightly greater with other users they follow than with the population of news outlet followers. Most users' overlap in liking of content by news outlets is roughly similar with other users they follow as with the population of liking users. Some users, however, exhibit much greater overlap in followership and liking of news outlets' content with other users they follow than with the population; and these users tend to follow/like content of only few news outlets. This could be indicative of echo chambers among certain subgroups of Twitter users and may warrant further investigation.
- 6.5 This study is exploratory in nature and is not designed to provide an exhaustive picture of measuring news consumption and media plurality on social media. We think there is scope for further research in this area, covering both Twitter and other social media platforms more generally. In relation to work on Twitter, we will consider two options for further work:
 - a) Firstly, we could carry out further analysis using the content of tweets, quotes, and replies. This could shed light on topics like the impact of news on users and the extent to which their sentiment points to potential issues, such as polarisation and misinformation.
 - b) Second, information that is not available on the Twitter API, but is available to Twitter, could be used to produce a more accurate picture of how news is consumed online. For

example, while we have data on likes, retweets, quotes, and replies, which we can use to study user engagement with news, those actions are very rare in our dataset.

Therefore, data on whether a user has read a tweet would give a more accurate picture of news consumption.⁶¹

6.6 As discussed above, we have been able to carry out this analysis on Twitter because of the data made available via the platform's API. Studying media plurality across a broader range of platforms, like Facebook or TikTok, is key to obtaining a more holistic picture of online media plurality, particularly since platforms may differ considerably in terms of their own choice architecture and algorithms and in their users' behavior and preferences. Therefore, findings on one platform may not be generalizable. It would therefore be of interest to repeat and adapt the approaches explored in this paper to other platforms to obtain a more complete picture of the state of media plurality in online news consumption.

6.7 We will also consider broader analysis beyond the approaches set out in this paper:

- a) First, experimental methods could be used to shed light on some of the issues discussed in this report.⁶² Understanding the impact of algorithms on media plurality is of importance and, as discussed in the Media Plurality Discussion Document, experimental methods could be used to assess whether differences in choice architecture or recommender systems could affect outcomes such as the diversity of news that people see.
- b) Second, the focus in this paper is on online behavior, but users also consume news offline. Combining information on both online and offline media consumption would provide considerable additional insights. For example, users may be in an echo chamber online but not be offline. While this is a possibility, there is currently limited evidence on this.

6.8 The above list provides some suggestions for potential future research, but it is not exhaustive. In the Media Plurality Discussion Document, we identify several gaps in our understanding of media plurality in an online world that could be filled through further research.

⁶¹ One could also refine this analysis by looking at which categories of news are read by a user.

⁶² There is a growing academic literature that uses experimental methods on social media to study issues including media plurality or online safety.

A1. Definition of the homophily metric

6.9 This appendix contains the precise definition of the homophily metric introduced in Section 5. We computed this metric according to users' following of the 12 news outlets or liking of their content during December 2021. It can, however, also be applied to other measures of the relationship between users and news outlets (e.g., a continuous rather than a binary measure of liking or a measure based on another type of relationship, such as retweeting). The homophily metric is defined in several steps.

6.10 Consider Twitter users i and j with their relationships to news outlets $k = 1, \dots, K$ characterised by outlet weights ω_{ik} and ω_{jk} . In this paper, $K = 12$, so that lower-case k ranges from 1 through 12, representing the 12 news outlets. The outlet weights are simply indicators with value 1 if a given user follows a news outlet or liked some of the outlet's content in December 2021, and 0 otherwise.

6.11 The similarity, $S(i, j)$, between users i and j in terms of their relationship to the K news outlets is then defined as follows. This similarity metric ranges from 0 to 1, with the lower bound representing no overlap in the relationship the users have to the news outlets and the upper bound representing complete overlap.

$$S(i, j) = \sum_k \min(\omega_{ik}, \omega_{jk})$$

6.12 In addition, suppose G denotes a generic group of Twitter users and consider two particular such groups: the friends of user i , $F_{<i}$, and some larger population of users, P . Then, we define the similarity between user i and group G as the average similarity between user i and other users j of the group G . The similarity between user i and the groups $F_{<i}$ or P are defined analogously by replacing G with $F_{<i}$ or P .

$$S(i, G) = \sum_{j \in G} \frac{S(i, j)}{|G|}$$

6.13 Finally, we define the homophily of user i , H_i , as the similarity, in terms of the relationship to news outlets, between user i and their friends, $F_{<i}$, relative to the corresponding similarity between user i and the larger population of Twitter users, P .

$$H_i = S(i, F_{<i}) - S(i, P)$$