Concentration of online news traffic and publishers' reliance on platform referrals: Evidence from passive tracking data in the UK

RASMUS KLEIS NIELSEN
RICHARD FLETCHER
University of Oxford, UK

Attention to online news is highly concentrated and increasingly shaped by platforms including search engines, social media, and aggregators that many use to find and access online news, potentially leaving some publishers highly reliant on platforms, raising the possibility of what has been called "platformization" or "infrastructural capture". We use passive tracking data from the UK to measure how concentrated attention to online news is across different types of access (direct, social media, search engines, aggregators) and to examine how reliant different individual news publishers are on platform referrals. We find that direct traffic to news sites is highly concentrated, whereas all the distributed forms of access analyzed have much lower levels of concentration. While we find that platform referrals are important for most publishers, we identify different profiles in terms of the volume of and reliance on referrals, suggesting that while some are very dependent on platforms, others are not. Overall, we find that while platforms themselves are part of the winner-takes-most concentration of attention overall on the internet, they simultaneously seem to contribute to less concentrated markets for attention to online news.

Keywords: Digital media, platforms, news, direct traffic, distributed traffic

_____

Rasmus Kleis Nielsen: rasmus.nielsen@politics.ox.ac.uk
Richard Fletcher: richard.fletcher@politics.ox.ac.uk
Date submitted: 2021-08-19

Many online dynamics are winner-takes-most, including audience attention to news (Hindman, 2008). Furthermore, audience attention to news is increasingly channeled by various platforms including search engines and social media that people use to access many different types of content (Hindman, 2018).

News publishers remain "primary gatekeepers" in that they continue to be those who decide what becomes what we commonly understand as "the news" (Nielsen, 2017; Singer, 2014), and some of them also retain strong direct connections with some of the public. In the UK, for example, in 2021, 51% of internet users say they have gone directly to a news site or app for news in the last week, and almost all of those who do also say that going direct is their main way of accessing online news (Newman et al., 2021).

At the same time, widespread use of social media, search engines and aggregators to find and access information means a wide variety of "secondary gatekeepers" – most importantly large platforms such as Google Search and Facebook – filter various kinds of content for their users, including news (Thorson & Wells, 2015; Wojcieszak et al., 2021). In the UK, in 2021, 27% say they have come across news while using social media in the last week, 24% say they have used search engines to access news, and 10% say they have used an aggregator to get news (Newman et al., 2021).

In contrast to *direct access*, where people access content by deliberately going to the specific publisher or actor who produced it, platforms enable *distributed discovery*, where people access content produced by third parties (Toff & Nielsen, 2018). Sometimes this is intentional engagement (i.e., by following a specific news brand on social media, or including its name in a search query), sometimes it is automated serendipity (i.e. a general search query returns a news brand the user would not otherwise have accessed) (Fletcher & Nielsen 2018), and sometimes it is incidental exposure (i.e., news seen while using a platform for other purposes such as socializing) (Boczkowski et al., 2018; Fletcher & Nielsen, 2017).

In an environment where many internet users rarely visit news sites directly, distributed discovery provides news publishers with an opportunity to reach more people, and to serve existing audiences across additional channels. Most publishers get a significant share of their online traffic via platforms, and some publishers a majority. This reflects how people use platforms, but also publishers' decision to invest in the opportunities that platforms offer to increase audience reach and serve existing audiences across more channels (Nielsen & Ganter, 2017; Nielsen & Ganter, 2022; Sehl et al., 2021).

At the same time, the fact that platforms deliver much of the audience for news also means that publishers are at risk of "infrastructural capture" (Nechushtai, 2017) or "platformization" (van Dijck et al., 2018) where they are penetrated by the commercial and technical logics of platform companies and become highly reliant on powerful secondary gatekeepers for distribution. Even for publishers who retain strong direct connections with much of their audience, the traffic they get via platforms is characterized by what scholars have variously called "contingency" (Nieborg & Poell, 2018) and "transience" (Barrett & Kreiss, 2019), where platforms continually rework and repackage their products and services based on economic considerations, data on user behavior, and sometimes external pressure. Publishers are keenly aware of how the content farm Demand Media saw a massive drop in online traffic after an update to Google Search in 2011, and how the social media-oriented Upworthy's reach was drastically reduced after Facebook changed their News Feed algorithms in 2014, and know they have to balance the opportunities that platforms offer for increasing their audience against the risks that come with relying on them (Nielsen & Ganter, 2017; Sehl et al., 2021; Nielsen & Ganter, 2022). More broadly, the ranking algorithms that platforms rely on sometimes reinforce existing inequalities (Noble, 2018; Hindman, 2018), underlining the importance of better understanding the winners and losers of this part of the online marketplace for attention. We focus here on platforms that survey research and industry data document are central to how people find and access news, though, to be clear, the category – large technology companies that own and operate digital platforms that enable interaction between at least two different kinds of actors (typically users and advertisers) and in the process come to host public information, organize access to it, and create new formats for it (Nielsen & Ganter, 2022) –

encompasses a wider range than those covered in the analysis below. We do not analyse platforms such as YouTube, Instagram, Snapchat, TikTok, and WhatsApp that, because they are less based around links, have to be studied in other ways than the approach we take here.

Research questions

In this paper, we focus on two key and so far unaddressed descriptive questions that help advance that understanding. First, if, as Hindman has convincingly argued, online attention to news "follows winner-takes-all patterns" (2008, p. 4), is that equally so for all the most widely used forms of access to online news (direct, social media, search engines, aggregators)?[1] Second, how different are individual publishers' relative volume of and reliance on traffic from the various platforms that enable distributed discovery? We address these questions using passive tracking data on a month's worth of desktop and laptop use from a nationally representative sample of UK adults in 2017. Using this dataset we are able to compare how concentrated traffic (measured as visits) is across different types of access, and to map different publishers' profiles in terms of how people access their online news in direct and distributed ways. For both questions, we document significant variation and provide a comparative and disaggregated analysis to advance our empirical understanding of patterns of media use that are often, at best, analyzed at a more aggregate level.

Previous work has documented very strong winner-takes-most dynamics online and highly concentrated patterns of audience attention (Hindman, 2008; 2018). This is most visibly illustrated by the very significant share of web traffic and time spent with the products and services offered by a few large platform companies, especially Google and Facebook, who together account for more than a third of the time people spend with digital media in the UK (and in turn have come to account for a large share of online advertising). But it also applies when it comes to online news more specifically. In the UK, for example, the web analytics company Comscore provides data on about 400 different sites categorized as 'News/Information', but the top five alone account for

---

[1] Hindman used the term attention to refer to the amount (proportion) of internet traffic that comes to a news website, whether direct or via platforms.

well over half of visits, a more concentrated marketplace for attention than, for example, twentieth century print newspaper readership. What we investigate here is whether all the different forms of access audiences rely on are equally concentrated, or whether there are large differences across direct traffic, and the various forms of distributed forms of discovery many rely on? This is our first research question (RQ1)*: How concentrated is online news traffic across different forms of access?*

Comparing degrees of concentration across different modes of access is important for understanding the overall structure of audience attention to news, and how consistent winner-takes-most dynamics are. But this in itself does not give us a more differentiated empirical understanding of individual news publishers' relative volume of and reliance on traffic from platforms. Industry analytics clearly document that platforms overall play a very significant role as drivers of traffic to news sites collectively. Parse.ly estimates that, among its clients, about half of online traffic to news is direct (including internal referrals), with the other half coming in the form of external referrals from platforms, especially search engines (mostly Google) and social media (primarily Facebook).[2] But how, if at all, does individual news publishers' reliance on platforms and consequent exposure to platform risk vary? This is our second research question (RQ2): *how different are individual news publishers' relative volume and relative reliance on traffic from distributed forms of access?*

Method

To address our two questions, we rely on passive tracking data collected in the UK in 2017. Our case country is a high-income democracy ranked 24th out of 180 countries in the 2022 World Press Freedom Index with a media environment characterized by high levels of internet use and a diverse and still relatively robust news industry, including a wide range of different private newspaper publishers, several significant digital-born news publishers, commercial broadcasters, and public service broadcasters—most prominently the license fee-funded BBC. National media systems vary in many important respects (Brüggemann et al., 2014), both in terms of the

---

[2] See https://www.parse.ly/resources/data-studies/referrer-dashboard

industry structure shaping the supply side and in terms of patterns of audience engagement on the demand side. From a comparative perspective, the UK has much in common with a number of other countries in Northern and Western Europe. But the presence of strong public service media means it is a different market from, for example, the United States, its newspaper industry as a whole remains stronger than its counterparts in, for example, Southern Europe, and, while platforms play a prominent role, they are less dominant when it comes to news than in countries like Japan or South Korea. It is an empirical question beyond the scope of this paper whether and to what extend our empirical findings can be generalized from the particular UK context.

Our passive tracking data come from YouGov's Pulse panel, which in March 2017 was made up of 13,709 people in total, and 6,811 active users. Demographic data are provided by panelists upon joining the panel, allowing YouGov to assemble a nationally representative panel in terms of age, gender and region. Panelists are offered incentives in return for installing Wakoopa tracking software, which passively tracks each panelist's internet use (they are free to pause or uninstall it at any time).

We tracked every article accessed from a desktop or laptop to 21 of the most popular UK news websites for a one-month period from 13 March to 10 April 2017. These websites were chosen because they represent some of the most widely used online news sites according to the 2017 Digital News Report that cover UK national (as opposed to local) news and/or UK political news specifically – which in turn provides the definition of the market we study here. However, even under this definition, some prominent websites are missing – a limitation we address later in the analysis. However, it is important to point out now that the news sites not included in our analysis have relatively small audiences. According to Comscore, the 21 sites we analyze account for just over 75% percent of *all* visits to any kind of online news in the UK by over 18s (including those to foreign and local news websites that are outside of our definition).[3]

During the month we covered, 3,071 panelists accessed at least one news article from the publishers included in the tracking. In total, we have data on 168,857 clicks

---

[3] Based on Comscore's MMX Desktop 'Total Visits' measure for March 2018.

on news articles and how people arrived at them. Because we want to compare different forms of discovery specifically, we analyze website visits and the access mode – defined by how the panellist arrived at the first article. This dataset contains 59,539 visits. We exclude from this analysis the 109,318 subsequent article clicks, which are internal referrals

For each news visit, the previous URL was also recorded so that it was possible to infer how the panelist accessed it—directly by clicking on a news article after navigating to a publisher's homepage, via various distributed platforms identified by their URL, or clicking through from a non-platform or non-news website (the last of which we do not include in the analysis below). Because it is very unlikely that panellists would type in an unwieldy news article URL, a direct visit occurs when a panellist clicks through to a news story from the website's homepage. A search visit occurs when the previous URL is an AOL, Bing, Google, or Yahoo! search page. Similarly, Facebook, Twitter and Google News visits occurs when the previous URL was from Facebook, Twitter or Google News. The overall distribution of visits by publisher and type of access (direct versus the main forms of distributed access identified in the dataset, namely Facebook, search, Twitter, and Google News) can be seen in Table 1. (The category "Other" includes referrals from other news sites, from email, as well as clicks where the source of the referral cannot be determined, sometimes called "dark social" by industry analysts.)

Table 1. Visits to each website by access mode.

Website	Direct		Facebook		Search		Twitter		Google News		Other	
	Visits	%	Visits	%	Visits	%	Visits	%	Visits	%	Visits	%
bbc.co.uk/news	11,161	53.3	1,294	13.7	1,019	11.0	374	16.5	157	19.0	2,446	14.6
theguardian.com	3,101	14.8	1,268	13.4	1,529	16.4	541	23.8	131	15.9	2,451	14.6
dailymail.co.uk	3,074	14.7	561	5.9	1,265	13.6	175	7.7	107	13.0	1,377	8.2
telegraph.co.uk	509	2.4	358	3.8	1,309	14.1	230	10.1	72	8.7	982	5.6
independent.co.uk	231	1.1	1,111	11.8	643	6.9	219	9.6	52	6.3	678	4.0
mirror.co.uk	343	1.6	656	7.0	727	7.8	94	4.1	113	13.7	390	2.33
thesun.co.uk	142	.7	688	7.3	816	8.8	67	3.0	19	2.3	306	1.82
express.co.uk	531	2.5	364	3.9	565	6.1	38	1.7	79	9.6	703	4.19
buzzfeed.com	254	1.2	758	8.0	121	1.3	87	3.8	0	0	836	5.0
msn.com	522	2.5	436	4.6	117	1.3	21	.9	1	.1	4,056	24.2
metro.co.uk	192	.9	437	4.6	384	4.1	45	2.0	17	2.1	259	1.5
huffingtonpost.co.uk	110	.5	397	4.2	302	3.2	67	3.0	10	1.2	1,367	8.2
news.sky.com	359	1.7	99	1.0	65	.7	44	1.9	10	1.2	108	.6
standard.co.uk	40	.2	156	1.7	205	2.2	70	3.1	40	4.8	150	.9
itv.com/news	0	0	249	2.6	91	1.0	59	2.6	14	1.7	52	.3
thetimes.co.uk	249	1.2	68	.7	24	.3	61	2.7	1	.1	252	1.5
thecanary.co	0	0	265	2.8	1	< .1	45	2.0	1	.1	10	.1
uk.news.yahoo.com	80	.4	63	.7	104	1.1	2	.1	2	.2	241	1.4
theladbible.com	0	0	187	2.0	2	< .1	16	.7	0	0	19	.1
breitbart.com	32	.2	16	.2	7	.1	14	.6	0	0	93	.6
channel4.com/news	0	0	2	< .1	5	.1	2	.1	0	0	2	<.1
Total	20,930	100.0	9,433	100.0	9,301	100.0	2,271	100.0	826	100.0	16,778	100.0

The passive tracking data we use has two distinct advantages for our purposes. First, because we have the previous URL, allowing us to infer how an article was accessed, we are able to compare how concentrated attention is across different forms of access. Second, because we have granular data on each publisher in our dataset – as opposed to the aggregated data displayed by industry analytics companies and frequently referenced in both professional and scholarly discussions – we are able to map individual publishers' relative volume of and reliance on traffic from platforms. This means we have a more inclusive and granular dataset to work with than the so far most advanced quantitative analysis of different publishers' performance on social media, for example the important work by Bailo et al. (2021) examining the performance of Australian news media on Facebook over time through a combination of CrowdTangle data on interactions (which does not allow comparisons of, for example, social and search) and aggregated SEM Rush data on sources of online traffic (which does not have the granularity of tracking data).

The data also has limitations, however. First, we only have data on desktop and laptop use, not mobile use, which the market research agency eMarketer in 2017 estimated accounted for a little more than half of digital media use in the UK.[4] As other researchers have shown, mobile news use often differs from desktop and laptop use (Dunaway et al., 2018), and it remains an empirical question whether the patterns identified here hold on mobile as well. Second, the data is not drawn from a random sample. YouGov uses quota sampling to build a panel that is representative of the UK population in terms of age and gender (the median age group is 35-44, and 51.5% are male), but Pulse remains an opt-in panel, and may be biased towards particular groups, attitudes and dispositions that in turn are associated with particular patterns of news use. In particular, the panel has received more formal education than the UK population on average (No qualifications: 3.3%, GCSE or equivalent: 11.1%, A-level or equivalent: 18.9%, Degree or above: 51.4%, Other: 13.4%). Third, as mentioned above, the dataset does not include some of the smaller news websites that comprise the UK-focussed national and/or political news market. Fourth, passive tracking does not capture news use on platforms that - due to how they are designed and operate - are much less likely to drive referrals to news sites, including for example YouTube, Instagram, and WhatsApp.

With these limitations in mind, our data still provides a unique opportunity to get a better empirical understanding of how concentrated online news traffic is across different forms

---

[4] See https://www.emarketer.com/content/average-time-spent-in-the-uk

of access and of how different individual news publishers' relative volume and relative reliance on traffic from distributed forms of access are, the two research questions we now turn to.

How concentrated is news traffic across different forms of access?

Following Hindman's work (2008, 2018), we use the Herfindahl–Hirschman Index (HHI) to measure degrees of concentration across different forms of access. Originally developed by economists to measure market concentration (also used to analyze media markets, for example by Noam 2016) and most often focused on share of revenue in a specific market, the index has also been used to measure concentration in the "marketplace for attention". This is done by treating individual publishers' share of attention as the equivalent of a market share, calculated either at the level of a defined market, either at the national level (Hindman, 2008) or local level (Hindman, 2018), or in terms of how concentrated individual users' attention to news is (Trilling & Schoenback, 2013). We follow Matthew Hindman's original approach here and focus on calculating the HHI score for different forms of accessing in the UK-focussed national and/or political news market. News use is of course not only a market (it also has wider social and political dimensions). But the HHI still provides a useful way of assessing how concentrated attention to news is at the publisher level.

For our purposes, the HHI is thus calculated as the sum of the squared market share (in percent) of the clicks going to each brand included in the analysis; or, more formally:

$$H = \sum_{i=1}^{N} s_i^2$$

where s_i is the share of brand *i* in the market for attention to news (defined as visits), and N is the number of brands. It thus has possible values between 0 and 10,000. In practice, analysis is often limited to the biggest players in a given market – something we will also do here given the nature of our data.

HHI scores are interpreted differently by different authorities. The UK Competition and Markets Authority (CMA), for example, generally regards markets with HHIs of more than 1,000 as concentrated, and those with HHIs of more than 2,000 as highly concentrated.[5] The US Federal Trade Commission (FTC) generally sets the bar higher, and regards HHIs between

---

[5] See
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/939636/State_of_Competition_Report_Nov_2020_Final.pdf

1,500 and 2,500 as moderately concentrated and scores above 2,500 as high concentration.[6] As there are no established cut-off points for interpreting HHI scores for attention to news differently from how HHI scores calculated on the basis of revenue for anti-trust purposes, we follow Hindman (2008, 2018) in including competition authorities' cut-off points as a point of reference.

To provide a baseline for our analysis, the HHI for the entire dataset, however accessed, distributed across the 21 news publishers, is 1,325. By UK CMA standards, this suggests a concentrated market for attention to online news. (The figure is driven in large part by the BBC's 28% share of all visits, followed by the Guardian at 15% and the Mail Online at 11%.) For a rough comparison, the HHI for overall television audience share as measured by BARB was 1,755 in the UK the year our data was collected, counting the ten largest broadcasters who account for 92.6% of viewing, and the HHI for audited weekday print newspaper circulation in September 2017 was 1,197.[7] (In television, the BBC again has the largest share, at 31.6%, followed by ITV at 21.7% and Channel 4 at 10.2%. In print newspapers, *The Sun* at 18%, the freesheet *Metro* and the *Daily Mail*, both at 16%, have the largest shares.)

As mentioned earlier, even if we define the market of interest as UK-focussed national and/or political news, our dataset is missing visits to some websites. To estimate the effect that these omissions might have on our results, we attempt to impute visits from missing sites based on Comscore data. Comscore provides basic statistics on the use of websites based on a combination of passive tracking and server-side measurement. Among other things, they provide estimates of monthly visits to web domains for the whole UK population. Thus, we can use the data from March 2018 (2017 data was not available) to estimate the number of visits that would have been recorded with a sample of our size. These visits were then added into the dataset, and a separate HHI produced based on the recomputed market shares. Desk research identified 17 further websites that could be included under our definition of the market, bringing the total to 38 (analysis is often restricted to the 50 largest firms within a market, see e.g., Ganapati 2021). The HHI for the 38-website dataset was 1,245 (see Supplementary Materials for more information). Adding extra websites, no matter how small, will inevitably lower the level of concentration – but the fact that none of the additional sites

---

[6] See https://www.justice.gov/atr/horizontal-merger-guidelines-08192010
[7] See https://www.barb.co.uk/download/?file=/wp-content/uploads/2018/04/Barb-Viewing-Report-2017_FINAL_LR-May-2018.pdf and https://www.pressgazette.co.uk/national-newspaper-abcs-bulks-helped-times-and-daily-telegraph-boost-print-circulations-in-september/

have a market share of more than 1% means that difference between the 21-website and 38-website datasets is small, and the classification of the market as concentrated does not change. On this basis, and because it is not possible to split the Comscore data by access mode, we proceed with the original dataset.

To be able to compare how concentrated attention to online news is across different forms of access, we treat each as a distinct market where publishers compete to attract visits directly or via different platforms. It is important to recognize here that, when compared to offline media and to direct discovery, distributed forms of discovery happen in very particular kinds of markets, enabled and structured by the technologies, community standards, and terms of service provided by individual for-profit platform companies who benefit commercially from enabling transactions via the products and services they offer (van Dijck et al., 2018; Poell et al., 2022). Treating referrals from, say, Facebook, as a form of market for attention and examining how concentrated it is should not be confused with naturalizing Facebook, or with assuming that the company won't change how it treats news—something it has done multiple times, sometimes with dramatic consequences.

Proceeding on this basis, the data presented in Table 1 then enable us to calculate the HHI for each form of access. The scores are shown in Figure 1, with the baseline figure for the whole dataset (online news visits overall) and HHI scores for television viewing and weekday print newspaper circulation included as benchmarks.
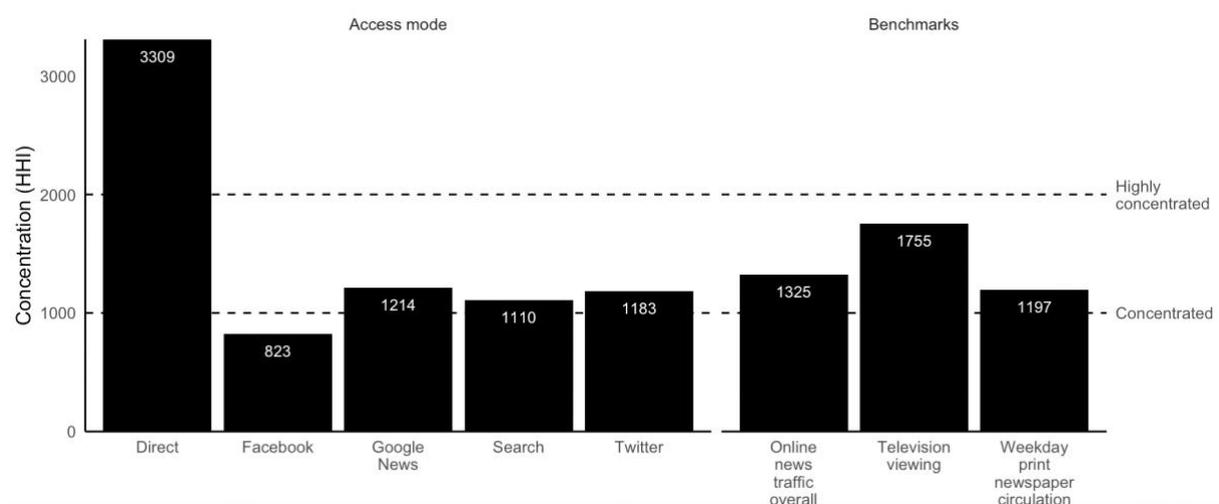

Figure 1. Concentration levels across different access modes.

What we find in response to our first research question is thus that (a) online news traffic in the UK is concentrated overall (by the CMA standard), primarily driven by highly concentrated direct access and (b) the HHI for every other form of online news access is less concentrated. The HHI scores for the most important forms of distributed access to online news are lower than for television viewing and about at the level of print newspaper circulation (a market that by CMA standards is concentrated, but falls under the FTC guideline threshold for moderate concentration).

We thus show that there are clearly very different degrees of concentration across different forms of access. While the winner-takes-most dynamics that Hindman identified remain strong online (both generally, in terms of the role of platform companies such as Google and Facebook, and for news specifically), all the main forms of distributed discovery of news analyzed here have much lower levels of concentration, with a larger number of publishers each capturing a significant share.

How different are individual news publishers' relative volume and relative reliance on traffic from distributed forms of access?

But how different are individual news publishers' profiles in terms of how people access their online news in direct and distributed ways, then? As mentioned, surveys document that many internet users rely on platforms to find and access online news, industry analytics provide an aggregate overview documenting the overall importance of distributed discovery, and individual news publishers have internal proprietary data that helps them understand their own audience profile. Our data allow us to go well beyond this to offer a more granular analysis comparing different publishers' profile in terms of their relative volume and relative reliance on traffic from distributed forms of access.

Figure 2 summarizes the descriptive data contained in Table 1 in terms of the absolute volume of direct traffic and absolute volume of traffic from distributed forms of access (social, search, and aggregator combined) for each publisher. We use log scales on this chart because of the large disparities in overall traffic between different publishers. There is a clear, but by no means direct relation between the volume of direct traffic and the volume of distributed traffic ($r = .67$, $p < .001$). While the publishers who attract the most direct traffic (the BBC, the *Guardian*, etc.) also attract a significant volume of distributed traffic, most other titles do

relatively better in terms of their share of distributed traffic than of direct traffic (e.g., the *Independent*, the *Daily Telegraph*, the *Daily Mirror*).
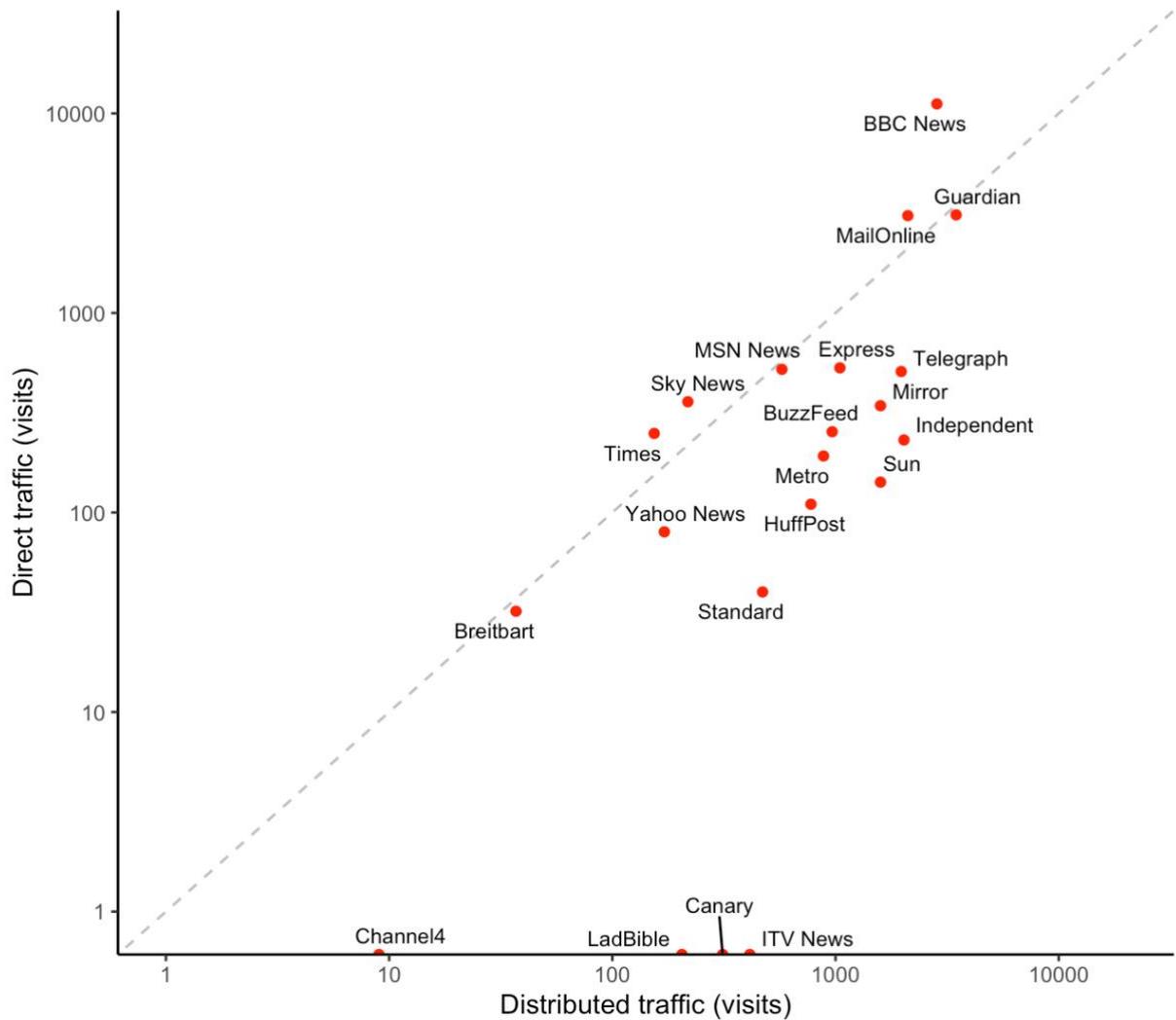

Figure 2. Publisher's direct traffic plotted against distributed traffic.

To visualize individual publishers' relative volume of and relative reliance on distributed forms of access, we now plot publisher's reliance on distributed traffic (measured as the percentage of total visits that come from distributed access) against total distributed traffic.
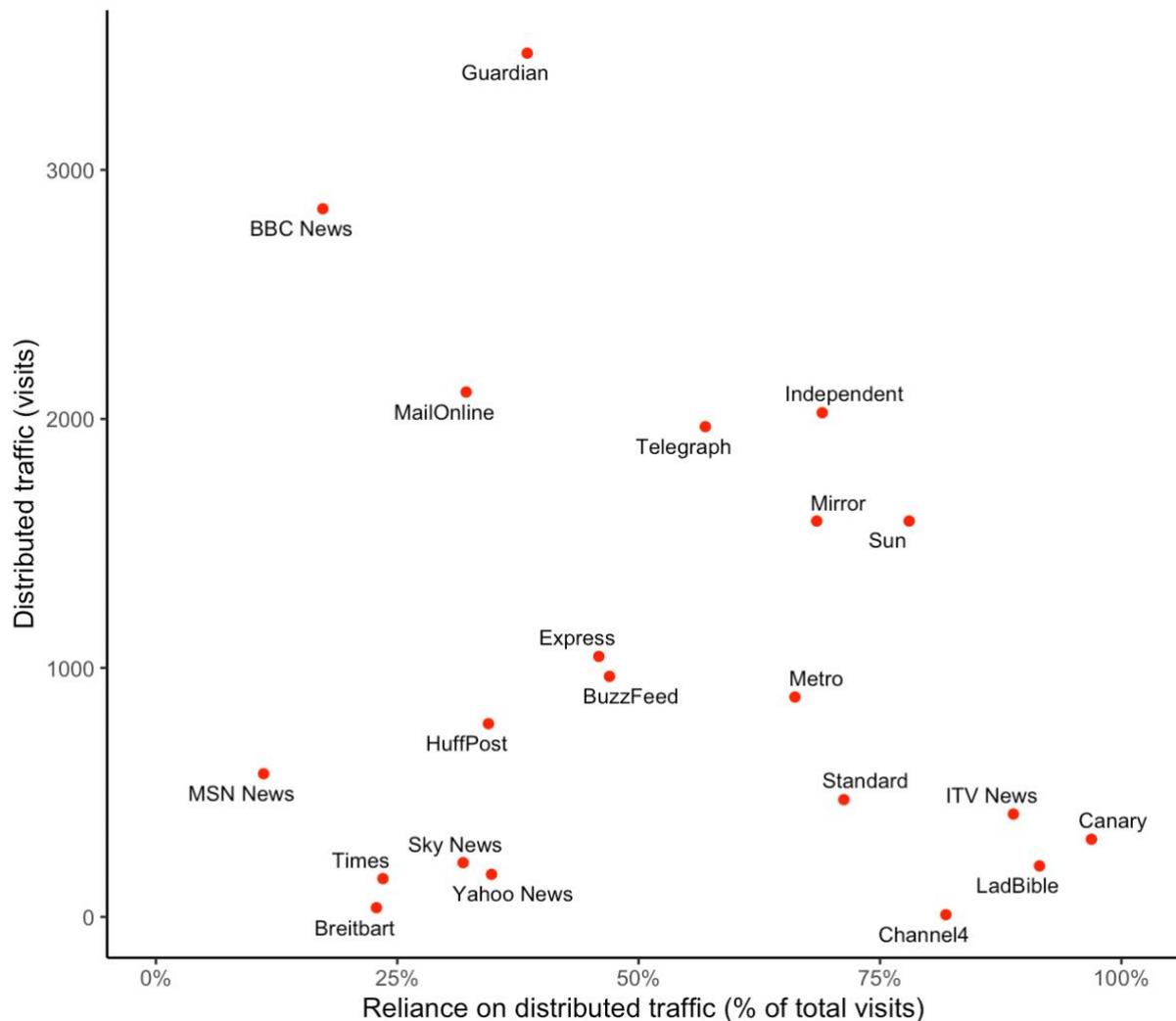
Figure 3. Publisher's reliance on distributed traffic plotted against distributed traffic.

Figure 3 enables us to go beyond the topline understanding provided by industry data on the overall role of distributed discovery for the news industry as a whole, and help us identify a wide range of different publisher profiles. While a number of publishers are in the middle of the figure, we can also identify four other real types. First, <u>high volume/low reliance</u> in the top left corner, where for example the *Guardian*, across social, search, and aggregators gets more distributed traffic than any other publisher in our dataset, but also attracts a very significant amount of direct traffic. Second, <u>high volume/high reliance</u>, where the *Independent* rivals publishers with far larger direct traffic like the *MailOnline* in terms of the volume of distributed traffic it attracts, but also is more reliant on it, as its more modest direct traffic means it gets well over half of its traffic this way. Third, <u>low volume/high reliance</u>, where for example ITV

News gets about one-tenth of the distributed traffic of the *Guardian*, but this still makes up a clear majority of its total visits. Fourth, <u>low volume/low reliance</u> where for example the *Times of London* – long an industry outlier in terms of its approach to platforms, going back to Rupert Murdoch's 2009 decision to remove the newspaper's online content from Google's search index – attracts only a limited amount of distributed traffic, and also has very little reliance on it, as it gets less than ten percent of its traffic this way.

What we find in response to our second research question is thus that (a) a number of publishers have roughly the reliance on distributed traffic that aggregate industry data would suggest but (b) there is significant variation in individual publishers' relative volume of and relative reliance on traffic from distributed forms of access, with four different empirical profiles including high volume/low reliance, high volume/high reliance, low volume/high reliance, and low volume/low reliance identified. All publishers who rely in part on platforms to reach people online are susceptible to the forms of "contingency" (Nieborg & Poell, 2018) and "transience" (Barrett & Kreiss, 2019) described by scholars highlighting how ranking algorithms, community standards, and terms of service can sometimes change at a moment's notice. Scholars have not yet developed ways of operationalizing terms such as "infrastructural capture" (Nechushtai, 2017) or "platformization" (van Dijck et al., 2018) in quantitative terms, but we believe our analysis can help ground discussions around these phenomena empirically by documenting that publishers' reliance on platforms clearly varies greatly, with some publishers attracting significant distributed traffic while retaining a large volume of direct traffic, and those who primarily get direct traffic, much less exposed than those more reliant on various platforms.

Discussion

In this paper, we have analyzed how concentrated online news traffic is across different forms of access (including direct as well as search, social, and aggregators) and how different individual news publishers' relative volume and relative reliance on traffic from distributed forms of access are. Using passive tracking data from the UK, we have shown that news traffic via various forms of distributed discovery has lower levels of concentration than direct traffic, and we have documented significant variation in individual publishers' relative volume of and relative reliance on traffic from distributed forms of access.

It is important to note that the lower levels of concentration documented are relative to the pronounced winner-takes-most dynamics in the highly concentrated market for direct traffic, and in several cases about comparable to the levels of concentration seen in broadcast

television and print newspaper circulation (themselves often regarded as concentrated or moderately concentrated markets). What we have documented here is not some sort of perfect pluralist heaven, but somewhat less concentrated traffic patterns with a slightly larger number of publishers capturing a significant share of attention. In terms of overall attention to online news, and individual publishers' audience profile, distributed access may, as they have been shown to do for individual users (Fletcher et al., 2021; Stier et al., 2021), increase diversity compared to direct access, but only slightly. Smaller titles are generally still small, just relatively less so. Bigger titles still dominate, only relatively less so.

It is also important to keep in mind that the distributed forms of access we have analyzed here are intermediated by platform companies (some of whom in turn operate in concentrated or very concentrated markets for, e.g., search engines or social media). While it is hard for outside researchers to assess exactly how much activity on each platform is directly or indirectly related to news – estimates are often in the broad range of how much of overall internet use is focused on news, so a few percent (Hindman, 2018; Nielsen & Ganter, 2022) – even as platform company executives from time to time stress that news is an important part of what they offer their users. The crucial thing to recognize here is that in most cases, it is news *in general*, not any one *individual* news publisher, that is important to platforms, and that they have clear commercial incentives to privilege at least some diversity – as Google's Vice-President of Content Partnerships David Eun has said, the task of platform companies' partnership teams is to "create a sense of infinite choice" for users.[8] This illustrates the tension between, on the one hand, the way platforms aim to serve their end users—through diversity and volume of content—and, on the other hand, the interests of individual publishers, who seek to stand out and fear being reduced to commodified, substitutable pieces of content.

The role of platform companies, with their track record of constantly changing their products and services, sometimes dramatically – as illustrated by the case of Demand Media and Upworthy – is also an important part of interpreting the very different publisher profiles we have identified: high volume/low reliance, high volume/high reliance, low volume/high reliance, and low volume/low reliance. Publishers do not simply get to decide to have a profile like the *Guardian*, it is the result of years of expensive efforts that not all news media may be in a position to invest. Platforms have the upper hand, and can change their products and services more or less as they see fit, but while the incentive structures they create play an

---

[8] Quote from https://www.youtube.com/watch?v=li5T5eM4B3g.

important role in shaping outcomes, so too does the tactical and strategic choices that publishers make. Demand Media, deliberately built to make the most of the opportunities afforded by search engines, and Upworthy, similarly built to make the most of social media (Nielsen & Ganter, 2022) both fit the high volume/high reliance profile that can make the most of platform opportunities but also, as a consequence, are most exposed to platform risk. Other publishers have very different levels of exposure – those with low volume/high reliance are also exposed, even if both their gains and their possible losses are smaller in absolute terms, whereas those with high volume/low reliance may see significant fluctuation in absolute terms, but much less exposure because of their larger direct traffic. Finally, those with low volume/low reliance have little exposure – an outcome that, at the *Times* of London, is no accident, but the direct result of taking a very different approach than competitors such as the *Guardian*. This also means that publishers have very different prospects if platforms where to, either for their own strategic reasons or in response to external pressures, reduce the amount of news they present to their users. The least reliant publishers, irrespective of their absolute volume of traffic, would feel the least impact overall, whereas the most reliant publishers would likely lose a very significant part of their overall audience.

While limited to desktop and laptop data and to a single case country of the UK (characterize among other things by the prominent role played by the BBC), our analysis help capture the complex impact platforms have on news publishing, and how different publishers' positions are in an increasingly platform-dominated media environment. The share of audience attention and advertising that especially the biggest platforms such as Google and Facebook capture is a serious challenge to the sustainability of many traditional news publishers. At the same time, the very same platforms that in that sense risk undermining diversity on the *supply* side (because they challenge the business behind news) also channel attention on the *demand* side in ways that are less concentrated than what people chose to go to direct to online news. While this development, and how publishers have adapted to it, leave some news publishers heavily exposed to platformization, this is far from universal, as others benefit from high volumes of referral traffic while maintaining low reliance on it. So though platforms themselves are part of the winner-takes-most concentration of attention overall on the internet, they simultaneously seem to contribute to less concentrated markets for attention to online news.

Funding

References

Bailo, B., Meese, J., & Hurcombe, E. (2021). The Institutional Impacts of Algorithmic Distribution: Facebook and the Australian News Media. *Social Media + Society, 7*(2): 20563051211024964. https://doi.org/10.1177/20563051211024963

Barrett, B., & Kreiss, D. (2019). Platform Transience: Changes in Facebook's Policies, Procedures, and Affordances in Global Electoral Politics. *Internet Policy Review, 8*(4): 1–22. https://doi.org/10.14763/2019.4.1446

Boczkowski, P. J., Mitchelstein, E., & Matassi, M. (2018). 'News Comes across When I'm in a Moment of Leisure': Understanding the Practices of Incidental News Consumption on Social Media. *New Media & Society*, *20*(10): 3523–39. https://doi.org/10.1177/1461444817750396

Brüggemann, M., Engesser, S., Büchel, F., Humprecht, E., & Castro, L. (2014). Hallin and Mancini Revisited: Four Empirical Types of Western Media Systems. *Journal of Communication*, October, n/a-n/a. https://doi.org/10.1111/jcom.12127

van Dijck, J., Thomas Poell, T., & de Waal. M. (2018). *The Platform Society: Public Values in a Connective World*. New York: Oxford University Press.

Fletcher, R., & Nielsen, R. K. (2017). Are People Incidentally Exposed to News on Social Media? A Comparative Analysis. *New Media & Society*, August, 1461444817724170. https://doi.org/10.1177/1461444817724170

Fletcher, R., & Nielsen, R. K. (2018). Automated Serendipity: The Effect of Using Search Engines on News Repertoire Balance and Diversity. *Digital Journalism, 6*(8): 976–89. https://doi.org/10.1080/21670811.2018.1502045

Fletcher, R., Kalogeropoulos, A., & Nielsen, R. K. (2021). More Diverse, More Politically Varied: How Social Media, Search Engines and Aggregators Shape News Repertoires in the United Kingdom. *New Media & Society*, July, 14614448211027392. https://doi.org/10.1177/14614448211027393

Ganipati, S. (2021). Growing Oligopolies, Prices, Output, and Productivity. *American Economic Journal: Microeconomics*, *13*(3), 309–327. https://doi.org/10.1257/mic.20190029

Hindman, M. (2008). *The Myth of Digital Democracy*. Princeton: Princeton University Press.

Hindman, M. (2018). *The Internet Trap: How the Digital Economy Builds Monopolies and Undermines Democracy*. Princeton: Princeton University Press.

Nechushtai, E. (2017). Could Digital Platforms Capture the Media through Infrastructure? *Journalism*, August, 1464884917725163. https://doi.org/10.1177/1464884917725163

Newman, N, Fletcher, R., Schulz, A., Andı, S., Robertson, C. T., & Nielsen, R. K. (2021). *Reuters Institute Digital News Report 2021*. Oxford: Reuters Institute for the Study of Journalism.

Nieborg, D. B., & Poell, T. (2018). The Platformization of Cultural Production: Theorizing the Contingent Cultural Commodity. *New Media & Society*, April, 1461444818769694. https://doi.org/10.1177/1461444818769694

Nielsen, R. K. (2016). News Media, Search Engines and Social Networking Sites as Varieties of Online Gatekeepers. In C. Peters, & M. Broersma (Eds.) *Rethinking Journalism Again: Societal Role and Public Relevance in a Digital Age* (pp. 81-96). Routledge.

Nielsen, R. K., & Ganter, S. A. (2018). Dealing with Digital Intermediaries: A Case Study of the Relations between Publishers and Platforms. *New Media & Society, 20*(4): 1600–1617. https://doi.org/10.1177/1461444817701318

Nielsen, R. K., & Ganter, S. A. (2022). *The Power of Platforms: Shaping Media and Society*. Oxford, New York: Oxford University Press.

Noam, E. M. (2016). *Who Owns the World's Media?: Media Concentration and Ownership around the World*. Oxford: Oxford University Press.

Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: University Press.

Poell, T., Nieborg, D. B., & Duffy, .B. E. (2022). *Platforms and Cultural Production*. Cambridge: Polity Press.

Sehl, A., Cornia, A., & Nielsen, R. K. (2021). How Do Funding Models and Organisational Legacy Shape News Organisations' Social Media Strategies? *Digital Journalism*. https://doi.org/10.1080/21670811.2021.1968920

Singer, J. B. (2014). User-Generated Visibility: Secondary Gatekeeping in a Shared Media Space." *New Media & Society, 16*(1): 55–73. https://doi.org/10.1177/1461444813477833

Stier, S., Mangold, F., Scharkow, M., & Breuer, J. (2021). Post Post-Broadcast Democracy? News Exposure in the Age of Online Intermediaries. *American Political Science Review*, October, 1–7. https://doi.org/10.1017/S0003055421001222

Toff, B., & Nielsen, R. K. (2018). 'I Just Google It': Folk Theories of Distributed Discovery. *Journal of Communication, 68*(3): 636–57. https://doi.org/10.1093/joc/jqy009

Trilling, D., & Schoenbach, K. (2013). Patterns of News Consumption in Austria: How Fragmented Are They? *International Journal of Communication, 7*(0): 25.

Wojcieszak, M., Menchen-Trevino, E., Goncalves, J. F. F., & Weeks, B. (2021). Avenues to News and Diverse News Exposure Online: Comparing Direct Navigation, Social Media, News Aggregators, Search Queries, and Article Hyperlinks: *The International Journal of Press/Politics*, May. https://doi.org/10.1177/19401612211009160

Supplementary materials

Table S1 contains information about the imputed 38-website dataset referred to in the main text. Comscore estimated visits for the entire 2018 UK over-18 population of 52,403,344[9] were divided by 7,694 to match the visits from the 6,811 active panelists from the original dataset. This should produce broadly comparable figures that can be added to the original dataset. The names of the additional sites in Table S1 are the entity names used within Comscore.

Table S1. Visits to each website and imputed visits based on Comscore data.

	21-website dataset		Imputed 38-website dataset	
Website	Visits	%	Visits	%
bbc.co.uk/news	16,451	27.63	16,451	26.75
dailymail.co.uk	90,21	15.15	9,021	14.67
theguardian.com	6,559	11.02	6,559	10.66
msn.com	5,153	8.65	5,153	8.38
telegraph.co.uk	3,460	5.81	3,460	5.63
independent.co.uk	2,934	4.93	2,934	4.77
mirror.co.uk	2,323	3.90	2,323	3.78
express.co.uk	2,280	3.83	2,280	3.71
huffingtonpost.co.uk	2,253	3.78	2,253	3.66
buzzfeed.co.uk	2,056	3.45	2,056	3.34
thesun.co.uk	2,038	3.42	2,038	3.31
metro.co.uk	1,334	2.24	1,334	2.17
news.sky.com	685	1.15	685	1.11
standard.co.uk	661	1.11	661	1.07
thetimes.co.uk	655	1.10	655	1.06
DAILYSTAR.CO.UK			501	.82
uk.news.yahoo.com	492	.83	492	.80
itv.com/news	465	.78	465	.76
NEWSNOW.CO.UK			426	.69
thecanary.co	322	.54	322	.52
theladbible.com	224	.38	224	.36
FT.COM			199	.32
RT.COM			195	.32
breitbart.com	162	.27	162	.26
AOL UK News			115	.19
CITYAM.COM			102	.17

	21-website dataset		Imputed 38-website dataset	
Website	Visits	%	Visits	%
INDY100.COM			93	.15
Politico			92	.15
LBC.CO.UK			85	.14
INEWS.CO.UK			78	.13
SPUTNIKNEWS.COM			26	.04
ALJAZEERA.NET			18	.03
FLIPBOARD.COM			11	.02
channel4.com/news	11	.02	11	.02
MORNINGSTAR.CO.UK			9	.02
EVOLVEPOLITICS.COM			8	.01
POLITICSHOME.COM			4	.01
GAWKER.COM			3	.01

Note. Visits in 'Imputed 38-website dataset' column based on data from Comscore's MMX Desktop 'Total Visits' measure for March 2018.