

University of Groningen

News devices

Bounegru, Liliana

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2019

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Bounegru, L. (2019). *News devices: how digital objects participate in news and research*. University of Groningen.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

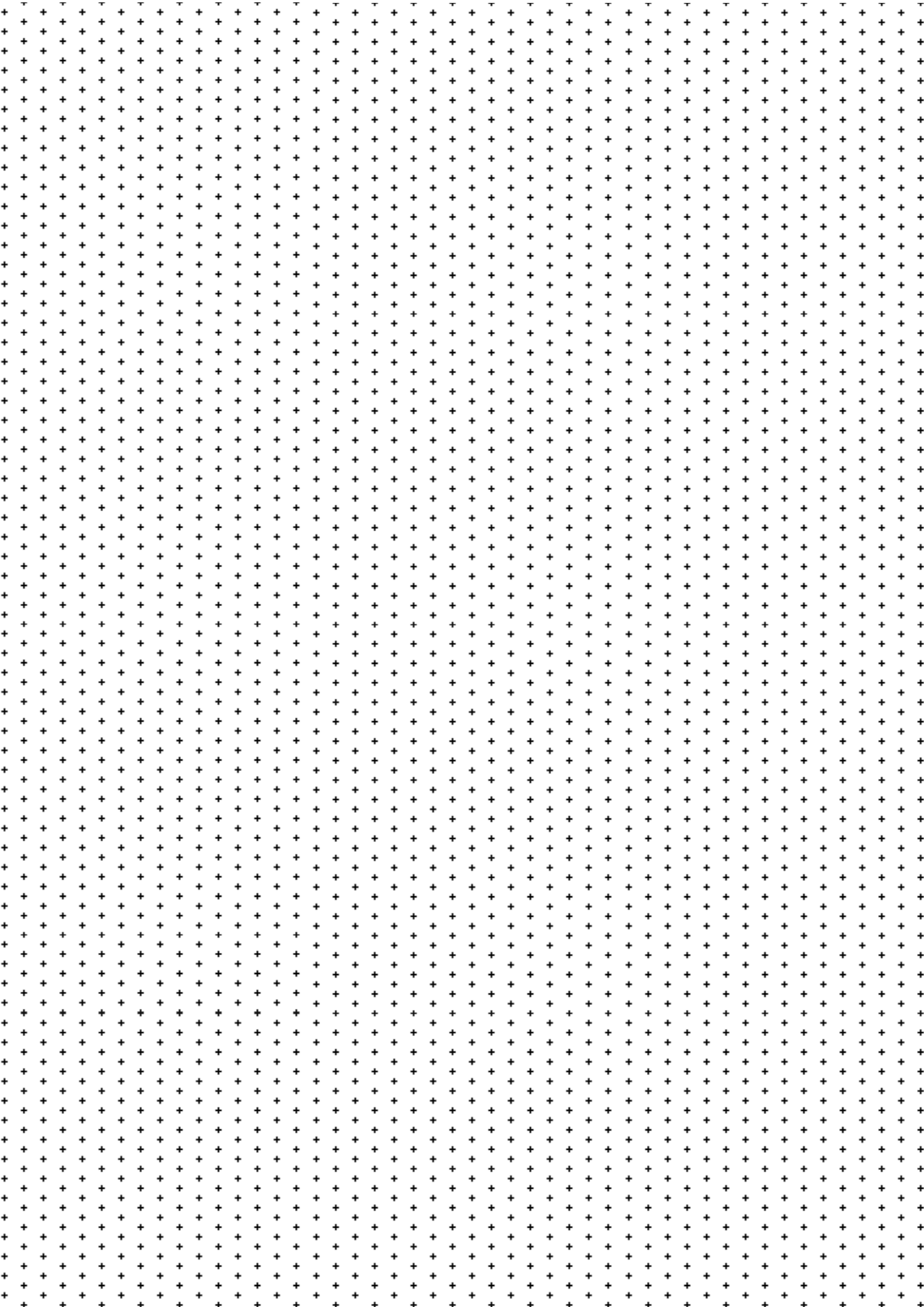
The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

5. Making Audience: Web Trackers as Audience Marketplace Devices in Professional and Junk News



In the previous chapter I engaged with GitHub as a connective coding device. I discussed how GitHub platformises journalism coding and renders it amenable to financial capitalisation through interplays between its technical infrastructure and economic imperatives. In this chapter I take on another digital device that renders another type of news work economically valuable, namely third-party web trackers.⁵⁹ In doing so I expand the range of actors that matter in relation to digital journalism to include three other important ones in the discussion: audiences, the online advertising and marketing industries, and other forms of online content production known as fake news.

The “fake news” scandal is linked to increasing concerns about the dawn of a “post-truth era” (Sismondo, 2017), characterised by the unsettling of established epistemic hierarchies, the blurring of boundaries between news and other less reputable forms of digital cultural production and an increasing suspicion of experts and expert knowledge on the side of the public.

This concern about disruptions of established knowledge hierarchies is well illustrated by reports that false stories about the 2016 US presidential elections, packaged as news, outperformed “real” news in the domain of Facebook engagement rates (Silverman, 2016). That is, dubious facts had circulated more successfully and reached more individuals on this social media platform than journalistic facts about the US elections. This, alongside several other scandals, have brought the platform’s reputation into decline.

But this is not the first time that the internet’s reputation is questioned due to its association with deceptive knowledge practices. Rogers (2002) notes that

⁵⁹ This chapter is based on research conducted as part of a collaborative multi-institutional project I co-led with Jonathan Gray, Tommaso Venturini and Michele Mauri in 2017-2018 with support from First Draft and in collaboration with several journalists, media organisations, public institutions and others. The first phase of the research that underpins this chapter was conducted during a data sprint at the University of Amsterdam in March 2017 together with Michele Invernizzi and Mischa Szpirt. An earlier version of this study was published as a chapter in the research report *A Field Guide to “Fake News” and Other Information Disorders* (Bounegru et al., 2018). Another version of this research will be published as part of an article in *New Media & Society* (Gray, Bounegru & Venturini, forthcoming). For the purposes of this chapter the analyses were redone, extended and rewritten.

from its early days the web has raised concerns as “a medium of dubious repute” (p.192), a “rumour mill” where “rogue” websites and “rumour merchants” engage in spreading questionable and toxic content. These early concerns faded at least in part with the advent of search engines such as Google, which organised web sources based link authority logics, although they raised other concerns about the information politics of such ordering mechanisms (Rogers, 2004).

Today concerns about the role of online platforms in spreading dubious content have returned to public attention, and this time have brought social media and their like economy (Gerlitz & Helmond, 2013) under intense scrutiny. As colleagues and myself argue elsewhere, social media platforms should be seen as a matter of concern not only because of their failure to appropriately respond to “fake news” but, perhaps more importantly, because of their uncanny success in enabling the circulation, virality, monetisation and consolidation of such forms of digital cultural production, a phenomenon which elsewhere we discuss as “infrastructural uncanny” (Gray, Bounegru & Venturini, forthcoming).

Indeed, the like economy, set up produce virality by intensifying the social validation of content and user engagement by means of social buttons and engagement counters spread across the web, is instrumental to such digital forms of “content farm[ing]” (Turow, 2011). Investigations exposed the profits that can be made from such forms of digital content production through an economic model based on the successful combination of low cost clickbaity sensationalist content which has the potential to go viral, with what in advertising is known as “click arbitrage”: buying website traffic cheaply through advertising on platforms such as Facebook and bringing these audiences on the site, where ad revenue can be made through ad networks such as Google’s AdSense (Nickel, n.d.; see also Dewey, 2016; Silverman & Alexander, 2016; Subramanian, 2017).

While debates about the economics of fake news have focused on exploring approaches to sanction disreputable publishers, such as by blacklisting them

from ad networks and flagging their content on Facebook, in this chapter I suggest that “fake news” can also be viewed as an “empirical occasion” (Marres, 2013; Marres & Moats, 2015). As an empirical occasion, fake news renders visible and amenable for social analysis not just the relations between *rogue* content producers, social media, audience participation and online advertising. It also renders analysable the relations between value creation, content production, social media, audience participation and online advertising that underpin digital cultural production *more generally*, including the more reputable kinds. In other words, “fake news” may be seen as an opportunity to explore the economics of various forms of digital cultural production.

Indeed, the dotcom crash has seen a shift in online business models towards advertising-supported models organised around the production and sale of audiences for most businesses with an online component (West, 2017). News publishers too have early on realised that people would not be willing to pay for news online and have turned to advertising models (Turow, 2011).

Richard Serra’s iconic 1973 short film “Television Delivers People” reminds us that advertising-supported cultural production generates not just programming, entertainment and news but also the audience product. In the context of online news, the implementation of the EU General Data Protection Regulation (GDPR) in May 2018 has made visible again the interdependence between news production and audience tracking and commodification, as some news websites unable to guarantee lawful use of their audience’s personal data have paused their services to EU citizens.



Unfortunately, our website is currently unavailable in most European countries. We are engaged on the issue and committed to looking at options that support our full range of digital offerings to the EU market. We continue to identify technical compliance solutions that will provide all readers with our award-winning journalism.

Figure 22: (a) Screenshot from Richard Serra's iconic 1973 short film "Television Delivers People". (b) Screenshot from the LA Times' website after the enforcement of the GDPR, indicating the website is unavailable to users accessing the website from the EU.

While audience commodification has traditionally been the object of political economy and critical media research, in this chapter I link research into the economic underpinnings of news with socio-technical approaches to digital media, in order to examine the tracking infrastructures and practices that underpin audience commodification. I focus on two forms of advertising-supported digital cultural production, professional news and junk news. The question that drives this chapter is: *What can the news device approach contribute to the study of the audience marketplaces in which advertising-supported digital cultural production is embedded?*

I start by revisiting digital transformations of the media audience marketplace (Napoli, 2003, 2011). Next I introduce a news device approach to studying the

audience marketplaces in which different forms of digital cultural production are embedded, focused on the tracking infrastructures of websites. Finally, I discuss the results of an empirical visual network exploration of audience marketplace practices and tracking infrastructures across a set of professional and junk news sites, and their implications.

5.1 Digital Transformations of Audience Marketplace

Practices

While audiences play multiple roles in news and their relations with journalism are multifaceted (for a discussion, see, e.g., Lewis & Westlund, 2015; Peters & Witschge, 2015), in this chapter the focus is on one particular aspect of this relationship, the audience as economic product or commodity. There is a long history in critical media and political economy research that examines the construction of the “audience commodity” as part of the analysis of advertising-supported media and cultural production (for an early account, see, Smythe, 1977). But in what follows I will primarily focus on the transformations to the audience marketplace brought about by the internet and digital platforms.

In the context of news, the turn from partisan newspapers to the penny press in the US in the 19th century marked a shift in media business models “from the sale of *products*—newspapers—to the sale of news corporations’ *audiences* to subsidise media production” (West, 2017, p. 5). Today a great part of news production in many countries is advertising-supported (Nielsen, 2016). The audience product in the context of these advertising-supported operations is at its simplest understood as a “representation of consumer attention to advertising messages” (Napoli, 2003, p. 22). The production and exchange of this product is sustained by a marketplace arrangement which requires multiple actors and mechanisms to measure audiences, calculate and negotiate their value, rate publications based on their audiences and exchange audience attention for money.

The internet and digital platforms and services are linked to significant developments in the audience marketplaces associated with various forms of cultural production. According to Napoli (2011), the internet today represents a laboratory for the media audience marketplace. In this laboratory audience marketplace participants experiment with various programmatic instruments, techniques and methods to construct and deliver the audience product by tracing, measuring and calculating user characteristics, behaviour, tastes and preferences and anticipating future behaviour. They also test techniques to target and serve ads and measure their effectiveness, to assign value to both audience segments and websites, and, on the user side, to resist such measurements through anti-tracking and ad-blocking software.

The currency around which the audience marketplace is traditionally organised is the exposure metric, also known as “eyeballs” (Napoli, 2011). One of the first cases of online advertising is reported to be dating from 1994 when HotWired displayed an AT&T banner ad on its page (Evans, 2008; Turow, 2011). Early online advertising was organised around banners or display ads which were sold based on measures similar to those operating in traditional media such as the impression-based model of “cost-per-mille”, or cost per one thousand individuals who were served the ad (Evans, 2008; Ratliff & Rubinfeld, 2010).

In the second half of the 1990s, a competing currency is established in the context of search engines, the cost-per-click (Evans, 2008). Associated with the rise of search engines and search-based advertising, this coin of exchange sees payments made only when viewers click on ads. The rise of search advertising has seen the revenue share from online display ads, used by the news media, declined by more than half in the first decade of the 21st century (Evans, 2008). The development of performance-based models such as the cost-per-click, and later of the cost-per-action, are part of what Napoli (2011) calls the “post-exposure audience marketplace”, where, alongside exposure, other currencies, measurement instruments and audience conceptions are being experimented with and are shaping what the audience product is becoming.

The configuration and actor composition of the post-exposure audience marketplace is complex. While traditionally participants in this marketplace consisted of media publishers, advertisers (and related service providers such as advertising agencies), audience measurement companies and audience members (Napoli, 2003), today we are faced with an increasingly complex configuration of “interlocking multisided platforms” (Evans, 2008; for a discussion of multi-sided markets, see Chapter 4).

While traditionally news media were two-sided markets which provide readers with information and entertainment and advertisers with an audience for their product (Nielsen, 2018; Rochet & Tirole, 2003; van Couvering, 2017), today they are undergoing a process of platformisation, whereby they are joining other types of cultural producers in the precarious position of content providers in online platform markets, such as that organised by Facebook (van Couvering, 2017; Nieborg & Poell, 2018). In doing so they are becoming increasingly reliant on platform instruments to make audiences measurable, calculable, analysable and economically valuable.

At the same time, online advertising and marketing industries are becoming increasingly complex and are too becoming dependent on social media platforms such as Facebook (see, e.g., Helmond, Nieborg, & van der Vlist, 2017).

According to Nieborg (2016), in recent decades “the arrival of new intermediaries, the changing role of incumbents, and the adoption of internet-enabled mobile devices resulted in an increasingly opaque multi-sided market structure” (p. 4). In this complex arrangement exchanges and connections between the supply and demand side of advertising inventory are mediated by an increasingly large number of intermediaries and data brokers, both more established (e.g. audience measurement firm Nielsen) and more recent (e.g. online platforms and their marketing and advertising services). An important development is the emergence of advertising networks such as Google AdSense to monetise the “long tail” of the internet, i.e. small websites and user generated content (desilva + phillips, LLC, 2008; Gehl, 2014). Ad networks

rely on third-party tracking techniques to enable advertisers to reach audience profiles across a large number of sites, as well as programmatic tools for publishers to sell their media inventory, and for advertisers to buy it and place ads. A more recent intermediary, ad exchanges, centralises the selling and buying of ad space and takes its automation a step further through real-time bidding, the buying and selling of impressions, and the possibility to reach the user loading a website in real-time by placing bids in automated auctions. Indeed, any unreserved ad space on a publisher's website may enter the automated real-time bidding system of ad exchange service where ad networks, demand-side platforms or other ad exchanges can place bids (amounts of money they are willing to pay) to fill the ad space in real-time, as a user's browser loads a webpage (Interactive Advertising Bureau, 2014).

In this context, yield optimisation companies or supply-side platforms, such as PubMatic or Rubicon, take on the role of supporting publishers in their programmatic interactions with advertising platforms by e.g. evaluating and filtering bids according to the publisher's criteria (Turow, 2011). On the advertisers' side, intermediaries include demand-side platforms, which provide media buying services to advertisers and mediate interactions with ad exchanges in the benefit of the advertiser.

Increasingly important in this ecosystem are also data intermediaries or data brokers such as BlueKai or Lotame. These are services that aggregate user information from multiple online and offline sources and offer or resell it to a number of other companies (Zuiderveen Borgesius, 2014), such as supply-side platforms. These use it to enrich user profiles made available to advertisers.

This complex market configuration is enabled by the stabilisation of cookie-based and other online user tracking mechanisms. Cookies are "small text files that sites place on a user's computer to identify the user's browser, computer operating system, IP address, and (if the user provides it) personal information" (Gehl, 2014, p. 105). They were originally developed to enable e-commerce sites to remember and record multiple items selected for purchase by a user in a virtual shopping cart (West, 2017). In the audience marketplace,

cookies enable publishers to keep track of visitors to their websites and enable ad networks to track users' browsing behaviour across website. Today it is common practice for websites to be paid for placement of third-party cookies on their websites to enable users to be tracked across the web (West, 2017). While the cookie is perhaps the most well-known web tracking device, web tracking can be variously implemented, including through web beacons or bugs, invisible one-by-one pixel graphic images that enable sites to transfer user data to third-parties such as ad networks, and the more persistent "flash cookies" (Gehl, 2014; West, 2017).

Such web tracking mechanisms have increasingly supported an orientation in the post-exposure audience marketplace towards behavioural targeting, highly controversial due to its implications for individual privacy online (see, e.g., Brotherton, 2012; Zuiderveen Borgesius, 2014). This advertising technique involves the "delivery of targeted advertising to different members of the audience based on their demonstrated patterns of media consumption or behavioral responses such as information requests and other possible advertisement responses, such as click-throughs or product purchases" (Napoli, 2011, p. 111). This approach shifts ad targeting from a focus on publications, to a focus on those individual profiles whose characteristics are highly valued by advertisers (Bakir & McStay, 2018).

From the point of view of the media's audience conceptions, these developments have been understood as an increasing "rationalisation of audience understanding" (Anderson, 2011a, p. 553), whereby audience perceptions have become "increasingly scientific and increasingly data-driven, with more impressionistic or instinctive approaches to audience understanding increasingly falling by the wayside" (Napoli, 2011, p. 11). More recently a *Digital Journalism* special issue captures this shift in perceptions with the notion of "measurable journalism", described as a "term that encapsulates the cultural and material shift to digital platforms capable of providing real-time, individualizable, quantitative data about audience consumption practices" (Carlson, 2018, p. 409).

The post-exposure audience marketplace where behavioural targeting techniques play a central role marks a shift not just towards more scientific conceptions of audiences, which have underpinned this market from the beginning, but towards *particular scientific visions* inscribed in audience measurement systems active in this marketplace. These rearticulate audience representation from “statistical abstractions” (Napoli, 2003) or portraits of audiences based on broad descriptive categories of shared characteristics, towards a form of “radical behavioralism that calculates society without representing it” (Cardon, 2016, p. 104). The latter focuses on traces that differentiate individuals not for the purposes of representation but in order to act on and influence their behaviour.

In this chapter I argue that the web tracking infrastructures that materialise these rearticulations of audience understanding can be repurposed to explore in more detail the post-exposure audience marketplaces in which different forms of digital cultural production are embedded.

5.2 A News Device Approach to Audience Marketplaces

5.2.1 Web Trackers as Digital Objects

As discussed briefly in the previous section, in the context of advertising-supported digital cultural production, digital objects such as cookies and other web tracking devices play an increasingly prominent role in audience measurement and other practices that make up the audience marketplace. As snippets of third-party code to be found in the source code of websites, trackers form an invisible “data mining infrastructure” whose role is to establish connections between websites and various third-party services, and to enable data flows between them (Gerlitz & Helmond, 2013). Mayer & Mitchell (2012) see web tracking as part of “the increasing trend of third-party websites recording and analyzing users’ browsing activities across unrelated first-party websites” (p. 1). Roesner, Kohno, & Wetherall (2012) give the example of “a website (like doubleclick.net) that has its tracking code included or embedded

in another site (like cnn.com)” (p. 1). Embedded in websites such as cnn.com, data flows enabled by such trackers may include recording visits to the website, extracting various kinds of user data, combining it with data from other databases, serving ads and measuring ad effectiveness. Trackers are not just invisible devices but they are also *dynamic* or *lively*, in the sense that user activities shape the data flows which are being initiated and their intensity (Gerlitz & Helmond, 2013). From this point of view, they should be understood not just as snippets of code in the source code of websites or as files on users’ computers in the case of cookies, but as mechanisms that enable communication between the website, the user and third parties (van der Velden, 2018).

The attachment of this *invisible and dynamic “data mining infrastructure”* (Gerlitz & Helmond, 2013) to a website is linked to the changing nature of the websites, from being “designed and hosted by a single person, group, or organization” to being “increasingly composed of content from myriad unrelated ‘third-party’ websites in the business of advertising, analytics, social networking, and more” (Mayer & Mitchell, 2012, p. 1). Similarly, Helmond (2017) captures this transformation of websites in the social media age from “self-contained units” to “assembled units”:

The website can be seen as an assemblage of modular elements that on the one hand enable interactions with other actors on the web and on the other hand permeate or redraw the boundaries of the website by setting up data channels for the exchange of content and data stored in external databases. (p. 6)

These controversial web tracking devices and the associated data mining practices that they enable are extensively studied and problematised in relation to a number of issues: online surveillance, privacy and security concerns, and related policy challenges (Binns et al., 2018; Englehardt & Narayanan, 2016; Libert, 2015; Mayer & Mitchell, 2012; Soltani, Canty, Mayo, Thomas, & Hoofnagle, 2009; van der Velden, 2014), web economies, expanding data industries, data capitalism and associated issues of uncompensated digital labour and power asymmetries (Andrejevic, 2014; Gerlitz & Helmond, 2013;

West, 2017), web history (Helmond, 2017), forms of discrimination and bias that may emerge from the processing of such traces (Barocas & Selbst, 2014), comparisons of tracking practices across domains and countries (Castelluccia, Grumbach, & Olejnik, 2013; Deville & van der Velden, 2016; Libert & Nielsen, 2018), as well as practices of resistance in these tracking environments (Brunton & Nissenbaum, 2013). This diversity of issues attached to trackers and tracking practices reflects their highly contested and “multi-valent” (Marres, 2011) nature, in the sense that a number of different spheres are co-articulated through them, from audience economics, to innovation in ad and information delivery (e.g. personalisation of both news and ads, recommendations), editorial decision-making and consumer surveillance. In this chapter however I focus primarily on web tracking as a device for understanding audience marketplaces.

5.2.2 Trackers as Audience Marketplace Devices

Many (although not all) of the trackers part of the invisible data mining infrastructure of advertising-supported websites are associated with the actors and practices of the audience marketplace. This fact draws attention to how the participation of news publishers in the audience marketplace is predicated on and shapes the very material infrastructure of news, the website. This changing character of websites, combined with infrastructural approaches in new media and internet studies (Helmond, 2015b; Plantin et al., 2016), may open up new research possibilities in the area of news audience economics (Napoli, 2003). These would see the news websites be treated as an object not only for the study of various aspects of digital news production, distribution and recommendation (see, e.g., Boczkowski, 2004a; Bødker & Brügger, 2018; Matheson, 2004; Karlsson, 2010; Stroud, Scacco, & Curry, 2016), but also for the study of particular aspects of the making of audiences into economically valuable products, such as the actor composition of audience marketplaces of a website and its evolution over time, comparative studies of audience marketplace configurations across different types of advertising-supported cultural production, and the relationships between news institutions and other

industries active in the audience marketplace. Audience marketplace configurations in this context can be understood as assemblages of actors that participate in audience commodification through tracking devices that facilitate tracking, calculating, analysing, evaluating and monetising various kinds of user data and online activities.

Drawing on Callon & Muniesa (2005), I conceptualise web trackers as online audience marketplace devices to draw attention to the work they do to materialise relations, exchanges and data flows between different participants in the online audience marketplace.⁶⁰ Participants include users, digital cultural producers, advertisers and advertising-related services, audience measurement companies and other data intermediaries. While audience marketplace devices may include traditional forms of measurement that have been migrated to digital environments such as surveys and focus groups, in this chapter I make a contribution towards understanding “natively digital” audience marketplace devices, i.e. those that are specific to the web medium as opposed to those that emulate more established methods of measurement (for more on this distinction see, Rogers, 2014). Trackers can be seen as audience marketplace devices not just in the sense of mechanisms that facilitate transactions and exchanges between different participants in the audience marketplace (e.g. capture of user data, delivery of advertisement, delivery of attention to advertiser, etc.), but also in the sense of acting as “algorithmic calculative devices”, which make various forms of algorithmic calculation possible (Amoore & Piotukh, 2016, p. 18; see also Callon & Muniesa, 2005). They do so by making user activities calculable by transforming them (e.g. the user gaze) into “locatable objects” (Introna, 2016) which can be detached, put in relation to other objects, calculated and assigned value (e.g. as impressions and currencies), and thus made economically valuable.

According to Callon & Muniesa (2005), to capture different aspects of economic calculations and their consequences, the researcher can take as a starting point the calculative agents who are active in the marketplace, the

⁶⁰ I am grateful to Tommaso Venturini for pointing me towards the work of Michel Callon and Fabian Muniesa on markets as calculative devices in this context.

construction of goods (in the case of my chapter, the institutionalised audience), or the calculated exchanges (in my case, the exchanges consist of the delivery of user attention and data to the content provider and to the advertiser), and I would add, controversies around currencies or coins of exchange in this marketplace (e.g. cost-per-click and cost-per-action).

In this chapter I take the first aspect, pertaining to calculative agents, as a starting point. More specifically I take the tracking practices of two forms of advertising supported digital cultural production, news and junk news sites, as a way to examine the audience marketplace configurations, third-party actor compositions and relations specific to various forms of digital cultural production. In focusing on the composition of the online audience marketplace, I also follow political economy researchers who have drawn attention to the fact that media critique focused on aspects of political economy should complement interrogation of scientific techniques and methods for audience construction (usually studied around the most well-known audience measurement firms, such as Nielsen), with studies that pay attention not just to a few leading actors but to *“the many systems available in the market”* (emphasis mine), i.e. studies of industries active in audience construction, and the relations between actors in these industries (Meehan, 1984, p. 218).

An approach that does not reduce the implications of tracking infrastructures to consequences for journalism practice seems particularly important in the case of news website tracking, because these sites have repeatedly been found to be one of the website categories with the largest volumes of online tracking activity (Englehardt & Narayanan, 2016; Libert & Nielsen, 2018). In a recent study focused on news sites in several European countries, Libert and Nielsen (2018) found that the volume of third-party tracking domains was consistently higher for news sites than for any other popular sites across the countries in the study. These tracking practices have significant implications outside journalism as well, as tracking activities facilitated by these sites feed the growing and controversial data industry (see, e.g., West, 2017).

Finally, the focus on arrangements of tracking agents is also important in addressing political questions pertaining to who tracks and analyses, and the implications of these activities. According to Callon and Muniesa (2005) calculative processes “are all costly activities that raise the question of calculative power” and of “asymmetries of calculation”, linked to the uneven distribution of calculative equipment (p. 1232). Similarly, in the case of web tracking and related data mining processes, Andrejevik (2014) has drawn attention to the asymmetries of participation in online data mining between “sorters” and “sortees”, i.e. “between those who collect, store, and mine large quantities of data, and those whom data collection targets” (p. 1673).

5.3 Studying Audience Marketplaces With Tracker

Signatures

My method consists of tracing the relationships between forms of digital cultural production and audience marketplace actors by examining the third-party tracking mechanisms embedded in websites and visible in their source code (drawing on an approach documented in Helmond, 2017). While in the audience research industry tracking activities have begun to be used to rank and evaluate media company performance as an indicator of their capacities to capitalise their audiences (Napoli, 2011), in this chapter I aim to rework tracking activities away from such quantitative evaluative practices and towards a qualitative exploration of tracking practices. This includes paying attention to the relationships between changing website infrastructures, the technicity of trackers and their cultures of use across different forms of digital cultural production, as well as legal considerations with regard to third-party tracking.

For this analysis I take as a starting point a corpus nineteen English language mainstream news sites and nineteen junk news sites whose election stories were found to receive high Facebook engagement scores in the months before the 2016 US presidential elections (Silverman, 2016).

Following Helmond’s (2017) suggestion that different page types might be

inscribed with different trackers, I include in my corpus the homepage as well as a selection of five popular article pages for each site, resulting in a corpus of 228 URLs. To identify the articles that receive the highest social media engagement scores I use the social analytics tool BuzzSumo.⁶¹ I focus on the tracking practices of popular sites and pages because the interest (and concern) in “fake news” was prompted by the viral character of these stories, as well as because popular sites tend to be richer in tracking elements (Macbeth, 2017). While tracking-poor sites are also relevant to examine, for the purposes of this chapter I focus on forms of digital cultural production that are tracking-rich. The articles included in the corpus span mostly 2016 and 2017, with a few junk news articles dating back to 2015.

To explore the third-party tracking networks associated with these forms of digital cultural production, I use the Tracker Tracker (Digital Methods Initiative, n.d.-b), a research tool that identifies third-party data-tracking technologies in a given collection of websites. The URLs were run through the Tracker Tracker tool in March 2017.⁶² Given the dynamic character of tracking, this discussion thus reflects third-party tracking domains loaded by the website at that time for a user who accesses the websites from Europe. Convenient as it might be for the researcher, the phenomena examined in this chapter are by far not stable objects of study. Junk news sites active around the 2016 US presidential elections are ephemeral constructions and due to the ensuing backlash many of them were retired not long after their stories became viral. Moreover, their tracking practices are also fluctuating, as responses to their viral character include blacklisting some of them from ad networks. Hence the audience marketplace configurations discussed in this chapter should be seen not only as economic-material arrangements but also as the outcome of public pressure and policy interventions. The instruments used to study them also shape the resulting picture of the phenomenon, as I will discuss next.

61 <http://buzzsumo.com/>

62 Preliminary work that informs this chapter has been conducted during two collaborative research projects which I led at the Digital Methods Initiative Winter School in January 2017 and at the Fake News Sprint in March 2017, both hosted at the University of Amsterdam.

5.3.1 The Tracker Tracker as a Research Device

Prompted by Brunton and Coleman's (2014) remark that a good research narrative includes not just a picture of the phenomenon as seen through the results of its study, but also the picture's "own blind spots, occlusions, and *range of focus*" (p. 79, emphasis mine), in what follows I will discuss a few aspects of my method that I think deserve further attention, with particular emphasis on the calibration mechanisms that make up the range of focus of this study.

As a media and social research tool, the Tracker Tracker repurposes the detection and classification capacities of the popular Ghostery privacy protection browser extension. In doing so, as discussed in Chapter 2, it is situated amongst device approaches to digital media research that seek to repurpose the analytical capabilities of existing digital services such as search engines and platforms (Rogers, 2013; Weltevrede, 2016). More specifically it is associated with a more recent phase of internet studies which aims to shift from studies of hyperlinks and single social media platforms to cross-platform analysis (Helmond, 2017; Rogers, 2018). In this case cross-platform analysis focuses on digital objects embedded in websites that create connections with third parties. Indeed, the Tracker Tracker tool has been used for a number of media research as well as journalistic projects (see, e.g., Deville & van der Velden, 2016; Gerlitz & Helmond, 2013; Helmond, 2017; Silverman, Singer-Vine, & Vo, 2017; van der Velden, 2014). Such approaches may also be thought of as "inventive methods", in the sense that they are characterised by a "variety and variability of purpose" (Lury & Wakeford, 2012, p. 5) and they may be put to use for a number of different research and other purposes.

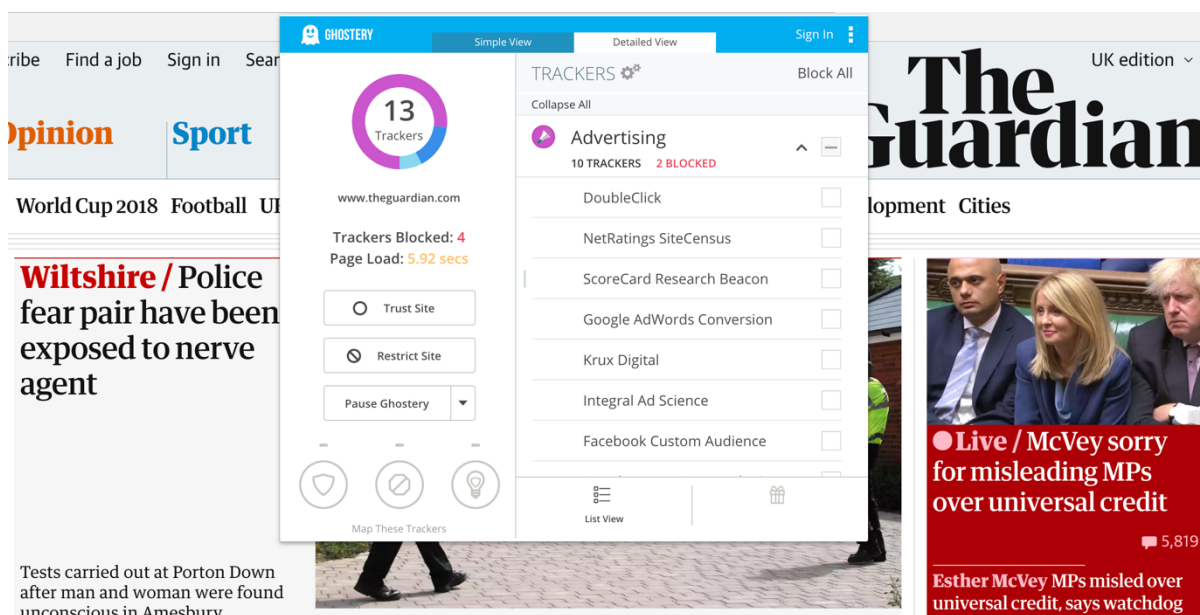


Figure 23: Screenshot of the Ghostery browser plugin visualisation of tracking elements active on The Guardian's homepage.

As a research device, the Tracker Tracker tool channels Ghostery's tracker detection and classification capabilities to enable researchers to extract third-party tracking elements for a single or a set of URLs in table or network format. While web tracking can be set up by first parties too (the website the user is voluntarily visiting), Ghostery focuses on third-party tracking elements embedded in first party websites as these are domains that the user connects to involuntarily.

Ghostery maintains a database of over two thousand trackers and over four thousand scripts associated with these trackers,⁶³ which it describes as one of the largest tracker databases.⁶⁴ Each tracker in the database has a URL and a profile.⁶⁵ The service matches requests to third parties sent during a page load against patterns, scripts or code snippets in the Ghostery database (e.g. `\".googlesyndication.com/simgad/|\".googlesyndication.com/pagead/|partner.googleadservices.com/gampad/` for Google AdSense). This allows the service to identify third-party tracking elements with which the page establishes connections (e.g. AdSense) and the companies that own them (e.g.

63 <https://www.ghostery.com/submit-a-tracker/>

64 <https://ghostery.zendesk.com/hc/en-us/articles/115000734653-View-all-Trackers>

65 E.g. see the tracker profile for the ad network Google AdSense: https://apps.ghostery.com/apps/google_adsense.

Google). A third-party element can have one or multiple scripts embedded in a web page (and thus can establish one or more connections to a web page), and a company can own one or more third-party elements. Such is the example of Google who owns multiple audience marketplace services that make use of tracking elements, including AdSense, Google Analytics and DoubleClick.

The Tracker Tracker tool enhances Ghostery by enabling researchers to compare the presence of trackers across websites (van der Velden, 2018) and to study “tracker networks” associated with websites (Helmond, 2017). The analysis of tracking networks builds on earlier forms of repurposing of natively digital objects to examine how web pages are associated, valued and ranked in online cultures, through e.g. hyperlink analyses, URL-hashtag analyses in the case of Twitter (analyses of associations between URLs based on hashtags), URL-page analysis in the case of Facebook (analyses of associations between URLs based on Facebook pages that share them). The Tracker Tracker also repurposes the tracker classification that Ghostery produces for the purposes of raising individual awareness of surveillant technologies and providing privacy protecting technologies and services.⁶⁶ Given that Ghostery’s database, known as Ghostrank, is compiled based on tracker data collected from users who have opted to share information with the service (Macbeth, 2017), it means that Ghostery will be less effective at detecting less frequently used tracking elements (Englehardt & Narayanan, 2016).

The lively or dynamic character of tracking discussed in the previous section also shapes the study of tracking practices. Indeed, the data flows that are set in motion and captured for analysis depend on a number of things. This includes the location of the user who accesses the site. For example, the BBC does not serve ads for users visiting its site from the UK but does serve ads for users visiting it from outside the UK (BBC, n.d.). Other studies have noted that additional tracking may be activated upon user activities such as clicking on an ad or a social widget (Roesner et al., 2012). Advertising networks also contribute to the fluctuation of trackers loaded on a page, as they load ads (and

66 In the course of my doctoral research Ghostery was acquired by Cliqz, a German company producing anti-tracking technologies and services. See: <https://www.ghostery.com/blog/ghostery-news/ghostery-acquired-cliqz/>

associated tracking elements) from various suppliers who bid to display an ad based on the profile of the user who visits the website (Macbeth, 2017). This means that different page loads may result in different trackers being detected (Libert, 2018). Finally, certain types of third-party trackers such as ad networks act as hosts or aggregators for several other third-party tracking elements which will not be directly embedded in the first-party website but are requested or referred to by embedded trackers (Roesner et al., 2012). From the point of view of the publisher, such behaviour results in opacity, as publishers might not be aware of all third parties that access user data through their site (Joseph, 2018).

Finally, the configuration of the Tracker Tracker tool itself shapes the detection of tracking practices. Pages are loaded through a PhantomJS browser without a user interface, which means that any prompts for user input are ignored, which may impact the trackers loaded on the page. This would include prompts to accept cookie policies and to log into a website to be able to access its contents.⁶⁷ For this reason this technique may be seen to come closer to other lower bound methods likely to under-detect tracking activities (such as those used by Englehardt & Narayanan, 2016; Libert, 2015; Roesner et al., 2012).

5.4 A Visual Network Exploration of Audience Marketplace Configurations From the Perspective of Web Tracking

The network files generated by the Tracker Tracker tool for the lists of junk and mainstream news sites (homepages and articles) are uploaded and appended in Gephi,⁶⁸ a visual network exploration tool which is widely used in digital humanities, digital social research and other fields. This operation results into a network of 504 nodes and 5,897 connections. Of these, 219 are first-party URLs⁶⁹ (113 pertaining to junk news sites and 106 to mainstream news

67 <https://tools.digitalmethods.net/beta/trackerTracker/>

68 <https://gephi.org/>

69 The Tracker Tracker tool returned no output for a few URLs, particularly homepages of mainstream news sites, which reduced the original collection of first-party URLs to 219.

sites), and 285 are tracking elements, which makes the network a bipartite network. While relations between junk and mainstream news have previously been considered, among other things, in terms of their comparative valuation on social media in terms of engagement (Silverman, 2016), the comparative speed of their spread on Twitter (Vosoughi, Roy, & Aral, 2018), their comparative degree of reliability (Les Décodeurs, 2017) and their linking behaviours (Venturini et al., 2018), in this chapter I explore connections between them based on their invisible tracking and data mining infrastructures.

Before moving into visually exploring the network in more detail, a first observation can be made based on the volume of tracking elements and connections, which points towards the *distributed nature of audience commodification*. Indeed, while these practices have always been a collective accomplishment, this analysis reveals it as accomplishment distributed across an increasingly large number of inter-dependent actors. Overall there are close to 300 tracking elements in the network that receive close to 6,000 connections from first-party websites. Audience construction and monetisation is not enacted through standalone tracking activities but it is distributed across inter-dependent tracking elements, which often collect multiple user data points across multiple websites and communicate and exchange data with each other. Indeed, partnership programmes are a common feature in the online advertising industry (see, e.g., Helmond et al., 2017). For this reason, it is important to explore relations between actors in this market and to examine audience commodification not as the outcome of one audience measurement systems but as the outcome of interactions between multiple such systems. As mentioned in section 4.1, in this increasingly complex configuration in which exchanges and activities are increasingly managed programmatically, we are not only faced with an increasing “rationalisation” of ways in which the news media understands its audience (Napoli, 2011), but also with an intensification of audience measurement and analysis, and with particular forms of rationalisation that draw on “radical behaviouralism” (Cardon, 2016).

Visual network exploration (Venturini et al., 2015), is particularly suited for the qualitative exploratory analysis of tracking practices grounded in the empirical

world because it allows the researcher to examine individual actors and associations between them at a disaggregated level. While visual network exploration has seen a “renaissance” over the past decades, early key figures of social network analysis such as Moreno (1977) also pointed towards the importance of network visualisation as a method of exploration in social research (on the visual exploration of networks see also Venturini et al., 2017, 2018; on the methodological reflexivity required when using data visualisation as an analytical device, see Gray, Bounegru, Milan, & Ciuccarelli, 2016).

To visually explore the audience marketplaces in which different forms of advertising-supported digital cultural production are embedded, I use a “force-directed spatialization” technique which “simulates a physical system in order to spatialise a network. Nodes repulse each other like charged particles, while edges attract their nodes, like springs” (Jacomy, Venturini, Heymann, & Bastian, 2014, p. 2). This way, the disposition of the nodes in the space of this network representation gains meaning and can be interpreted as “a proxy of their structural similarity: two nodes being the closer the more directly or indirectly they are connected” (Venturini et al., p. 4). The outcome of this technique is the visual representation of relations between trackers and websites in the form of a network composed of regions of various node and edge densities (also known as clusters) separated by empty or sparsely populated areas known as structural holes (Burt, 1992). Clustering and the structural holes that separate them are illustrative of asymmetrical associations between actors in the network.

This technique is particularly useful for the visual exploration of tracking practices because its resulting clustering is similar to the outcomes of cluster detection or community detection techniques (Noack, 2009). Drawing on earlier work by colleagues and myself, I see visual network exploration as an iterative process involving “a constant toing and froing of categorization and observation, typology and topology” (Venturini et al., 2018, p. 269).

Given the (relatively) large number of nodes in the network for a qualitative analysis, to facilitate exploration I size nodes according to their number of

connections. Because the network is directed, i.e. links have directionality (in this case from third-party trackers pointing to websites), they can be sized by in- or out-degree, i.e. the number of connections they receive and the number of connections they send, respectively. Which measurement is relevant depends on the aims of the research. If the focus is on investigating individual news publishers and the connections that they establish with third parties, then nodes can be sized by in-degree to highlight the publishers which receive most connections from third-party trackers. In the case of my research, the question does not pertain to tracker ecologies at the level of the individual websites. For this type of analysis to yield interesting results, different website selection criteria than the ones used for this study might be better suited. Instead, my aim is to explore tracking practices across different forms of digital cultural production. Because of this interest, I size nodes by out-degree in order to be able to interpret the prominence of trackers across these information spheres.

These operations having been completed, the reader may notice from Figure 24 below that the network is not divided into two equally sized regions, which we could be expecting given that our starting points were two equal sets of URLs. Like many real-world networks, the layout of the audience marketplace network at first sight does not seem to present a distribution of nodes into any number of neatly separated clusters. Instead, the first characteristic we may notice is a large component at the bottom and a small one at the top which suggests an uneven distribution of associations between nodes, in need of further exploration.

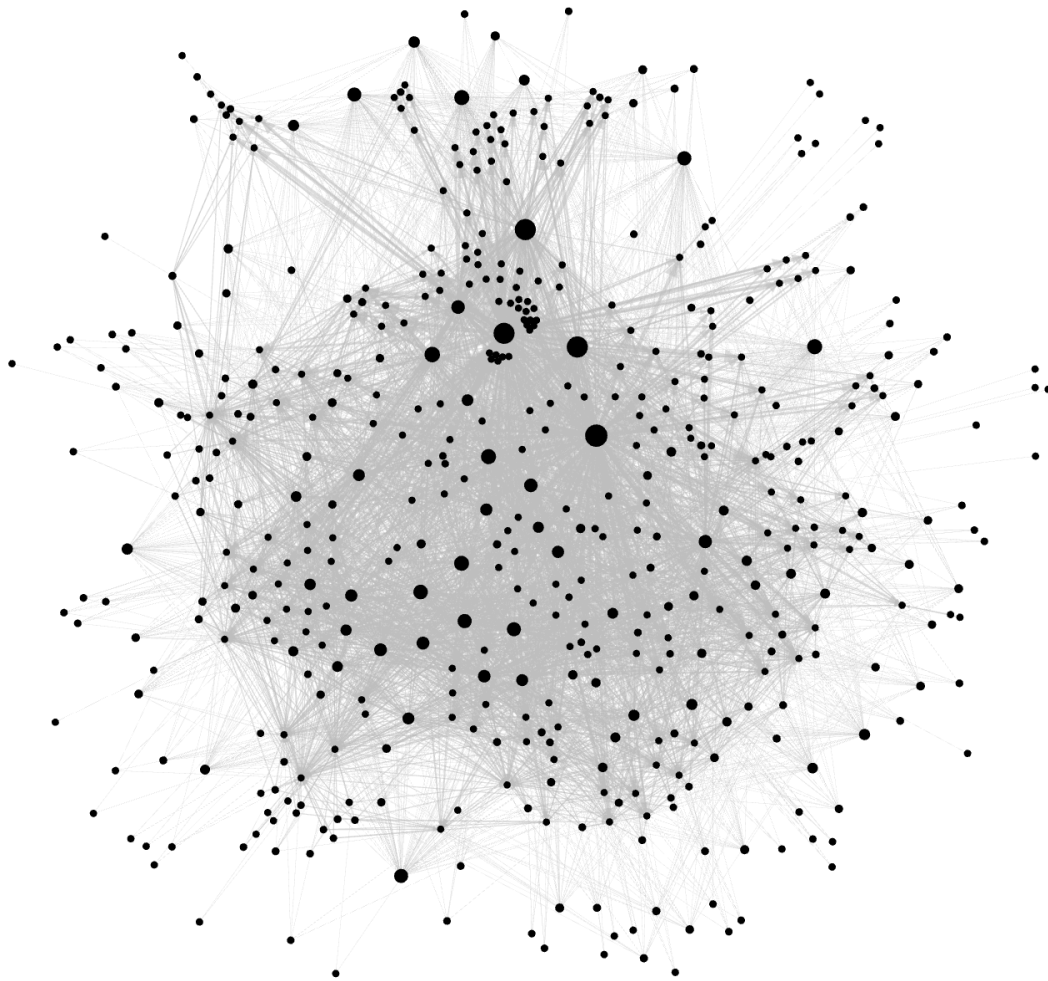


Figure 24: Mainstream and junk news sites and their third-party tracking networks spatialised with a force-directed layout algorithm in order to illustrate structural similarities visually, as node proximities.

In this case we could assume that this particular spatialisation of the network is indicative of differences in tracking practices associated with the two different types of URLs which were the starting points of this analysis: mainstream news and junk news. For the time being the original binary classification of URLs in the corpus will be used. As we will see, the classification will be revisited and enriched later in order to support the network exploration.

And indeed, if we colour nodes of the first-party domain type by the kind of websites they belong to and highlight outlinks sent by third-party domains by the type of their target first-party sites, we see that this website classification can explain the structure of the network, but it can do so only partially (see

Figure 25). While the majority of the URLs in the small component seem to be junk news URLs, the big component is divided between junk and mainstream news URLs.

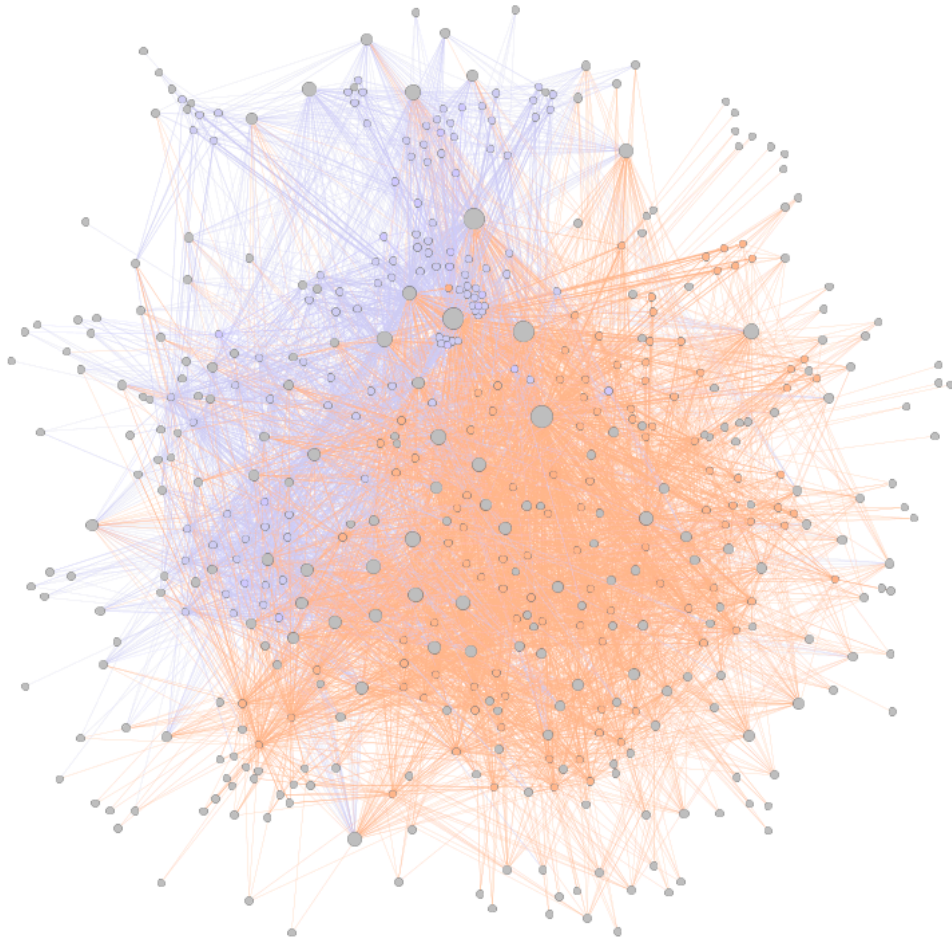


Figure 25: Mainstream and junk news sites and their third-party tracking networks, where junk news URLs are represented as purple nodes and mainstream news URLs are represented as orange nodes. Edges that link first-parties to third-parties are coloured by the first-party domain type.

Nevertheless, there are things we can learn from this analytical operation. Overall third-party trackers (of which there are a total of 285) receive more connections from mainstream news sites (3,763) than from junk news sites (2,134). News sites also connect to a larger number of unique tracking elements (243) than junk news sites (144). Mainstream news websites thus appear to be both more intrusive and more diverse in their tracking practices.

But, as I will return to later in this section, this may be taken as indicative of the frailty of business models and revenue streams of advertising-supported news media, increasingly pressured to monetise all their inventory.

Returning to the network exploration, as the URL classification does not fully explain the spatialisation of the graph, further analysis of the graph structure is needed. A closer look at the node composition and the types of connections making up the large graph component, shows that it consists of both mainstream news and junk news URLs, thus pointing towards a network composed of not two regions (around mainstream and junk news sites), but of multiple regions and thus tracking styles. When examining the densities of nodes in this component more closely, at least three more regions become apparent in the large pole of the network, two organised primarily around junk news URLs and at least one organised primarily around mainstream news URLs. Gephi's community detection algorithm, applied as a node colour layer, seems to confirm these regions (see Figure 27).

To strengthen this topological interpretation, I experiment with further node categorisation and analysis. While so far the topological interpretation relied on the classification of URLs, could the typology of tracking elements add something to the interpretation of graph regions? To explore this question, I first turn to the typology of trackers that the Tracker Tracker tool outputs based on Ghostery's classificatory work. As Ghostery is regularly updating tracker categories, I update the network with the most recent tracker classification, which is the most relevant for the purposes of this study because it focuses on the purpose of the tracker. Ghostery's aim to increase individual awareness of services which access their data when they browse the internet, becomes immediately apparent in the tracker categories. These categories collapse the multiple services and functions at work in the audience marketplace into generic labels such as advertising, aimed at a non-specialist public. To test the robustness of this classification I manually check the tracker category against service descriptions on their own websites as well as other online source and in a few instances reclassify the tracker to reflect its main purpose and categorise tracking elements that do not receive a label from

Ghostery. In what follows I will discuss a few observations based on this analysis.

5.4.1 Advertising-Dominated Tracking Networks

Layering the Ghostery classification over the third-party services in this network representation suggests digital cultural production to be deeply entangled with the online advertising and marketing industries (see Figure 26). The audience marketplaces in which these forms of digital cultural production are embedded are dominated by advertising industry-related tracking, which make up over two thirds of the third-party domains in the network and are dispersed over all regions of the graph. Audience measurement instruments such as site analytics and social media tracking elements are following advertising industry-related tracking in terms of usage. This observation points towards the reliance of these information spheres on a great number of advertising related services, which may in turn be an indication of the precarity of these information spheres which need to resort to multiple services to monetise their inventory.

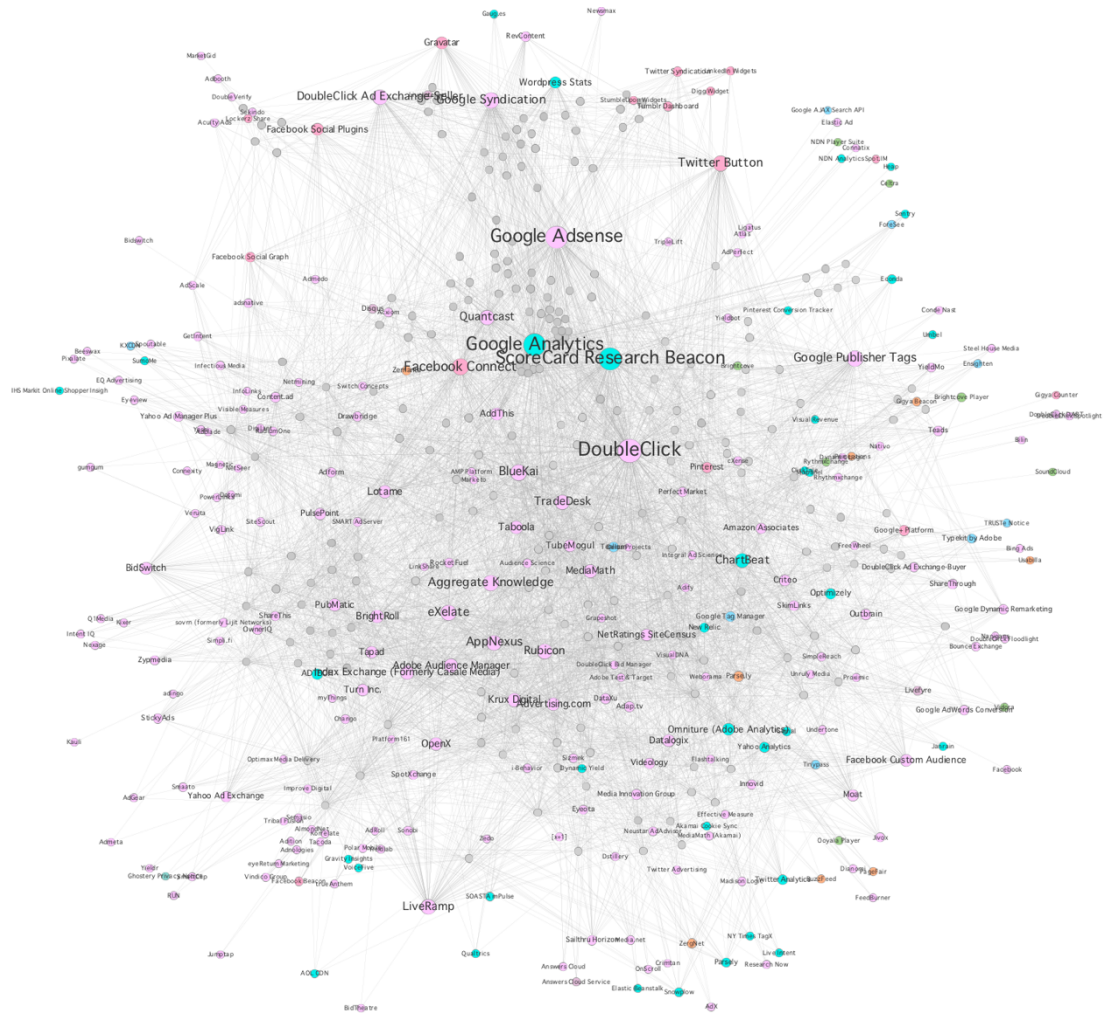


Figure 26: Mainstream and junk news sites and their third-party tracking networks, where third-party tracking elements are coloured by their type. Notable are the nodes in purple (representing advertising industry-related trackers), those in aqua representing in-site and cross-site analytics trackers and those in pink representing social media tracking elements.

Notable is also the bridging position of the in-site analytics service Google Analytics between junk news sites and mainstream news sites. This position indicates its centrality to both digital cultural production types. Indeed, Google Analytics has repeatedly been found to be one of the most widely used third-party domains across the web (Englehardt & Narayanan, 2016; Macbeth, 2017). But this external classification understandably does not correspond to the network structure and thus does not help to further elucidate the structure of the audience marketplace network, leaving space for further investigation.

To advance this analysis, I proceed by examining the nodes in each region in

more detail. I focus my examination of each region of the graph primarily on trackers associated with the audience marketplace and less on trackers essential to website functioning or to their privacy features.

As I have noted in another article co-authored with several colleagues, by now the iterative nature of such exploratory work should be apparent, as well as the importance of grounding classificatory work into network topology and letting network structure demarcation be informed by classificatory work:

It is important to notice that the operation of classifying the nodes and of reading the disposition of classes are not separated, but performed at the same time. As it will become clear in the next pages, our technique does not consist simply in the projection of a set of pre-existing categories on a connectivity-based layout, but on recursively using the categories to make sense of the layout and the layout to define the categories. (Venturini et al., 2018, p. 8)

Finally, one classification criterion might not be enough to make sense of the disposition of nodes in a network which depicts a complex empirical phenomenon, and certainly not in the case of bipartite networks where each node type asks to be treated according to its specificity. Hence multiple features of a first-party site or a third-party tracking element might need to be considered to make sense of the topology of the network: from genre and business model in the case of first-party sites, to the more specific role in the advertising industry and the audience marketplace for third-party tracking elements.

5.4.2 Variations in Tracking Styles and Audience Marketplace Configurations

The analysis of the key regions of the graph (see Figure 27) suggests that different forms of digital cultural production have their own infrastructures and practices for measuring, analysing, intensifying and monetising the activities of their users. In doing so it shows that it is not only tracking services

that shape news infrastructures and audience construction as discussed in section 4.2, but that technical artifacts and social arrangements are mutually constitutive, as different types of digital cultural production may develop their own tracking styles and practices.

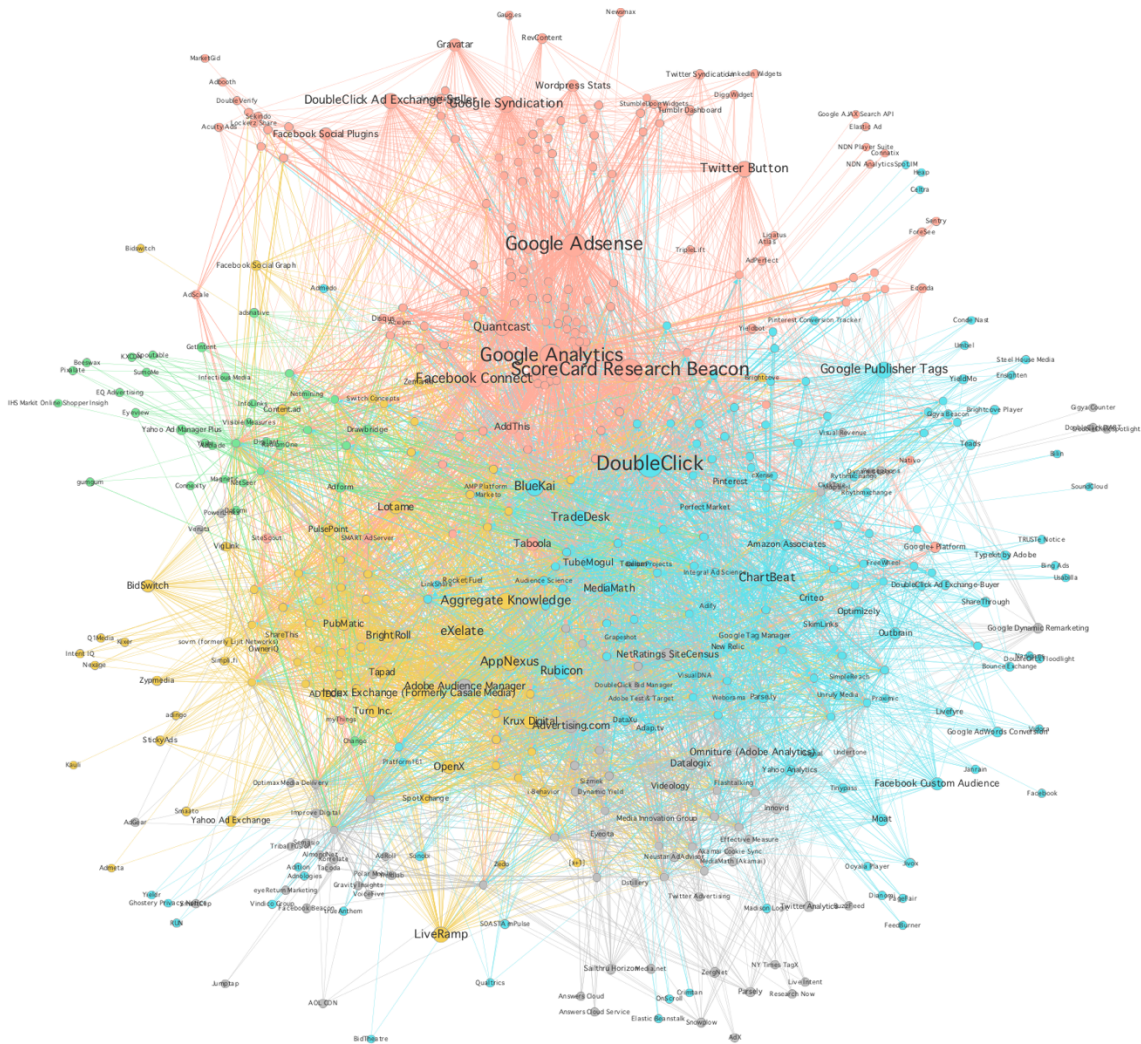


Figure 27: Mainstream and junk news sites and their audience marketplace configurations as seen through third-party tracking networks in which they are embedded. Graph regions or clusters are visually distinguished by colour.

In one of the largest regions of the graph (represented in aqua), we recognise a *professionalised audience marketplace configuration*, associated with high quality publishers (around half of the total number of mainstream news URLs), specialised services and a large number of intermediaries. The audience commodity assembled and exchanged in this space could be imagined as a “premium”, highly groomed product, an elaborate construction where users

are known through many data points, collected and combined from multiple sources due to the rather permissive tracking practices of publishers in this space. Indeed, what data points are actually collected by each tracker is certainly an important next step that the type of analysis illustrated in this chapter should explore in future studies.

Audience construction and monetisation in this region is distributed, diversified and intensified through scores of advertising industry intermediaries. These include major ad exchanges such as DoubleClick but also multiple demand-side platforms (e.g. TradeDesk, MediaMath, Madison Logic, TubeMogul and Data Xu). Demand-side platforms are used by advertisers and advertising agencies to automate the buying of targeted inventory from a number of sources. These also include supply-side platforms used by publishers to automate the sale of residual inventory which could not be sold directly to advertisers (such as Rubicon, Teads, Sonobi and Unruly Media). Alongside these we recognise several data brokers or data aggregators specialising in collecting, combining and selling audience data, such as BlueKai, Signal and AudienceScience. Their datasets help to enhance the audience commodity sold to advertisers. In this space audiences are not just measured to sell inventory but also in terms of click-through and conversion rates. This helps to measure ad effectiveness and is done through services such as Moat, Google AdWords Conversion, and Simple Reach, the latter of which specialises in social media.

In this space we also recognise industry-specific services such as the web analytics startup specialising in news publishers ChartBeat. Audiences are met with content personalisation and recommendation, both of news or marketing, through services such as Optimizely, Parse.ly, Perfect Market and Cxsense. The personalisation of content delivery (both news and ads) is a strategy used to retain visits and increase advertising revenue (Turow, 2011). Paywall system trackers (such as Tinypass) are an indication of the mixed-strategy business models of websites in this cluster. A number of website utilities such as tag or code managers for third-party trackers (Google Tag Manager, Tealium) and website testing and optimisation tools (New Relic), and fonts (Typekit by

Adobe) also indicate a professionalised content production space. Native advertising services (e.g. Outbrain, SkimLinks, ShareThrough) are also present as well as ad retargeting providers (e.g. Criterio), and trackers specialising in the advertising of Amazon products (Amazon Associates).

Compared to the professionalised audience marketplace configuration associated with mainstream news sites, the junk news region of the graph (represented in orange), depicts a *less professionalised audience marketplace configuration*. This configuration is perhaps specific to the long tail of the internet and to small scale, amateur-run publishing operations. The audience construction practices of these operations are also less professionalised and less distributed and seem more rudimentary. In this region we recognise the majority of junk news URLs (over 70% of them). With tracking styles similar to junk news URLs are a couple of mainstream news article URLs belonging to the Guardian and NY Daily News, both reliant on digital advertising for at least part of their funding.

In this space audience monetisation relies on fewer intermediary data and advertising services, and is prominently performed through services that have emerged to monetise user generated content and amateur digital content production or the long tail of the internet. These services include ad networks and ad exchanges, such as Google AdSense, DoubleClick Ad Exchange-Seller, SiteScout and AdScale.

Google AdSense is used by the majority of websites in the network, including by high quality publishers who employ it to monetise their residual inventory or less sell-able audiences, i.e. ad space which has not been directly bought by advertisers. This service uses contextual targeting, in the sense that ads are placed based on a match with a website's content (desilva + phillips, LLC, 2008). In this space we also recognise native advertising networks such as Taboola, RevContent and Nativio. These distribute advertiser sponsored content or "content ads", displayed on websites in the form of recommendations of articles related to those the user is currently viewing. Such services have been criticised for their distribution of false stories and clickbait,

i.e. content with sensational and misleading headlines, on mainstream news sites (Change Advertising, 2016; Griffith, 2014). They have also been criticised for facilitating ad revenue for sites that distribute misleading information (Moses, 2016).

Notable is also the presence of several tracking elements associated with social media, social sharing and social bookmarking platforms. Their presence attests to the platformisation or the expansion of social media audience amplification and measurement mechanisms into junk news spheres, and also to their centrality in attracting traffic to these websites to be monetised through advertising. Such services include Facebook Connect, a service that allows users to connect to third-party websites using their Facebook identity. This service bridges between this junk news region and other regions of the graph. Other social media and social bookmarking services prominent in this region include the Twitter Button, AddThis, Digg, StumbleUpon, LockerzShare, Tumblr, LinkedIn and Google+. The presence of the WordPress blogging platform analytics tracker, Stats, and of the Gravatar service which is integrated with Wordpress also attest to the more amateur character of junk news publishing operations.

Besides Facebook Connect, through their position in the graph we recognise another two audience measurement services in particular as bridging between the different clusters in the network. These are the site-centric traffic analytics service Google Analytics (present in the majority of websites in the network but on more junk news sites than on mainstream news sites) and the cross-site user-centric analytics service ScoreCard Research (also present on the majority of sites), which records audience navigation patterns across websites. The latter is used to generate website rankings which inform decisions about advertising rates and the two are often used in combination (Cardon, 2016). We also recognise the measurement company Quantcast which offers its services to less popular sites which do not perform well in more established website rankings such as Nielsen or comScore (Turow, 2011).

A third large region of the graph (represented in yellow) assembles mainstream

news URLs whose tracking styles appear to resemble at least in part those of junk news URLs. On the junk news side we recognise a mix of satirical and entertainment websites. On the mainstream news side we recognise a public radio network funded through a mix programming fees, grants, sponsorship and advertising, and a political news website funded through subscriptions, advertising and events.

What seems to be specific to audience marketplace configuration in this region is *the prominence of data brokers*, also known as data management platforms or data aggregators. Such services include Aggregate Knowledge, eXelate (owned by Nielsen), LiveRamp and Krux Digital. The role of these data intermediaries is to store, combine, analyse and segment audience data from multiple sources. A particular type stands out, data onboarding services such as LiveRamp, which bring offline data (from customer databases, loyalty programmes or subscriptions), often including personally identifiable information, into online campaigns (Scudder & Wiener, 2012; Joe, 2015). For example, the data collected through Politico's subscription system, one of the mainstream news sites in this cluster, may constitute such a source of offline data. Such data aggregation services inform ad targeting on the advertiser side, and on the publisher side they contribute to the construction of the audience profiles which publishers sell to advertisers (Marvin, 2016).

The audience construction practices of the junk news URLs associated with this audience marketplace configuration may be seen as more professionalised than those of the URLs in the second discussed region of the graph (represented in orange). This is because the use of data aggregators enables publishers to produce more detailed knowledge about audiences to support the sale of inventory. We also recognise video advertising as prominent in this audience marketplace configuration through services such as SpotXchange, StickyAds and Zypmedia. Finally, VigLinks, a content monetisation service that enables the monetisation of referral links, seems to be specific to junk news sites (Khan, 2012).

A final small but notable region (represented in green) consists of a satirical

website (satiratribune.com) with a more *unique tracking profile*. The profile is rather international given its small size, featuring the Russian demand-side platform GetIntent, but also Swiss, Danish and Canadian alongside US services.

Given the fluctuations in tracking practices of junk news sites noted in section 4.3, the network topology discussed above may also be interpreted temporally and indicate distinctions between sites that have been suspended from the services major ad networks and have been pressured to diversify their monetisation strategies.

5.4.3 Asymmetries of Participation in the Audience Marketplace

Another issue that this analysis points towards is that of asymmetries of participation in the audience marketplace. One way in which this is manifested is through concentration or monopolistic tendencies resulting from the platformisation or extension of big platforms in the space of audience measurement and monetisation. This is a tendency which the audience measurement industry has always presented (Napoli, 2011). This phenomenon sees a small number of companies providing services to a large number of websites in the network (on this point see also Libert & Nielsen, 2018). As this analysis has shown, big online platforms such as Google, Facebook and Twitter have come to reshape not just news production and circulation but also audience commodification, and to dominate the online audience marketplace. Three Google services, DoubleClick, Google Analytics and Google AdSense are each present on over 80% of the URLs in the network. Facebook Connect is present on over half of the URLs in the network and the Twitter button on over 40% of them. Another service present on the majority of URLs is ScoreCard Research (over 80%). Although focusing primarily on audience analytics to support editorial decision-making, the presence of the startup ChartBeat alongside the big platforms on over 50% of URLs, is worth noting. Media concentration is also geographically circumscribed around the US.

While the notion of media concentration points towards asymmetries between service providers in the audience marketplace, another asymmetry pertains to the relationship of news publishers and the digital advertising industry. It is striking that advertising revenue has been in decline for news publishers, given the scale at which news sites have become vehicles for intrusive digital advertising and data industry tracking and data collection practices (over two thirds of all tracking elements in this study are advertising-related). For this reason, the tracking practices of mainstream news sites should be addressed not only in relation to user privacy and data protection concerns as initiatives such as the EU General Data Protection Regulation aimed to do, but as Englehardt & Narayanan (2016) also gesture towards, they should also be seen as an indication of the precarity of business models and revenue streams of news organisations which place increasing pressure on monetising content with advertising. While regulations such as the GDPR are aimed at increasing transparency and accountability on the audience - data collector and processor side of the advertising market, the relation between publishers and advertisers and the pressures that the business practices of the online advertising and digital marketing industries place on publishers would also need to be attended to if aggressive online data collection practices and privacy invasion are to be tempered. Indeed, absence of transparency and accountability on this side of the market is of concern as well, and publishers have repeatedly expressed their concern about the low returns they receive from selling online advertising inventory and the lack of transparency and accountability that a highly complex intermediation structure generates (Davies, 2016; Pidgeon, 2016).

Finally, the last aspect of this asymmetry pertains to the publisher-audience relation. While audiences are typically seen to pay for news in two ways, with the time spent reading the news and sometimes by buying access to news content with money (Nielsen, 2016), this analysis reminds us of a third less acknowledged form of payment that audiences make in news, namely their data. While the relationship between news and its publics is typically discussed as citizen participation in democratic societies through news or as audience participation in journalism through interactivity features of digital media (for a

discussion of this see, e.g., Peters & Witschge, 2015), this analysis reminds us that news and its institutions have always been co-produced with publics. Audiences have always actively participated in news through the provision of their data which informs multiple aspects of news, from the financial stability of institutions, to editorial decisions and the authority of a news publication, even if professional claims to journalistic authority do not allow these contributions to be recognised as such.

5.5 Conclusion

This chapter tested the news device approach in the context of another area of news, the marketplaces through which audience products are made and exchanged. In doing so it can be seen as a contribution to socio-material approaches to news work and their calls for bringing such approaches to bear on the business side of news (see, e.g., Lewis & Westlund, 2015). By studying the audience marketplace configurations of news websites, and more specifically of news pages that attract high social media engagement scores, in relation to those of junk viral news sites, it also aims to contribute to debates about fake news and the economics of advertising-supported viral content production.

The approach developed in this chapter explores the capacities of the website as a site for studying the business side of news. More specifically it focuses on the third-party user tracking mechanisms embedded in websites and their participation in the production and exchange of audience products. By combining audience economics with an infrastructural or material approach to new media studies, I illustrated an approach by means of which such digital objects can be used to study the composition and relations that make up the post-exposure audience marketplace configurations in which advertising-supported mainstream news and viral junk news sites operate.

In relation to the central question of this thesis, pertaining to how digital objects participate in and format news work and ways of studying it, this

chapter invited an understanding of how digital devices and their associated industries make a difference to news not only through the shaping of practices but also through the shaping of the infrastructures of news, in this case the website, through the integration of third-party elements.

Another contribution is made by extending the study of interactions between news and digital devices to also include interactions between these devices and knowledge creation. I describe how a tracking detection device, Ghostery, can be configured into a research device with the help of the Tracker Tracker research tool and analytical techniques developed to address the research question of this chapter. Such techniques enrich tracker detection at the level of an individual website with the analysis of associations between tracking elements based on their shared use by websites. I suggest that attention needs to be paid to how the research method shapes the research object and describe how the composition of audience marketplaces is configured by the dynamic character of tracking. In doing so I suggest that the study of audience marketplace configurations cannot be separated from the study of tracking practices, which shape the representations of audience marketplaces we arrive at.

A contribution is made to the study of audience marketplaces by providing an account from the viewpoint of tracking infrastructures of websites. While this may be seen as a partial view, and indeed it is, as any other representation, it may also be seen as enriching ways of knowing audience marketplace practices with another perspective, the view put forward by the tracking infrastructures of websites. From the point of view of tracking infrastructures of websites, junk news, and to a much greater degree mainstream news, appear to be deeply entangled with the complex structures of the online advertising and marketing industries. The economic dependence of news on the online advertising industry is manifested at the level of the material infrastructure of news, the news site, through an invisible tracking infrastructure, through which relations, exchanges and data flows are established between different participants in the online audience marketplace. This dense and invisible tracking infrastructure is problematic not only from the point of view of audiences in relation to issues

of user privacy, security and labour but, as far as news publishers are concerned, it is indicative of the precarity of business models and revenue streams of news organisations which place increasing pressure to sustain resource intensive news production through complex and invasive advertising structures, that rely on aggressive data collection practices, with little transparency and accountability towards other participants in the audience marketplace.

I illustrated a research technique by means of which specificities of tracking and audience marketplace practices can be surfaced and compared across digital cultural production. The analysis suggested that different forms of digital cultural production have their own practices and infrastructures for intensifying, measuring, analysing and monetising the activities of their audiences. This suggests that while the online advertising industry and its tracking infrastructures is shaping not just audience economics but also news infrastructures, tracking infrastructures are also shaped by cultural production dynamics. This study illustrated a few such audience marketplace arrangements, from amateur configurations associated with the internet long tail, to more complex configurations associated with professionalised forms of information production, such as the news media.

Asymmetries between participants in the audience marketplace cut across all these configurations, from monopolistic tendencies of big online platforms in the online advertising industry, to economic pressures on publishers which increasingly become vehicles for the aggressive data collection practices of advertising industry actors, and finally audience members who increasingly bear responsibility for the implications of these market configurations for their privacy and security, aided by data protection software and regulations (such as the recently enforced EU General Data Protection Regulation).

This analysis provided a picture of audience marketplace practices in which legal initiatives such as the recently enforced EU General Data Protection Regulation are aiming to intervene. Future research may use this analysis as a baseline to examine the changes that this regulation has brought to audience

marketplace configurations for users accessing publisher websites from Europe. Finally, while this analysis focused on display advertising, future research should also examine the audience marketplaces associated with mobile news sites as well as apps.