Digital News Consumption and Copyright Intervention: Evidence from Spain Before and After the 2015 "Link Tax"

Sílvia Majó-Vázquez[ORCID]

Reuters Institute for the Study of Journalism, Deparment of Politics and International Relations, University of Oxford, 13 Norham Gardens, Oxford, OX2 6PS, United Kingdom

Ana S. Cardenal[ORCID]

Deparment of Law and Politics, Universitat Oberta de Catalunya, Av. Carl Friedrich Gauss, 5 Parc Mediterrani de la Tecnologia, Edifici, B3 08860, Castelldefels, Spain

Sandra González-Bailón[ORCID]

Annenberg School for Communication, University of Pennsylvania, 3620 Walnut Street, Philadelphia, PA 19104-6220

We analyze patterns of digital news consumption before and after a "link tax" was introduced in Spain. This new legislation imposed a copyright fee for showing snippets of content created by newspapers and resulted in the shutdown of Google News Spain. The Spanish copyright law is a precedent to the Copyright Directive currently submitted to the European Parliament, which is planning to impose a similar "link tax." We offer empirical evidence that can help evaluate the impact of that sort of intervention. We analyze data tracking news consumption behavior to assess changes in audience reach and audience fragmentation. We show that the law has no discernible impact on reach, but we identify an increase in the fragmentation of news consumption.

In October 2014, the Spanish Congress approved a reformed Intellectual Property Law (IPL) regulating the creation of online content. The law, which took effect in January 2015, introduces a nonwaivable copyright fee to be paid by online aggregators for linking to content created by newspapers and publishers. The law is designed to affect any activity that relies on hyperlinks to online news content created by third parties, but makes an explicit exception for social media platforms like Facebook or Twitter (Ministerio Educación, 2014; Xalabarder, 2014). Immediate reactions to this law included Google closing down its News portal and removing Spanish media outlets from the service (Google, 2014; Rushe, 2014).

Prior to this episode, similar regulatory efforts had been attempted in Belgium, France, and Germany. However, in those instances the decision to tax links was made optional. The Spanish legislation, in other words, opened the door for a more aggressive regulation of online content. An important consequence of this legislation is that it affects not only giants like Google but also smaller actors, especially "native" digital newspapers or news communities like *Menéame* (the Spanish version of *Digg*).

The negative effects of the law are already visible, according to a report commissioned by the Spanish Association of Periodical Publishers (AEEPP), and they affect both legacy and new media actors (Méndez, 2017; Posada de la Concha, Gutiérrez, & Hernández, 2015). More recent research offers a better quantification of the impact. In one study, the evidence suggests that the shutdown of Google News in Spain resulted in a decrease of daily visits to Spanish news outlets by 11% (Calzada & Gil, 2016). Another study shows a decrease of 20% in overall news consumption among Google News users in the period after the search engine decided to close its portal (Athey, Mobius & Pal, 2017). This decrease, the study finds, was concentrated around small publishers: large publishers did not see significant changes in their overall traffic as a consequence of the shutdown. This negative assessment was also echoed recently by the U.S. government in a 2016 report stating that the "link tax" requires "careful monitoring" since it threatens the success of emerging business models (USTR, 2016, p. 179). And yet, in spite of growing concerns on the impact that this sort of intervention can have on the online media environment, similar regulations are now being considered on the European level (European Commission, 2016, 2017; Schechner & Woo, 2016). As currently phrased, the legislation proposed by the EU executive could allow press publishers to charge for the reproduction of headlines or the mere indexing of their articles (Meyer, 2016).

Underlying these regulatory efforts there is an assumption that access to online information is mainly driven by the linking activity of news providers (Calzada & Gil, 2016; Posada de la Concha et al., 2015; Xalabarder, 2014; Athey, Mobius & Pal, 2017). Whether links are used to steal or boost audience flow depends on who makes the argument, *i.e.* legacy media or digital-born media organizations. This paper presents novel data that can help evaluate the empirical consequences of the "link tax" legislation, and determine whether the regulation of linking activity has a discernible impact on how audiences consume news online. This is an important question if we also consider the role that social media play in granting access to news. Social media were explicitly excluded from the Spanish legislation but they are becoming increasingly important as entry points to news and political information (Newman, Fletcher, Kalogeropoulos, Levy, & Nielsen, 2017).

The regulation of online content relates, more broadly, to academic discussions about how digital technologies are affecting news consumption and access to political news, which is one of the most important components of democratic life (Delli Carpini & Keeter, 1996). Early accounts of the dangers of personalized information (Katz, 1996; Resnick, 1997; Sunstein, 2009; Turow, 1998) and more recent discussions of the effects of the 'filter bubble' (Pariser, 2011) share a common argument: that new technologies are increasing audience fragmentation, and that this has negative consequences for society and democracy. In the background of these discussions lies a model of democratic engagement that stems from epistemic and deliberative conceptions of the public domain (Berelson, 1952; Converse, 1964; Gutmann & Thompson, 2009; Habermas, 1994). This democratic ideal puts information at the center of civic life: information becomes a vehicle for political engagement; the key to gain political knowledge; and the foundation for political action (Baum & Groeling, 2008; Delli Carpini & Keeter, 1997; Knobloch-Westerwick & Johnson, 2014; Prior, 2007; Verba & Nie, 1987). Access to a common and rich informational space is, consequently, an important democratic condition: the quality of decision-making depends on having a space for the discussion of public affairs which, in turn, requires access to political information. It is no wonder, then, that a lot of attention is being paid to how digital technologies are reshaping the public domain and reconfiguring access to news.

The extent to which the web and related technologies preserve this ideal of a public sphere (Habermas, 1994) depends on the behavior of two actors: the providers of information and the consumers of that information. The "link tax" responds, to a great measure, to the lobbying efforts of European publishers. As stated by one of the politicians helping the recent Copyright Directive advance through the channels of the European Parliament, "Newspapers help set the agenda, so politicians have to listen to them" (Scott & Clark, 2015). The economic costs of having to comply with a "link tax" are clear: they are the main reason why Google decided to withdraw its services from Spain (Google, 2014); the social news website *Menéame* has already stated that it will not be able to play the costs derived from the "link tax" (Méndez, 2017); and *El País*, one of the legacy newsapers that promoted the law, recently withdrew its support to the regulation (Editors, 2017). But what is not so clear is how much of a difference links make to the other side of the equation, i.e. audiences, and to how they consume news and political information.

On the one hand, there is evidence that most online news consumption results from individuals visiting the home pages of their favorite news outlets, which tend to be mainstream media (Flaxman, Goel & Rao, 2016). There is also evidence that traffic to mainstream news sites from links and search engines accounts for a very small fraction of their total traffic (Allcott & Gentzkow, 2017). What this suggests is that imposing a copyright fee on links might not drastically change the way in which consumers access news online. On the other hand, there is evidence that the Google News shutdown did reduce the consumption of news, although this decrease affected mostly the smaller publishers (Calzada & Gil, 2016; Athey, Mobius & Pal, 2017).

In this paper, we conduct empirical analyses that add additional evidence to prior work and shed light on audience behavior in a unique regulatory framework. Using data from two different sources that track online browsing patterns, we show that the "link tax" did not have a significant impact on the reach of news sites, defined as the fraction of the online population accessing those sites. However, we also show that there is an increase in audience fragmentation, which we define as a reduction in the audience overlap of news media sites. Crucially, the approach we take to measuring audience fragmentation makes an important methodological improvement to how audience networks were mapped in prior work (explained in more detail in Mukerjee, Majó-Vázquez, & González-Bailón, 2017). Ultimately, our goal is to discuss whether the increase in fragmentation could result from the "link tax." First, though, we start by contextualizing our research in the larger literature on news consumption in the digital age.

Digital News Consumption

Digital technologies have brought fundamental changes to the way people consume political information. The Internet allows citizens to have greater control over news selection; it offers a broad range of sources to keep up with political events; and it transforms the gatekeeping process so that it is no longer the monopoly of traditional media (Benkler et al., 2015; Farrell & Drezner, 2008; Groshek & Tandoc, 2017). Early theoretical accounts of these changes were, for the most part, optimistic. However, scholars soon honed in on some of the pernicious consequences of digital technologies, identified with traditional conceptions of the public sphere in the backdrop (Chaffee & Metzger, 2001; Gitlin, 2002; Prior, 2008; Sunstein, 2009; Turow, 1998). These critical accounts pay special attention to the increasing fragmentation of the online domain and the negative effects that fragmentation has on democracy.

The main claim of these negative accounts is that the fragmentation of news production provokes a fragmentation in the audience, which, in turn, weakens the foundations of a common space for the discussion of public affairs. However, this prior work usually disregards the decisions that users make on how to navigate online content; it offers, in other words, theoretical approaches to media fragmentation that do not take into consideration the consumption patterns revealed by users. Scholars who assess the

behavior of online audiences have, in fact, offered evidence that contradicts the fears of fragmentation (Athey, Mobius, & Pal, 2017; Flaxman, Goel, & Rao, 2016; Gentzkow & Shapiro, 2011; Taneja, 2016; Trilling & Schoenbach, 2013; Webster & Ksiazek, 2012; Webster, 2014). These studies show that media diets are diverse and that they include the most prominent news media, which become the *de facto* common ground for news exposure. Determining how these two streams of evidence come together (i.e. the increasing fragmentation of news production, and the common grounds revealed by media diets) requires more research. The debate on how digital technologies shape audience behavior is, in other words, still open and unresolved.

Fragmentation in news consumption

In this study, fragmentation relates to the demand side of information. We define fragmentation not in terms of the aggregate number of media options available to consumers but in terms of how audiences distribute across those sources to obtain their news. We take a structural approach that looks at media networks and at how audiences navigate news sites. More formally, we measure fragmentation as a reduction in the number of ties capturing audience overlap between news media sites. If these ties exist it means that audiences consume news from a variety of sources; conversely, an absence of overlapping ties means that audiences self-select to consume news from a smaller range of sites.

Authors studying digital news from the supply side have overemphasized the specialization of news outlets, and assumed that the enhanced capacity for content personalization inevitably leads to increasing audience fragmentation (Chaffee & Metzger, 2001; Napoli, 2008; Sunstein, 2009). For example, Tewksbury (2005) sustains that the specialization of online media outlets leads to fragmentation because they attract homogeneous groups of users. Parallel to these studies, there is a related area of work that considers whether audience fragmentation responds to the informational structure built by news providers through their links (Ackland & Gibson, 2004; Williams, Trammell, Postelnicu, Landreville, & Martin, 2005). Theoretically, the provision of a link should be a journalistic activity: Through links, readers can trace sources of information; expand on the context that gives meaning to data; and elaborate on interpretations. But due to the crisis affecting the media industry worldwide (Pew Research Center, 2016; Newman et al., 2017), business criteria tend to dominate journalistic decisions. As a consequence, news sites rarely link to external competitors: They treat links not as journalistic objects that add value to the stories, but as economic assets or functional devices that can, in principle, keep audiences within corporate boundaries (Dimitrova, Connolly-Ahern, Williams, Kaid, & Reid, 2003; Karlsson, Clerwall, & Örnebring, 2014). The "link tax" that motivates this study, and the lobbying efforts that underlie its political sponsorship, are a good example of how business criteria dominate journalistic decisions – and of how the industry is trying to monetize emerging patterns of news consumption.

According to De Maeyer (2012), the tendency to avoid linking to outside sources shapes the network of news providers as "walled gardens" (Napoli, 2008, p. 63). This metaphor presumes that users will not venture beyond the walls established by hyperlinks. Yet again, this argument is not supported by evidence explicitly measuring the extent to which audiences respond to the connections created by media organizations. The "link tax" offers a good natural experiment to determine the impact that the existence of those links (or their absence, due to imposed copyright regulations) have on audience behavior.

The emergence of new players

These contradictory scenarios demand attention to the role that digital actors play in shaping online news consumption. Digital technologies have multiplied the number of sources that grant exposure to news – but also the means to get access to those news. News sources today can be classified in four main

types: (a) legacy media (i.e. the online version of newspapers that predate the internet, or other main-stream sources, like public service broadcasters or commercial TV channels); (b) digital-born media (i.e. outlets that were born with the internet, like BuzzFeed or the Huffington Post); (c) social media platforms (i.e. Facebook, Twitter); and (d) search engines (i.e. Google). Sources (c) and (d) are mostly referral sites, which means that they direct users, through links, to sources (a) and (b). The "link tax" discussed in this paper affects the links created by news aggregators, which, depending on interpretation (and the law is ambiguous in this respect) could affect media classified as search engines (category d) or digital-born news outlets (category b). Social media (category c) were explicitly excluded from the regulation, although they are also now under public scrutiny for their role in the spread of misinformation (e.g. Brinkhurst-Cuff, 2017). In fact, Facebook recently decided to use third-party news organizations and fact checkers to fight fake news (Mosseri, 2016; Sharockman, 2016; Snopes, 2016). This decision clearly points to the distinctive role that legacy and digital-born media play in the production of news, a role that social media platforms cannot play on their own.

Figure 1 provides a ranking of the most important social media sites for news in Spain, according to the Digital News Report published by the Reuters Institute for the Study of Journalism (Newman et al., 2017). For reference, rankings are also provided for the UK and the US. As the figure suggests, Facebook is reported to be the most important social media platform to obtain news, especially in Spain. It is unclear, however, what fraction of this population also consumes news through other means, i.e. subscriptions or direct search; or whether Facebook refers those users more frequently to legacy media or digital-born outlets. Since Facebook does not produce original content (or not yet), users obtaining news though this platform are still referred to the outlets that produced the news. One caveat is that news content published directly through the Instant Articles service is hosted by Facebook and hence traffic is not referred to the website of news organizations (Goel & Ravi, 2015). As of today, there is no systematic data tracking how many users encountering news through Facebook actually click on the links to read the original content. But we do know through several sources, including the Reuters surveys and the observational data analyzed in this paper, that Facebook (and other social media) are definitely not the only sources of news online – they are not even the main source of referrals: search engines and direct access are still the most prominent sources of traffic (Flaxman, Goel, & Rao, 2016; Allcott & Gentzkow, 2017).

The question of how diverse the average media diet is can be answered by looking at two trends: changes in the audience reach of different news outlets; and changes in the audience overlap between those news sources. In this paper, we focus on categories (a) and (b) above, i.e. legacy and digital-born news outlets. We do not include social media platforms in our analyses because, as already stated, they were explicitly excluded from the Spanish "link tax" law. In addition, even if they are important entry points to news, they are still just, for the most part, referral sites. This is probably the reason why social media were excluded from the "link tax" in the first place. It is also the case that social media like Facebook have been more willing to reach agreements to share revenues with news organizations, of which the Instant Articles service is an example. In any case, most links circulating through social media are created by users themselves (which accentuates, in fact, the filter-bubble effect to a greater extent than automated feed algorithms, at least in the US, see Bakshy, Messing & Adamic, 2015; also Lazer 2015). With these caveats in mind, the following section introduces the data and the methods we use to analyze changes in media reach and audience overlap. Ultimately, our goal is to provide empirical evidence that casts light on the impact that copyright regulations have on news consumption patterns.

Data and Methods

The first step in our analyses was to compile a list with the most prominent Spanish news sites in terms of their traffic, that is, their percentage reach of the online population. We used Alexa Internet's traffic
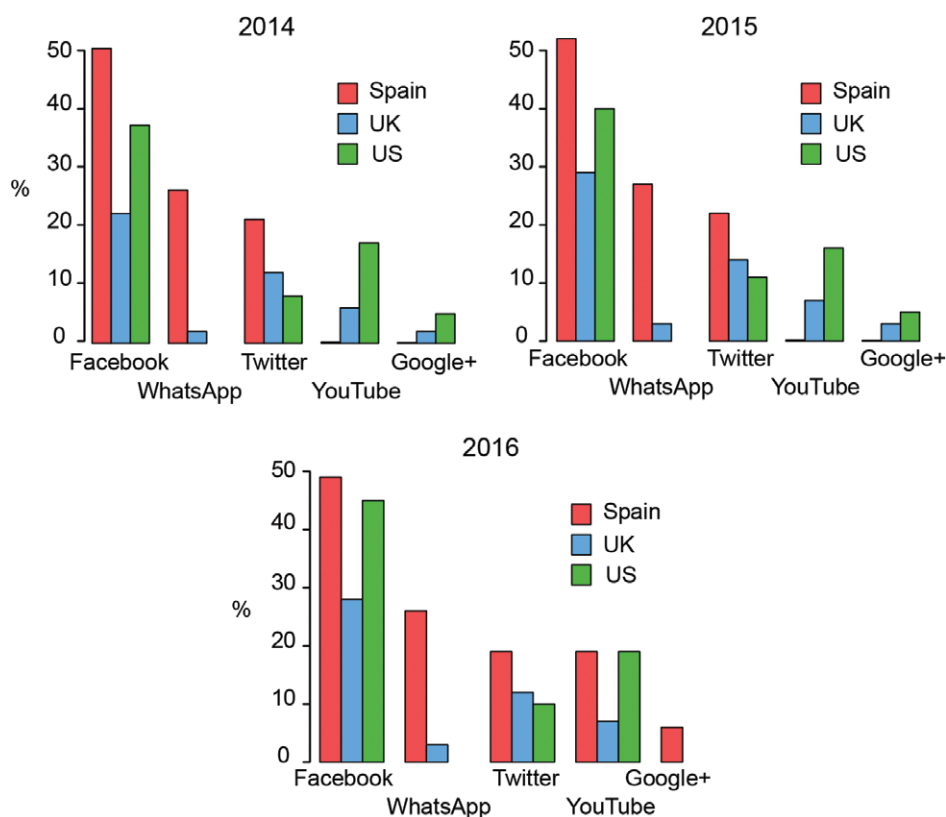
Figure 1 Top Social Networks Used to Access News.
Source: Reuters Institute for the Study of Journalism survey on Digital News. The question asked was "Which, if any, of the following have you used for news in the last week?"

rankings and comScore audience measurement statistics to create this list. These two companies use different methodologies to create their estimates. Alexa's ranks are based on the browsing behavior of people in their global panel, which is a sample of all Internet users. The panel used by comScore, on the other hand, is restricted to users based in Spain, and their estimates refer to the Spanish online population. In spite of these differences, both rankings have a correlation of 0.86, giving us confidence that the news sites selected (legacy and digital-born) are the most influential in terms of readership. One caveat worthwhile making is that rankings and audience metrics for the sites at the tail of the distribution (i.e. those with low traffic) are more unreliable and volatile. Since we analyze longitudinal data, a few of the sites in the original list appeared intermittently in the data. We removed these sites and, in the end, we were left with 93 sites, 36 classified as digital-born and 57 as legacy media.

Reach Data

We obtained daily reach data from Alexa for the period 1 September 2014 to 30 September 2015 (the "link tax" came into force on 1 January 2015; Google News closed on 16 December 2014). This gave us a time series with 395 observations for every news site. We aggregated these data for the two categories we are considering here: legacy and digital-born media. The purpose of analyzing these time series is to
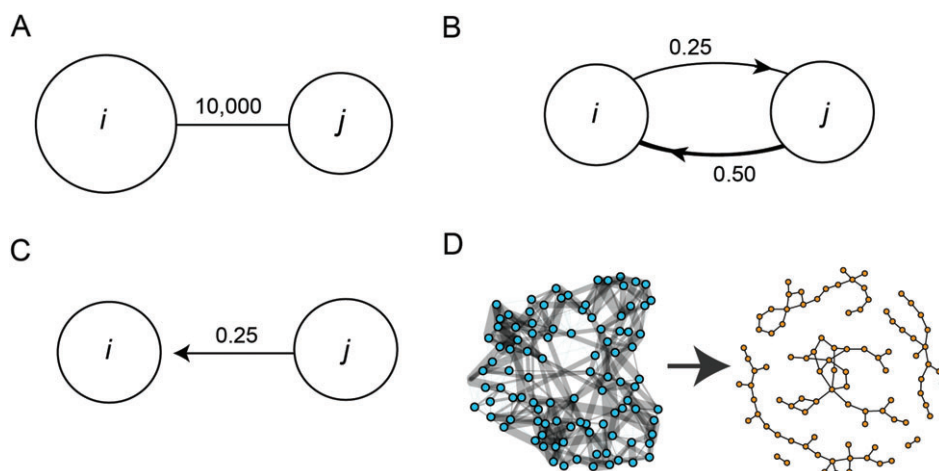
Figure 2 Networks of Audience Overlap.

determine if there is a visible shift in the average reach of these sites after the copyright law came into force in 1 January 2015. In the light of prior research (Calzada & Gil, 2016; Athey, Mobius, & Pal, 2017), our hypothesis is that the "link tax" had a negative impact on the traffic of digital-born outlets (the younger and smaller publishers) and no impact on the more established digital-born outlets. To determine if this hypothesis holds, we use time series modelling and, in particular, a technique known as intervention analysis (Cryer & Chan, 2008). The goal of this technique is to estimate if there is a statistically significant trend change after an intervention, controlling for seasonality and autocorrelation, which are typical features of time series data.

Audience Overlap Data

We obtained audience overlap data from comScore. This statistic gives us information on the size of the audience that access any two news outlets in a given month. Audience overlap offers the building block to reconstruct networks that can then be used to analyze the cohesiveness or fragmentation of news consumption patterns, as depicted in Figure 2. Panel A contains (an idealized) example of dyadic overlap between two news sources. According to this example, 10,000 online users access both site *i* and site *j*: this is the absolute duplication, or the fraction of the online population that gets exposed to these two news sources. Since media sites differ greatly in their audience reach, this raw overlap can be expressed as a percentage proportional to the number of people who read a given source, as depicted in panel B: here the network captures relative duplication and it tells us that 25% of site *i*'s audience also reads site *j*, but 50% of those reading site *j* also read site *i*. This asymmetry in audience behavior is best depicted in panel C, which tells us that site *j* shares 25% more audience with site *i* than vice versa.

Both the symmetrical and asymmetrical versions of this network (panels A and B) have been used in prior work (Ksiazek, 2011; Webster, 2014; Webster & Ksiazek, 2012; Taneja & Webster, 2016) following the methodology introduced in (Ksiazek, 2011). In this paper we use the undirected version of panel A to build our audience overlap networks, especially since we are also analyzing differences in reach through the time series. We aggregate audience data for the months of September 2014 and September 2015 – that it, at the beginning and at the end of the time series tracking changes in reach. Overall, in

analyzing these networks we follow the intuition of previous work but, crucially, we propose an important improvement to the methodology, as explained in more detail elsewhere (Mukerjee, Majó-Vázquez, & González-Bailón, 2017). The analyses conducted in prior work are limited in one important respect: the construction of audience overlap networks does not eliminate overlapping ties that might result from random noise in the data or just random browsing behavior. In particular, there are two problems with the approach followed in previous studies: first they calculate the difference between observed and expected duplication (as one would in an ANOVA table) using percentages, not frequencies, as the row and column marginals; and second, they do not test whether that difference (assuming it is meaningful) is statistically significant. Hence, prior work analyzed networks that contained substantial amounts of noise.

The elimination of nonsignificant ties is a procedure known in the networks literature as "thresholding" (Borge-Holthoefer & González-Bailón, 2017), and is illustrated by the idealized example in Figure 2, panel D. Here, we apply a technique that relies on the conventional *phi* coefficient but instead of being applied to a binary contingency table, it is applied to the matrix representing the number of common visitors to two news sites. If D_{ij} is the absolute audience duplication between news sources i and j, A_i and A_j are the total reach of each source, and N is the total online population, then the coefficient is expressed as:

$$\phi_{ij} = \frac{D_{ij}N - A_i A_j}{\sqrt{A_i A_j \left(N - A_i\right)\left(N - A_j\right)}}.$$

The coefficient is positive (or negative) when the overlap is larger (or smaller) than expected by chance. To determine the statistical significance of these correlations, the technique makes then use of the *t* statistic, which is also a conventional statistical tool to determine how likely it would be to observe the measured overlap if the null hypothesis of no overlap were true. The *t* value is calculated with the formula:

$$t_{ij} = \frac{\phi_{ij}\sqrt{N-2}}{\sqrt{1 - \phi_{ij}^2}}.$$

For a significance level $p < 0.01$ (two tailed) *t* values need to be $t_{ij} > 2.58$; overlapping ties with values below that threshold are eliminated as nonsignificant.

Figure 3 illustrates how much the networks change before and after thresholding. Panel A shows that the test eliminates ∼25% of all the overlapping ties. Most importantly, the resulting networks are significantly more centralized, which is a measure we can use to capture status or asymmetry in relational structures (Freeman, 1979). In the context of these networks, the centrality of sites flags where audiences concentrate and which sources are more prominent – which offers a good metric to compare the two categories of news sources we are interested in: legacy and digital-born news media.

Results

Changes in Percentage Reach

Figure 4 shows the time series of the aggregated percentage reach for legacy and digital-born media. As stated above, we focus on the period September 2014 to September 2015 but the figure also shows, for illustration, the longer trends in reach. The time series show a dip that coincides with the beginning of our observation window. We do not have a substantive reason to explain for this shift, but we suspect it is related to how the rise of mobile devices impacts Alexa's estimates (which rely on tracking software
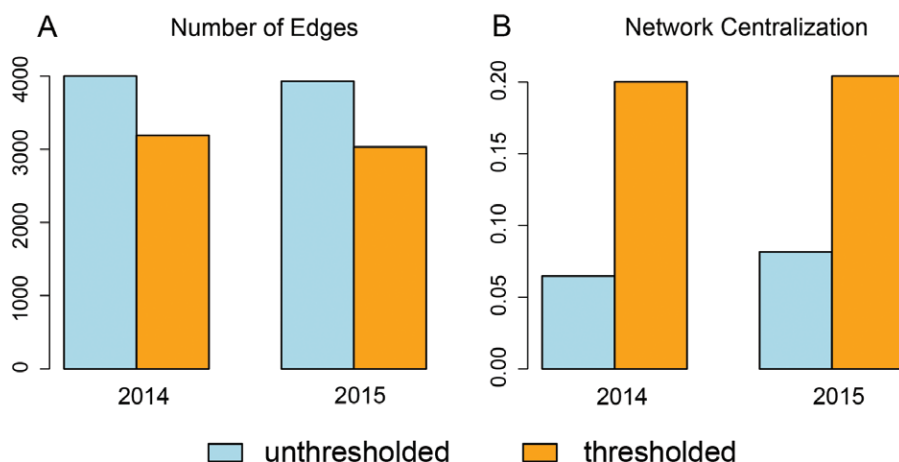
Figure 3 Audience Overlap Networks Before and After Thresholding.

installed on desktops). The shorter time series we analyze are bounded by our two measurements of audience networks (assembled with comScore data and analyzed below).

The blue dots track the average reach for legacy media sites; the red triangles track the average for digital-born news outlets. The first message contained by this figure is that legacy media have consistently better rankings in audience reach: they are read by twice as many internet users for the whole time period. The second message is that the "link tax" intervention seems to have made a small difference to the preintervention trend. The fitted line plots the predicted values of a linear model that includes a dummy variable for 1 January 2015, i.e. the date when the "link tax" took effect. There is an overall decaying trend, but the line suggests a bump in the reach of legacy media and, to a lesser extent, digital-born media that coincides with the implementation of the law. Of course, this is a very rough approximation to the data and the question is whether this change of trend remains significant once we control for autocorrelation and seasonality.

Table 1 shows the results of ARIMA models fitted to the data aggregated weekly (and selected following an assessment of usual diagnostics). The estimated parameters suggest that, for both time series, the intervention does coincide with a slight increase in reach; however, as the standard errors and z values indicate, this increase does not reach the threshold of statistical significance. In other words, there is no strong evidence that the law changed patterns of news consumption – at least, as assessed through this global aggregated measure of audience reach.

The analysis of the comScore data also show that there are no substantive changes in reach, although in this case we identify some significant (albeit small) shifts. Figure 5 shows the difference in the percentage reach of news sites in 2015 compared to their reach on 2014 (panel A). There is evidence that legacy media sites increase their average reach during this year by about 0.4 percentage points, a small increase that is nonetheless statistically significant (panel C). This increase sets legacy media significantly apart from digital-born sites (panel B). Digital-born sites remain, on average, on the same reach levels but they become significantly less visible relative to legacy media sites. Overall, legacy news organizations seem to have improved in their ability to attract audiences while digital-born outlets remain on similar levels compared to 2014. It is important to remember, however, that the outliers who dominate both groups are the net winners of these trends: the least
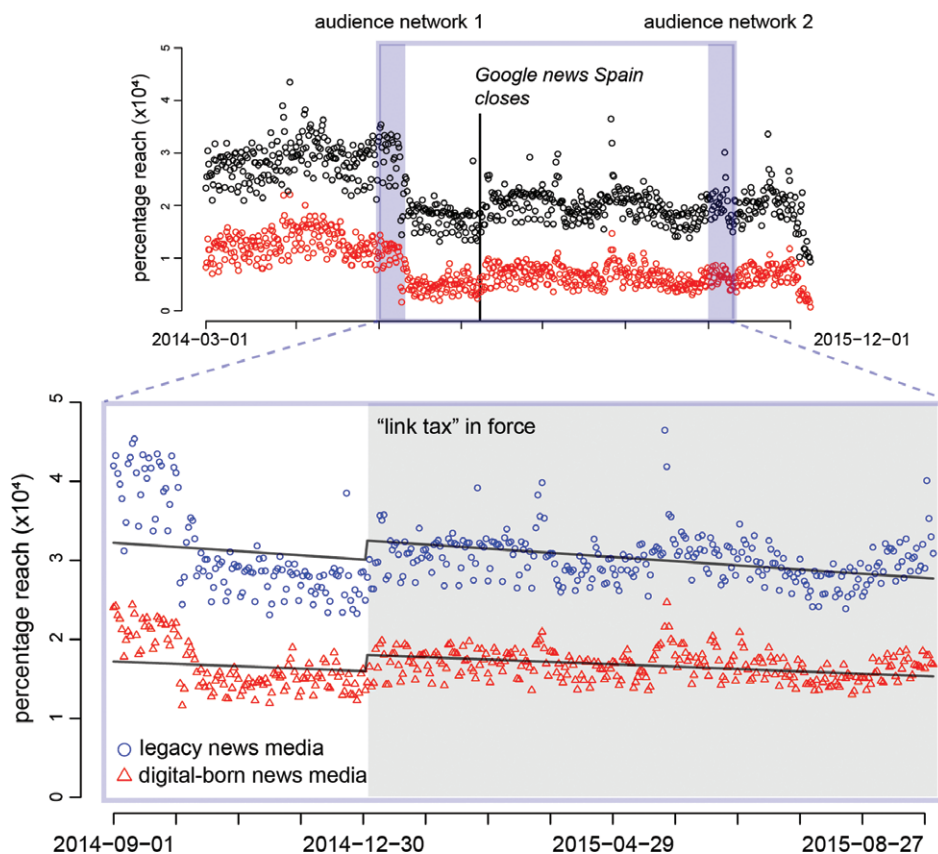
Figure 4 Changes in Percentage Reach for Legacy and Digital-Born Media (Alexa Data).

popular legacy news sites perform worse, in terms of reach, than the most popular digital-born sites.

Changes in Audience Overlap Networks

Figure 6 summarizes changes in the audience overlap networks – which, again, we use to assess changes in audience fragmentation. In particular, the figure pays attention to degree centrality, which measures the number of other sites a focal media outlet shares audience with. Another way to think of centrality in this network is in terms of public recognition: sites that are very well connected are also those that become the gravity centers of people that will then consult other, more diverse, secondary sources. Panel A shows there is a lot of variance in how central sites are in 2015 compared to 2014, and that there is a general tendency to lose overlapping ties. Both legacy and digital-born media lose audience overlap with other outlets (panel B) but this decrease is higher for digital-born media. In both cases, the decrease in centrality is statistically significant (panel C).

Table 2 gives some more statistics to compare these networks. Because of the lower density in the network of 2015, there is also a lower transitivity, that is, less clustering – which we interpret as additional evidence of increasing fragmentation. Still in both networks transitivity levels are higher than those expected in random networks (the confidence intervals give the transitivity scores corresponding

Table 1 Time Series Intervention Parameter Estimate

	ARIMA Model	Parameter	Estimate	SE	Z value
Legacy (weekly data)	(1,0,0)	Intercept	3.113	0.231	13.502
		AR1	0.872	0.080	10.956
		Intervention	0.017	0.147	0.910
Digital-Born (weekly data)	(1,0,0)	Intercept	1.650	0.127	12.983
		AR1	0.836	0.102	8.187
		Intervention	0.059	0.106	0.579

to the 2.5th and 97.5th percentiles of the simulated distributions, drawn from N = 1,000 random graphs that preserve the same size, density, and degree sequence). This means that, even though audience overlap goes down over time, it still reveals a clustering in how audiences behave that is not attributable to random browsing. The lack of significance of the homophily and degree correlation scores means that audiences do not self-select according to the type of media (legacy or digital-born) or to their prominence (core vs peripheral).

Discussion

Overall, our analyses do not show strong evidence that the "link tax" drastically changed the audience reach of news sites, but they show evidence of increasing fragmentation. The structure of the audience overlap network becomes sparser and more fragmented over time. A sparser and more fragmented network means that news sites share less audience, thus becoming more differentiated in who consumes
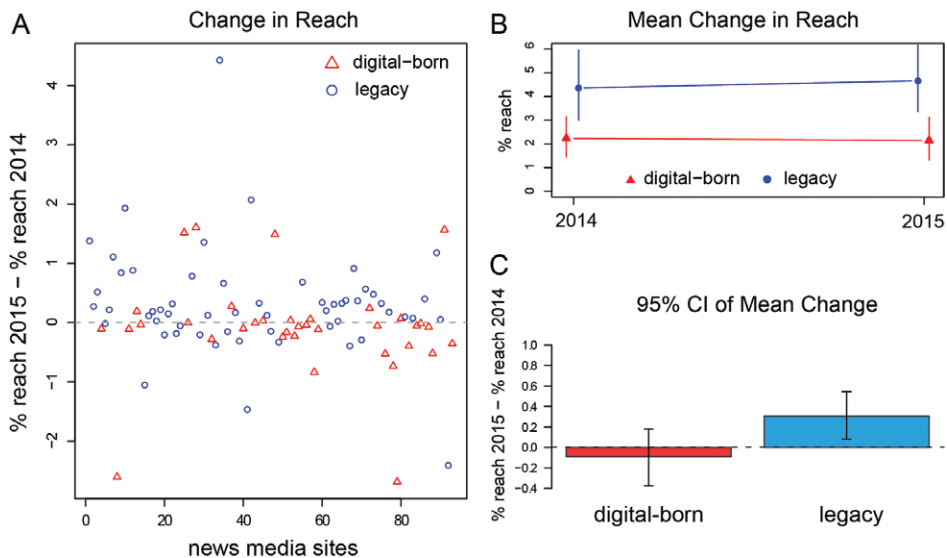
Figure 5 Changes in Percentage Reach for Legacy and Digital-Born Media (comScore Data).
Note: confidence intervals (CI) are bootstrapped based on non-parametric resampling of the data.
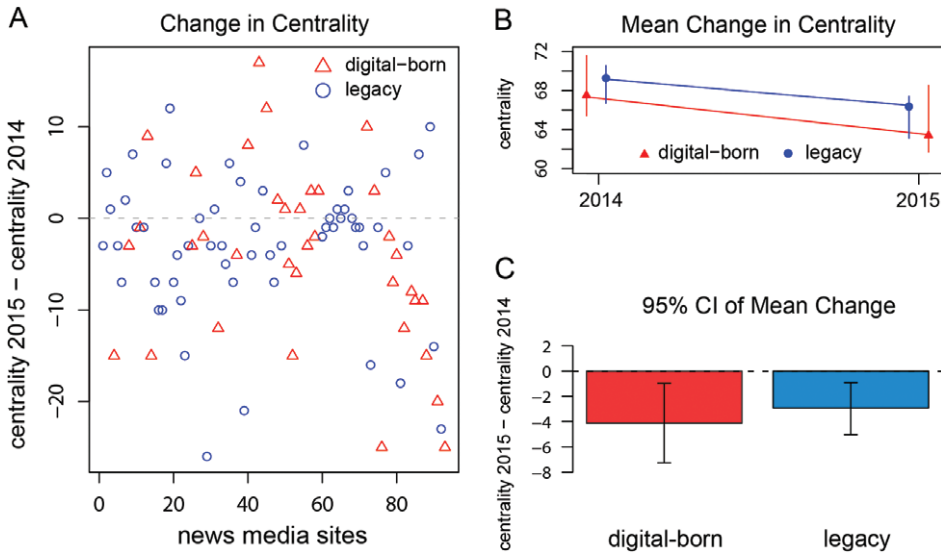
Figure 6 Changes in the Centrality of Sites in the Audience Overlap Networks.
Note: confidence intervals (CI) are based on bootstrapping and network permutations where media type labels (i.e. "digital born", "legacy") are randomly assigned.

Table 2 Network Statistics for the two Audience Overlap Networks

	2014		2015	
	observed	bootstrapped CI	observed	bootstrapped CI
Number of Nodes	93		93	
Number of Edges	3189		3032	
Transitivity	0.781	(0.516, 0.536)	0.758	(0.502, 0.522)
Homophily	0.007	(−0.028, 0.004)	−0.018	(−0.028, 0.005)
Degree Correlation	−0.143	(−0.033, 0.035)	−0.177	(−0.036, 0.034)

Note: bootstrapped confidence intervals (CIs) are based on random networks (N = 1,000) that preserve the same number of nodes, edges and degree distribution of the observed networks. For the homophily statistic, CIs are based on random permutations (N = 1,000) of the node labels classifying the sites as legacy or digital-born media.

their news. One possible explanation for this reduction in the number of overlapping ties is that, in the absence of aggregators like Google News, consumers do not explore so widely the available news sources – a possibility that prior research supports (Calzada & Gil, 2016; Athey, Mobius & Pal, 2017). The audience data we analyze would reflect this effect as a decreasing overlap between sites. Another explanation is that more people are consuming news through mobile devices that do not leave a trail in the comScore or Alexa panels (yet). For the period of analysis considered here, however, smartphones and tablets still trailed behind desktops as entry points to the web (Newman et al., 2017).
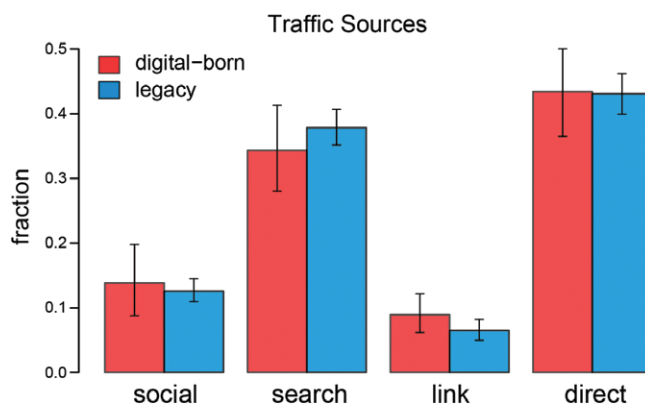
Figure 7 Referrals to News Media Sites.
Source: Alexa Internet.

One additional piece of evidence that can help discriminate between those possibilities comes in the form of referral data, as already discussed in the introduction to the paper. Figure 7 shows the sources of traffic for the news sites analyzed here, in the period following the "link tax" (and according to the Alexa global panel). As the numbers reveal, most people arrived at the news outlets either directly (by clicking on bookmarks or typing the address in the browser) or through search. Links and social media referrals amount to a small fraction of the total number of sources. In other words, the importance of links to direct online traffic might be overrated.

The figure also suggests that social media are not that prominent as gateways to news. Of course, referral data does not account for exposure to news within the platform – that is, for the number of social media users that never click on links but still read headlines and snippets from their feeds; and those who read third-party news content hosted by social media (i.e. through the Instant Articles service). Be that as it may, social media content has not been the subject of regulation – or not yet. The "link tax" makes it more difficult for media outlets to refer to content they did not create, but the impact this regulation has on user behavior is vague. For the time period we consider, branded legacy media is consistently the preferred source of information. There is also a range of secondary news outlets that users access to complement, not substitute, their primary sources. Given the clear and consistent evidence about this trend, both in this paper and in prior research, the question then is: why are not more people willing to pay for the news services of legacy media brands? This question is especially pertinent in Spain where legacy media are weaker in terms of subscriptions, compared to other European countries (Nicholls, Nabeelah, Nielsen, 2016). The causes that explain a weaker legacy media lie outside the scope of this paper, but it is unlikely that those causes respond only to the emergence of digital-born news outlets. The alternative hypothesis, i.e. that digital-born media emerges stronger where legacy brands are weaker, seem equally (if not more) plausible.

Future research should consider the impact of other regulatory initiatives – particularly aggressive in Europe – in redirecting the dynamics of news production and consumption. This policy dimension will become increasingly relevant as the impact of new technologies on the media landscape become more palpable. New technologies have allowed the emergence of unprecedented gateways to news consumption, and regulators are still trying to determine the extent of the transformations and whether (and how) to harness those changes, especially as they relate to shifting power structures. Modifications to intellectual property laws, as the Spanish "link tax" exemplifies, are also manifestations of a more general

problem that will require more attention in the immediate future, namely the impact of technology in public life – especially as algorithms become increasingly relevant in the curation of news and access to politically relevant information. As social media sites become even more prominent in directing traffic to political news and even in hosting third-party news content, it will also be important to understand their role in the public domain, determine the extent of their accountability, and translate that knowledge into effective interventions.

Conclusion

Our analyses contribute new evidence to still open debates about how digital technologies affect access to news. We used data from two different sources to track online browsing patterns, and we analyzed the impact of a copyright regulation that taxed links to news. We found that this intervention did not have a significant impact on the reach of news sites. However, we also found an increase in audience fragmentation, defined as a reduction in the audience overlap of news media sites. Our approach to measuring audience fragmentation makes an important methodological improvement to how audience networks had been mapped in the past. We discussed whether the increase in fragmentation could result from the "link tax" or from the parallel rise of social media as sources of information. We conclude that more research is necessary to determine which factors are the most important drivers in observed changes in news consumption patterns.

Acknowledgments

References

Ackland, R., & Gibson, R. (2004). *'Mapping Political Party Networks on the WWW', Australian Electronic Governance Conference 2004*, University of Melbourne, Melbourne, pp. 1-33.

Allcott, H., & Gentzkow, M. (2017). "*Social Media and Fake News in the 2016 Election.*" Working Paper, Stanford University.

Athey, S., Mobius, M., & Pal, J. (2017). "The Impact of Aggregators on Internet News Consumption." Stanford University Graduate School of Business Research Paper No. 17-8.

Baum, M. A., & Groeling, T. (2008). New Media and the Polarization of American Political Discourse. *Political Communication*, 25(4), 345–365. https://doi.org/10.1080/10584600802426965

Benkler, Y., Roberts, H., Faris, R., Solow-Niederman, A., & Etling, B. (2015). Social mobilization and the networked public sphere: Mapping the SOPA-PIPA debate. *Political Communication*, (ahead-of-print), 1–31.

Berelson, B. (1952). Democratic theory and public opinion. *Public Opinion Quarterly*, 313–330.

Borge-Holthoefer, J., & González-Bailón, S. (2015). Scale, Time, and Activity Patterns: Advanced Methods for the Analysis of Online Networks. In N. Fielding, R. Lee, & G. Blank (Eds.), *Handbook of Online Research Methods* (Second). Thousand Oaks: Sage Publications. Retrieved from https://ssrn.com/abstract=2686703

Brinkhurst-Cuff, C. (2017). "MPs to investigate threat to democracy from 'fake news'" The Guardian, January 29. https://www.theguardian.com/media/2017/jan/29/fake-news-mps-investigate-threat-democracy.

Calzada, J., & Gil, R. (2016). "What Do News Aggregators Do? Evidence from Google News in Spain and Germany." *SSRN* https://ssrn.com/abstract=2837553.

Chaffee, S. H., & Metzger, M. J. (2001). The end of mass communication? *Mass Communication & Society*, 4(4), 365–379.

Converse, P. E. (1964). The nature of believe system in mass public. In D. Apter (Ed.), *Ideology and Discontent*.

Cryer, J. D., & Chan, K.-S. (2008). *Time Series Analysis: With Applications in R*. New York: Springer.

Delli Carpini M. X., & Keeter, S. (1996). *What Americans Know about Politics and Why it Matters*. New Haven: Yale University Press.

De Maeyer, J. (2012). The journalistic hyperlink: Prescriptive discourses about linking in online news. *Journalism Practice*, 6(5–6), 692–701.

Dimitrova, D. V, Connolly-Ahern, C., Williams, A. P., Kaid, L. L., & Reid, A. (2003). Hyperlinking as gatekeeping: Online newspaper coverage of the execution of an American terrorist. *Journalism Studies*, 4(3), 401–414.

Editors. (2017, 24 March). Concordia digital. *El País*. Madrid. Retrieved from https://elpais.com/elpais/2017/03/23/opinion/1490295040_130405.html

European Commission. Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market. (2016). Brussels. Retrieved from https://ec.europa.eu/transparency/regdoc/rep/1/2016/EN/1-2016-593-EN-F1-1.PDF

European Commission. (2017). Draft Opinion of the Committee on Culture and Education for the Committee on Legal Affairs. Brussels.

Farrell, H., & Drezner, D. W. (2008). The power and politics of blogs. *Public Choice*, 134(1–2), 15–30.

Freeman, L. C. (1979). Centrality in social networks: Conceptual clarification. *Social Networks*, 2(3), 215-239.

Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, 80(S1), 298-320. https://doi.org/10.1093/poq/nfw006.

Garrett, R. K. (2009). Echo chambers online? Politically motivated selective exposure among Internet news users. *Journal of Computer-Mediated Communication*, 14(2), 265–285. https://doi.org/10.1111/j.1083-6101.2009.01440.x

Gentzkow, M., & Shapiro, J. M. (2011). Ideological segregation online and offline. *The Quarterly Journal of Economics*, 126(4), 1799–1839. https://doi.org/10.1093/qje/qjr044

Gitlin, T. (2002). Public sphere or public sphericules? In J. Curran & T. Liebes (Eds.), *Media, ritual and identity* (p. 168). Routledge.

Goel, V., & Ravi, S. (2015, 13 May). Facebook begins testing instant articles from news publishers. *The New York Times*. Retrieved from https://www.nytimes.com/2015/05/13/technology/facebook-media-venture-to-include-nbc-buzzfeed-and-new-york-times.html?mcubz=1

Google. (2014). Google Noticias España. Retrieved 1 February 2017, from https://support.google.com/news/answer/6140047?hl=es

Groshek, J., & Tandoc, E. (2017). The affordance effect: Gatekeeping and (non) reciprocal journalism on Twitter. *Computers in Human Behavior*, 66, 201–210.

Gutmann, A., & Thompson, D. (2009). *Democracy and disagreement*. Harvard University Press.

Habermas, J. (1994). Three normative models of democracy. *Constellations*, 1(1), 1–10. https://doi.org/10.1111/j.1467-8675.1994.tb00001.x

Karlsson, M., Clerwall, C., & Örnebring, H. (2014). Hyperlinking practices in Swedish online news 2007-2013: The rise, fall, and stagnation of hyperlinking as a journalistic tool. *Information, Communication & Society*, 18(7), 847–863.

Katz, E. (1996). And deliver us from segmentation. *Annals of the American Academy of Political and Social Science*, 546, 22–33. Retrieved from https://www.jstor.org/stable/1048167

Knobloch-Westerwick, S., & Johnson, B. K. (2014). Selective exposure for better or worse: Its mediating role for online news' impact on political participation. *Journal of Computer-Mediated Communication*, 19(2), 184–196.

Ksiazek, T. B. (2011). A network analytic approach to understanding cross-platform audience behavior. *Journal of Media Economics*, 24(4), 237–251. https://doi.org/10.1080/08997764.2011.626985

Méndez, M. Á. (2017, 7 February). Nuevo intento de imponer el canon AEDE: piden a Menéame 2,5 millones de euros al año. *El Confidencial*. https://www.elconfidencial.com/tecnologia/2017-02-07/canon-aede-meneame-internet-facebook-agregadores_1327333/

Meyer, D. (2016, 14 September). Why Europe's new copyright proposals are bad news for the Internet. *Fortune*. https://fortune.com/2016/09/14/europe-copyright-google.

Ministerio Educación, C. y D. (2014). *El Ministerio de Educación, Cultura y Deporte valora positivamente el consenso alcanzado por los grupos parlamentarios del PP, el PSOE y CiU en la votación de enmiendas sobre la lucha contra la piratería en la Ley de Propiedad Intelectual*. Madrid: Ministerio Educación Cultura y Deporte. Retrieved from https://goo.gl/1f2haH

Mosseri, A. (2016). News Feed FYI: Addressing Hoaxes and Fake News. Retrieved 1 January 2017, from https://newsroom.fb.com/news/2016/12/news-feed-fyi-addressing-hoaxes-and-fake-news/

Mukerjee, S., Majó-Vázquez, S., & González-Bailón, S. (2017). "Networks of Audience Overlap in the Consumption of Digital News." *Available at SSRN* http://ssrn.com/abstract=3032609.

Napoli, P. (2008). Hyperlinking and the forces of massification. In J. Turow & L. Tsui (Eds.), *The hyperlinked society: Questioning connection in the digital age* (pp. 56–69). Ann Arbor, MI: The University of Michigan Press.

Neuman, W. R. (2001). The impact of the new media. In W. L. Bennett & R. Entman (Ed.), *Mediated politics: Communication in the future of democracy* (pp. 299–320). New York: Cambridge University Press.

Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D. A. L., & Nielsen, R. K. (2017). *Reuters Institute Digital News Report 2017*. Oxford: Reuters Institute for the Study of Journalism, University of Oxford.

Nicholls, T., Shabbir, N., and Nielsen, R. K. (2016). *Digital-born news media in Europe*. Oxford: Reuters Institute for the Study of Journalism.

Pariser, E. (2011). *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin.

Pew Research Center. (2016). *State of the News Media* 2016.

Picard, E. (2008). Hyperlinking and advertising strategy. In J. Turow & L. Tsui (Eds.), *The hyperlinked society: Questioning connection in the digital age.* (pp. 159–164). Ann Arbor, MI:University of Michigan Press.

Posada de la Concha, P., Gutiérrez, A., & Hernández, H. (2015). *Impacto del Nuevo Artículo 32.2 de la Ley de Propiedad Intelectual,* NERA Economic Consulting, July 9.

Prior, M. (2007). *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections*. Cambridge University Press.

Resnick, P. (1997). Filtering information on the Internet. *Scientific American*, 276(3), 62–64. https://doi.org/10.1038/scientificamerican0397-62

Rushe, D. (2014, 11 December). Google News Spain to close in response to story links "tax". *The Guardian*. Retrieved from https://www.theguardian.com/technology/2014/dec/11/google-news-spain-to-close-in-response-to-tax-on-story-links

Schechner, S., & Woo, S. (2016, 14 September). EU executive proposes new copyright, communications laws. *The Wall Street Journal*. https://www.wsj.com/articles/eu-leaders-propose-new-copyright-communications-laws-1473850691.

Scott, M., & Clark, N. (2015, 28 August). European publishers play lobbying role against Google. *The New York Times*. https://www.nytimes.com/2015/08/29/technology/european-publishers-play-lobbying-role-against-google.html?_r=1.

Sharockman, A. (2016, 15 December). Fighting back against fake news. Politifact. Retrieved from https://www.politifact.com/truth-o-meter/article/2016/dec/15/fighting-back-against-fake-news/

Snopes. (2016). Snopes.com participates in Facebook's effort to combat hoaxes, fake news. Retrieved 1 January 2017, from https://www.snopes.com/2016/12/15/snopes-com-facebook-fake-news-initiative/

Sunstein, C. R. (2009). *Republic. com 2.0* (second). Princeton University Press.

Taneja, H. (2016). Mapping an audience-centric World Wide Web: A departure from hyperlink analysis. *New Media & Society*, 1–18. https://doi.org/10.1177/1461444816642172

Taneja, H., & Webster, J. G. (2016). How do global audiences take shape? The role of institutions and culture in patterns of web use. *Journal of Communication*, 66(1), 161–182.

Tewksbury, D. (2003). What do Americans really want to know? Tracking the behavior of news readers on the Internet. *Journal of Communication*, 53(4), 694–710.

Tewksbury, D. (2005). Seeds of audience fragmentation: The specialization in the use of online news sites. *Journal of Broadcast & Electronic Media*, 49, 332.

Trilling, D., & Schoenbach, K. (2013). Patterns of news consumption in Austria: How fragmented are they? *International Journal of Communication*, 7, 25.

Turow, J. (1998). *Breaking up America: Advertisers and the new media world*. Chicago, IL: University of Chicago Press.

USTR. (2016). The 2016 National Trade Estimate Report. https://ustr.gov/sites/default/files/2016-NTE-Report-FINAL.pdf.

Van Alstyne, M., & Brynjolfsson, E. (1996). Electronic communities: Global village or cyberbalkans. In *Proceedings of the 17th International Conference on Information Systems*. *New York*: *Wiley* (p. 32). Retrieved from https://web.mit.edu/marshall/www/papers/CyberBalkans.pdf

Verba, S., & Nie, N. H. (1987). *Participation in America: Political democracy and social equality*. University of Chicago Press.

Webster, J. G. (2005). Beneath the veneer of fragmentation: Television audience polarization in a multichannel world. *The Journal of Communication*, 55(2), 366–382.

Webster, J. G. (2014). *The marketplace of attention: How audiences take shape in a digital age*. Mit Press.

Webster, J. G., & Ksiazek, T. B. (2012). The dynamics of audience fragmentation: Public attention in an age of digital media. *Journal of Communication*, 62(1), 39–56. https://doi.org/10.1111/j.1460-2466.2011.01616.x

Xalabarder, R. (2014). *The remunerated statutory limitation for news aggregation and search engines proposed by the Spanish government: Its compliance with international and EU law* (WP14-2004). Barcelona.

About the Authors

Sílvia Majó-Vázquez is Research Fellow at the Reuters Institute for the Study of Journalism at the University of Oxford. Her research areas include news audience behavior and the role of digital-born and legacy media in European news markets. She applies tools from network science to the study of news

consumption and production. She has been a research fellow at the DiMeNet group of the Annenberg School for Communication at the University of Pennsylvania and a pre-doctoral researcher at the Internet Interdisciplinary Institute (In3-UOC).

Ana S. Cardenal is an Associate Professor of Political Science in the Law and Political Science Department at the Universitat Oberta de Catalunya (UOC). She holds a Ph. D in Political Science by the Universitat Autónoma de Barcelona (UAB), a degree in IR by Johns Hopkins University, and a B.Sc. in Journalism (UAB). She has been a visiting researcher at Stanford's Center for Latin American Studies and a Fulbright Fellow at New York University's Politics Department. Currently, she leads a research group at UOC on digital media, public opinion, and political behavior.

Sandra González-Bailón (corresponding author) is an Assistant Professor at the Annenberg School for Communication, University of Pennsylvania, where she leads the DiMeNet research group. Prior to joining Penn, she was a Research Fellow at the Oxford Internet Institute, where she is now a Research Associate. Her research lies at the intersection of network science, data mining, computational tools, and political communication. Her book *Decoding the Social World* is forthcoming with MIT Press (fall 2017).

Address: Annenberg School for Communication, University of Pennsylvania, 3620 Walnut Street, Philadelphia, PA 19104-6220, USA. E-mail: sgonzalezbailon@asc.upenn.edu.