

The Backbone Structure of Audience Networks:
A New Approach to Comparing Online News Consumption across Countries

Sílvia Majó-Vázquez*, Rasmus K. Nielsen*, Sandra González-Bailón**

* Reuters Institute for the Study of Journalism, University of Oxford

** Annenberg School for Communication, University of Pennsylvania

Abstract: measures of audience overlap between news sources give us information on the diversity of people's media diets and the similarity of news outlets in terms of the audiences they share. This provides a way of addressing key questions like whether audiences are increasingly fragmented. In this paper, we use audience overlap estimates to build networks that we then analyze to extract the backbone – that is, the overlapping ties that are statistically significant. We argue that the analysis of this backbone structure offers metrics that can be used to compare news consumption patterns across countries, between groups, and over time. Our analytical approach offers a new way of understanding audience structures that can enable more comparative research and, thus, more empirically grounded theoretical understandings of audience behavior in an increasingly digital media environment.

Keywords: online news; audience networks; fragmentation; comparative research; legacy media; digital-born media.

Corresponding author: Sandra González-Bailón, Annenberg School for Communication, University of Pennsylvania, 3620 Walnut Street, PA 19104, Philadelphia, U.S.

Email: sgonzalezbailon@asc.upenn.edu

Acknowledgements: work on this paper has been funded by NSF grant #1729412.

The consumption of political news is a core element of democratic engagement. A long tradition of media effects research has shown that the consumption of news has a positive impact on political knowledge, political participation, and civic engagement and thus play an important role in the democratic process (Dahlgren, 2005; Delli Carpini, 2004; Norris, 2000; Prior, 2007). Much of this research, however, has been focused on traditional forms of offline media like television and printed newspapers. Today, news is increasingly accessed online: digital media have already surpassed television as the most widely used source of news in many countries (see Newman, Fletcher, Kalogeropoulos, Levy, & Nielsen, 2017). This development presents political communication research with new challenges, including developing methodologies for understanding whether the move to a digital media environment is accompanied by audience fragmentation – and trends of balkanization, echo chambers, and filter bubbles as feared by some (Berry & Sobieraj, 2014; Garrett, 2009; Jamieson & Cappella, 2008; Baum & Groeling, 2008; Katz, 1996; Stroud, 2011; Sunstein, 2009; Turow, 1998). Another important question is whether these trends emerge in every national context and for every audience group, e.g., younger and older demographic segments of the population.

This paper introduces a methodological approach that, we argue, will help generate a better understanding of the structure of online news consumption and enable more comparative work (across countries, across demographic groups, and over time). Our approach borrows techniques that are well established in the field of network science but uncommon in political communication research. These techniques, we contend, can help develop a better, empirically-grounded and theoretical understanding of news audience formations in digital media environments. The approach we propose relies on the analysis of audience networks, which measure the amount of audience that news sites share. We specifically focus on the strength of the audience overlap across news sites. These audience networks are maps where the nodes represent media sites and the ties measure the number of individuals that consume news from a given pair of sources. The core of our method relies on the analysis of these networks once we have extracted the connections that are statistically significant – what we call the backbone of the network. This approach has never been used to analyze news consumption patterns but it is crucial, we argue, to obtain robust measures that can be compared across countries and media contexts.

Our approach is based on the analysis of the digital traces that people leave behind when accessing news content online. This offers an alternative source of evidence to surveys, which have traditionally been used to measure news consumption but have known weaknesses due to the limitations of self-reported data (Prior, 2009; Scharkow & Bachl, 2017). Trace data offer an alternative way to measure news consumption based on what is observed, not recalled. However, for all the enthusiasm that surrounds the increasing availability of trace data and so-called “digital footprints”, it is important to underline that data tracking audience behavior is not informative on its own: new data requires new methods to extract meaningful information and filter out the noise (King, 2014). We offer one such method in this paper.

In the analyses that follow we demonstrate how to apply the method we propose by examining three countries (the US, the UK, and Spain) that represent different regulatory models and media systems. A call for comparative analyses has been repeatedly made in the literature to avoid making inferences about diverse media markets using one single case study, usually the US. The selection of our cases was driven by the fact that they represent different regulatory frameworks and journalistic practices. The UK, for instance, has a long history of public service media that is widely used and well-funded. The US media market, on the other hand, is dominated by private organizations and characterized by an atomized supply. Finally, the Spanish case is also characterized by government intervention in the media market but this intervention is much weaker than in the UK; additionally, recent years have seen a fast proliferation of digital-born outlets, some of them with a similar capacity to attract audience attention than legacy media. We expect to find substantive and significant differences in how audiences navigate alternative media landscapes: the UK case, for instance, should be characterized by media networks where audiences are more centralized, given the prominence of public service broadcasting. Our method helps us test that intuition and quantify differences across media contexts with metrics that can be compared in a standardized fashion.

All in all, this paper makes three contributions. First, it presents a statistically robust method that extracts the backbone of audience overlap networks and preserves the most valuable information to understand news consumption online. Second, it demonstrates how this method can be used to assess patterns of online news consumption across different countries and media systems. And third, it looks at the structural position of legacy media and digital-born outlets

across media environments and demographic groups. These analyses offer quantifiable evidence to assess the role that emerging news providers play in different political contexts. Following a convention in the literature, we talk about digital-born outlets to refer to news sources that were born with the internet and do not have an offline edition. Legacy media, on the other hand, are the news sources that precede the internet – even if today most of them also have a strong digital presence. One of the questions our method aims to answer is the extent to which digital-born outlets are catching up in terms of reach and overall centrality. The answer to this question can help evaluate previous theoretical work arguing for a relocation of roles among types of media outlets, suggesting that legacy media are having their positions overtaken by new media (Castells, 2009; Chadwick, 2013; Jarvis, 2016; Pavlik, 2001). Although this paper is clearly methodological in scope, our argument also emphasizes the importance that better measurements have for theoretical development. Specifically, the method developed here provides a more sophisticated way of understanding audience fragmentation and the role played by digital-born media in granting access to news. Both are central issues for public opinion formation and, consequently, for the field of political communication.

1. The Rising Prominence of Digital News

The claim that digital technologies are increasingly relevant in granting access to news has now become a truism. As figure 1 shows, during the last five years online media have been an important (when not the main) source of news for the large majority of the population (around 80%) in the three countries we consider, according to Reuters Digital News Reports; online media vastly surpass print media and they are more widely used than TV in Spain and the US.

-- Figure 1 about here --

< Figure 1. Main Sources of News in the US, UK, and Spain >

These percentages, however, do not give us much information about the most prominent sources online, how heterogeneous these sources are in terms of total audience reach, or how

digital-born outlets compare with legacy media brands. The already classic long-tail argument suggests that online sources are very heterogeneous in how much attention they capture (Anderson, 2006; Hindman, 2009). Figure 2 confirms this claim. The histograms summarize the total reach of the news sites we consider in this paper, i.e. all sites classified under the “News/Information” category by comScore, a media measurement and analytics company that manages representative panels of internet users in the three countries we consider (panel sizes are $N \sim 210,000$ for the US, $N \sim 67,000$ for the UK, and $N \sim 30,000$ for Spain). The list of news sources, which was also checked manually to ensure all sites were relevant, has sizes $N = 332$ for the US; $N = 133$ for the UK; and $N = 185$ for Spain. These lists include legacy and digital-born outlets. All these sites have a reach of at least 0.01% of the total online population (below this threshold, comScore statistics become unreliable). The panelists agree to install software in their desktop computers that keeps track of browsing activity, and these logs are then combined with traffic data collected from the web pages. Our study, in other words, draws from monthly statistics of web use based on observed behavior collected both at the site and the user ends.

What figure 2 reveals is that a small number of media outlets capture most of the attention online, and these outlets happen to be legacy brands. The questions that interest us here are: Do the audiences of these outliers also consume other news sources and, if so, is there any evidence of self-selection in how they navigate the rest of the digital news environment? And are there any visible differences across countries in the prominence (i.e. centrality) that digital-born outlets have in relation to legacy brands? The following section gives more details about the data and methods we use to answer those questions.

-- Figure 2 about here --

< Figure 2. Total Audience Reach for News Sources in the US, UK, and Spain >

2. Data and Methods

2.1. Audience Networks

Audience duplication data was obtained from comScore in the form of monthly statistics estimating the number of users that access any two sites, e.g. how many people who accessed

The New York Times also accessed The Washington Post during a given month. As already illustrated in Figure 2, news sites differ greatly in their reach: cnn.com, for example, has a monthly reach of 57% of the US online population; the bbc.com has a reach of 43% of the UK online population; and the legacy newspaper elmundo.es has a reach of 27% of the Spanish online population. At the bottom of the audience reach ranking we find local or niche sites. These sites are less prominent in absolute terms but they are important to understand the diversity of media diets (at least, in terms of alternative sources that people navigate).

We use the audience overlap metric to build networks as detailed in Figure 3, which also summarizes our data collection strategy. In panel A we illustrate the timeline of our observation windows. For the UK and Spanish cases, we analyze audience duplication data for the months of May, June and July, that is, a month before, during, and after the Brexit referendum and the Spanish 2016 General Elections. For the US case, we analyze audience data for the months of October, November and December, that is, the period surrounding the 2016 Presidential Elections. Since audience overlap statistics fluctuate, we used three-month averages to build the networks that we analyze.

In this case, we examine audience behavior around major political events when the need for information increases and media diets are expected to be more diverse; but the method could equally well be used between elections to analyze how audiences change in response to major political events. In our networks, nodes are news sites and the ties map the strength of the overlap between those sites: the stronger a tie is, the more people access a given pair of news sources. We then sliced the networks by age groups, as depicted in Figure 1, panel B. This is to illustrate how our method can be used to compare audience behavior within countries as well as across countries. The age groups are the same for the three countries, with the exception of the youngest cohort, which has an age bracket of 15-24 in the US and the UK but 18-24 in Spain.

-- Figure 3 about here --

< Figure 3. Summary of the Audience Data Analyzed >

Our analytical goal is threefold: (1) to introduce a methodology for the analysis of audience overlap networks that filters out insignificant ties (according to a network-based null model and a probabilistic threshold of statistical significance); (2) to quantify news consumption

patterns in a way that can be compared across countries, between groups, and over time; and (3) to determine whether there are statistically significant differences in how people consume news online, paying special attention to the position of digital-born outlets vis-à-vis legacy media. We want to map, in other words, the media landscape as it emerges from people's choices in their search for news information online. The advantage of having those maps is that they can then be characterized and assessed in a systematic fashion to inform our understanding of news consumption from a comparative and relational perspective.

The use of duplication data to build audience networks was first introduced in a paper published in 2011 (Ksiazek, 2011), which was soon followed by a number of other studies that used the same methodology (e.g., Taneja, 2016; Taneja & Webster, 2016; Webster & Ksiazek, 2012). More recent research has proposed changes to the original methodology, which was limited in important ways: for instance, the strength of the overlap was disregarded from the analyses, and there was no assessment of the statistical significance of the observed overlap (Mukerjee, Majó-Vázquez & González-Bailón, 2018a; Mukerjee, Majó-Vázquez & González-Bailón, 2018b). We build on that work here to introduce a new technique that identifies the backbone, or the most significant overlap, in networks of news consumption. Unlike prior work, this technique defines the null model at the node (ego-centric) level, not at the dyadic level, and it offers a way to sort signal from noise while taking into account the structural properties of the observed network. This, we argue, is an important requirement when working with datasets that track digital traces: they might not suffer from the problems of imperfect recall but they offer, nonetheless, noisy measurement. Using these type of techniques is becoming increasingly relevant in the field of political communication and, in particular, in research that aims to determine the impact that online technologies have on audience fragmentation.

2.2. Backbone Extraction

Depending on how people consume news online, the resulting networks of audience overlap can look very different. Figure 4, panel A summarizes the possibility space within which observed audience networks can emerge. On one extreme (network 1) we have a scenario where there is no overlap, so the nodes (i.e. the news sites) share no audience and consequently are disconnected. This would be a case of extreme fragmentation and audience self-selection. On the

other extreme (network 5), we have a scenario of complete overlap, where all sites share audience with all other sites in the network. This would signal omnivorous news consumption practices. Of course, most empirical networks are likely to fall between these two extremes – the empirical question we want to answer is where exactly, considering the possibility space. The figure gives three schematic examples of intermediate cases: one in which the network is highly centralized around a hub (network 2); a more decentralized version where audience overlap is more evenly distributed (network 4); and a case where there are two clusters of sites that share audience amongst them but are disconnected from each other (network 3). The analyses we present below aim to differentiate these possibilities and determine if news consumption in specific media environments can be better defined by structures like (2), (3) or (4) – in line with the theoretical intuitions and hypotheses derived from how different regulatory frameworks operate, as explained in section 1.

-- Figure 4 about here --

< Figure 4. Schematic Representation of Backbone Extraction Technique >

A step prior to the analysis of these networks, however, involves filtering them so that only the overlap that is statistically significant (that is, unlikely to result from random chance) is retained. The filtering technique we propose in this paper is known as backbone extraction or disparity filter (Serrano, Boguñá, & Vespignani, 2009; see also Bessi and Briatte, 2016; Welbers and van Atteveldt, 2016; and Teixeira, 2018 for alternative implementations of the code in R). This technique eliminates ties that do not depart significantly from what would be expected under the null hypothesis of random weight distribution. For illustrative purposes, panel B of Figure 4 shows a simulated network before and after the backbone has been extracted. The thickness of the lines is proportional to the tie weight, which in our case measures the strength of audience overlap; the color of nodes in this visualization indicates clustering, that is, sites that are better connected to each other than to other sites. The backbone network is sparser because it has eliminated many of the weakest ties. Of course, what counts as a strong or a weak tie depends on the node adjacent to that tie: news sites with a large audience reach (i.e. the BBC) will have stronger connections to other sites than smaller outlets with less audience to share.

The backbone extraction technique takes into account the fact that the significance of tie strength is relative to the node being considered. Panel C in Figure 4 summarizes the null model that allows the technique to take into account disparity in the distribution of weights and determine statistical significance. First, the weights of all ties surrounding a node are normalized so that they fall in the interval $[0, 1]$ (network a). Then those weights are distributed uniformly so that each tie has the same strength (network b); these randomized weights, which express the null hypothesis, are then compared with the observed weights and only in cases where the difference is larger than a critical value, the ties are retained as statistically significant (network c). As with the more conventional t -tests, the critical value depends on the probability p used to define the threshold of significance. In this paper we use a threshold $p < 0.05$ – which means that the probability of observing a given overlap is very unlikely under the null hypothesis of random overlap distribution, so the tie is retained. As stated above, this approach is different from that used in previous published work (e.g., Majó-Vázquez, Cardenal & González-Bailón, 2017; Mukerjee, Majó-Vázquez & González-Bailón, 2018a; Mukerjee, Majó-Vázquez & González-Bailón, 2018b) in that it defines the null model on the ego-network level, not on the dyadic level; this analytical choice takes into account the fact that the distribution of overlapping ties surrounding a news site is shaped by that site’s total reach and overall centrality in the network.

2.3. Network Measures

Table 1 compares the audience networks before and after backbone extraction. In general, audience overlap networks are very dense, but many of those overlapping ties disappear in the backbone representation – this is the reason why the backbone networks are comparatively sparser. Importantly, they are also substantially more centralized (that is, closer to network (2) in Figure 4A). About 30% of all the news sites included in these networks are digital-born outlets; in the Spanish case, however, the percentage is much higher: more than half of the outlets are digital-born, the vast majority of them led by journalists who used to work for legacy organizations (Minder, 2015; Schoepp, 2016).

-- Table 1 about here --

< Table 1. Statistics for Audience Overlap Networks before and after Backbone Extraction >

3. Analyses

Figure 5 plots the centrality scores of news sites in the backbone networks. Digital-born outlets are significantly more central in the US than legacy media: on average, they share audience with 12 more outlets. The UK and Spanish cases reveal the opposite tendency: legacy media sites are more central, having overlapping ties with a higher number of other outlets. In the UK this difference is not significant, but it is significant in the Spanish case: legacy media sites share audience with 5 more outlets, on average, than digital-news sites. We can interpret these centrality scores as proxies to inequality and diversity in audience base: in the Spanish case, the difference in centralization suggests that legacy media sites are more attractive to a wider range of the online population; in the US case it is digital-born outlets that are more attractive. Sites with higher centralization, in other words, have a more diversified portfolio of users (where diversity is measured by the number of other outlets those users also consume).

-- Figure 5 about here --

< Figure 5. Differences in the Network Centrality of News Media Sites >

These patterns persist when we take age into account – a demographic that has been theorized in prior work as marking a divide in news consumption (American Press Institute, 2015; Antunovic, Parsons, & Cooke, 2016; Shehata, 2016). Digital-born outlets are more central in the US in every age group (particularly so for people aged 55 and above), and they are less central for every age group in Spain. In the UK, there are still no significant differences, regardless of who access the sites (junior or senior users).

Overall, there are clear differences in the structure of the networks across countries – more so than across age groups. As Figure 6 shows, the US network is the least centralized: users consume news in a more distributed way, i.e. they have a more diverse news diet, than those in the UK. Going back to Figure 4A, the US network would be closer to structure (4), the UK network would be closer to structure (2). The Spanish case stands in between. In all cases,

however, the centralization scores are significantly higher than expected by chance (although they do not change drastically across age groups).

-- Figure 6 about here --

<Figure 6. Differences in the Network Centralization across Countries and Age Groups>

Figure 7 plots the modularity scores of the networks assembled, again, by age groups. These scores offer a network statistic that identifies the existence of clusters in a network where nodes are better connected by audience ties (as illustrated by the color-coded groups in Figure 4B; the technique we use here is based on random walks, see Pons & Latapy, 2006 for technical details). The higher this score is, the higher the modularity of a network – and the better the groups are defined. As the figure shows, modularity is significantly high in the US, with a clear departure from what the random null model suggests. This measure of fragmentation is particularly high amongst the youngest groups. In the UK, modularity is substantially below the expected random values – the fact that it is so close to zero suggests that there is no evidence of fragmentation in how people select their news sources. The Spanish networks exhibit a similar lack of fragmentation. Overall, none of the networks we consider resembles the hypothetical case depicted by structure (3) in Figure 4A – all three audience networks are highly connected in a single component, with different levels of centrality and clustering.

4. Discussion

The methodological approach illustrated here has much to offer to the field of political communication. Understanding the structure of online news audiences is increasingly important: the turn to digital media for news has potentially profound implications for political knowledge, political participation, and civic engagement. Broadly, our approach affords systematic comparison of audience networks in three ways: (1) across countries, for cross-nationally comparative research that can help us avoid the risk of “naïve universalism” and generalizing from a case of one country; (2) across different audience groups that we might hypothesize will have significantly different ways of engaging with online news; and (3) over time, to determine

if the networks change substantially during the political cycle. In this paper we have analyzed audience structures around major political events, but the method applies equally well to other periods and it allows comparison between different stages of the political process, which can help advance our understanding of how certain events impact audience formation.

To illustrate our approach, we have relied on data from a third party provider. This data is only available at the aggregated level and, as previously highlighted for other type of proprietary data (e.g., Goldman, Mutz, & Dilliplane, 2013), it also presents some limitations for reproducible research due to the terms of use associated to the license. However, the panels we use are still more representative of the underlying population than most of the data accessed through the APIs that social media platforms provide (see Taneja, 2016 for a broad discussion on this). Moreover, online audience metrics are constantly audited by external companies that validate sampling and measuring processes for the advertisement industry. Still, future research should aim to consolidate alternative sources of trace data so that robustness of results can be tested.

Future research should also consider platforms other than the web to analyze news consumption patterns. The general approach to backbone extraction we apply here can also be applied to other forms of trace data, including the analysis of audience structures on different social media platforms like Facebook and Twitter. This exercise would provide further potential for comparing audience structures not only across countries, different groups, and over time, but also in different technological environments. Given the prominence of social media in granting access to news, and their walled-garden philosophy with respect to more open technologies like the web, analyzing news consumption patterns in these platforms should be a priority for political communication researchers. This, of course, requires the consolidation of channels that allow researchers to access the necessary data – a discussion on how to accomplish this is already taking place (e.g., King & Persily, 2018).

Our method provides a more sophisticated approach to the central issue of audience fragmentation, which is one of the core questions facing our field and also of increasing public interest. Our findings suggest that, despite the fears expressed in some quarters, “infinite choice” does not, in fact, “equal ultimate fragmentation” (Anderson, 2006, p. 181). To properly understand audience behavior in a changing media environment, including the degree of fragmentation, we need theoretical innovation (Bennett and Iyengar 2008) but we also need

methodological innovation. Many of the foundational questions in political communication research rests on issues of methods and measurement (de Vreese & Neijens, 2016), and this applies more than ever in an increasingly digital media environment – especially if we are to link audience behavior to media effects. Here, we have suggested one way of sifting through digital traces to identify meaningful patterns in news consumption. Our approach allows us to scale up the analyses and generalize the findings across political contexts. This comparative approach is necessary if we are to build theories that are applicable to diverse media environments.

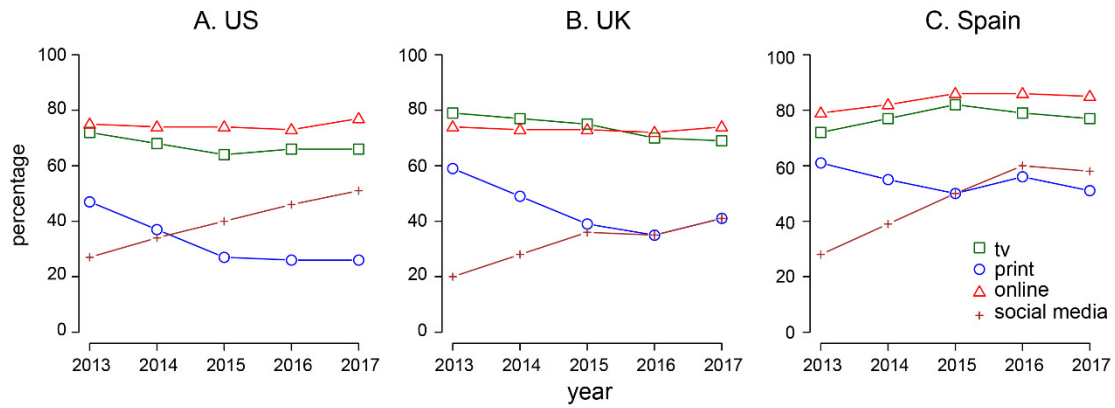
References

- Anderson, C. (2006). *The Long Tail. How Endless Choice is Creating Unlimited Demand*. London: Random House.
- Antunovic, D., Parsons, P., & Cooke, T. R. (2016). "Checking" and googling: Stages of news consumption among young adults. *Journalism*, 1464884916663625.
- Baum, M. a., & Groeling, T. (2008). New Media and the Polarization of American Political Discourse. *Political Communication*, 25(4), 345–365.
<http://doi.org/10.1080/10584600802426965>
- Bessi, A., & Briatte, F. (2016). Disparityfilter: Disparity Filter Algorithm for Weighted Networks. R package version 2.2.3, <https://CRAN.R-project.org/package=disparityfilter>.
- Castells, M. (2009). *Communication Power* (first). New York: Oxford University Press, Inc.
- Chadwick, A. (2013). *The hybrid media system: Politics and power*. Oxford University Press.
- Dahlgren, P. (2005). The Internet, public spheres, and political communication: Dispersion and deliberation. *Political Communication*, 22(2), 147–162.
- de Vreese, C. H., & Neijens, P. (2016). Measuring Media Exposure in a Changing Communications Environment. *Communication Methods and Measures*, 10(2-3), 69-80.
[doi:10.1080/19312458.2016.1150441](https://doi.org/10.1080/19312458.2016.1150441)
- Delli Carpini, M. X. (2004). Mediating Democratic Engagement: The Impact of Communications on Citizens' Involvement in Political and Civic Life. In L. Kee Kaid (Ed.), *Handbook of Political Communication Research* (pp. 395–434). Lawrence Erlbaum Publishers.
- Freeman, Linton C. 1979. "Centrality in Social Networks: Conceptual clarification." *Social Networks* 2 (3):215-239.
- Goldman, S. K., Mutz, D. C., & Dilliplane, S. (2013). All virtue is relative: A response to Prior. *Political Communication*, 30(4), 635–653.
- Hindman, M. S. (2009). *The Myth of Digital Democracy*. Princeton, NJ: Princeton University Press.
- Jamieson, K. H., & Cappella, J. N. (2008). *Echo Chamber: Rush Limbaugh and the Conservative Media Establishment*. Oxford: Oxford University Press.
- King, G. (2014). Restructuring the Social Sciences: Reflections from Harvard's Institute for Quantitative Social Science. *PS: Political Science & Politics*, 47(1), 165–172.
- Ksiazek, T. B. (2011). A Network Analytic Approach to Understanding Cross-Platform Audience Behavior. *Journal of Media Economics*, 24(4), 237-251.
[doi:10.1080/08997764.2011.626985](https://doi.org/10.1080/08997764.2011.626985)

- Majó-Vázquez, S., Cardenal, A. S., & González-Bailón, S. (2017). Digital News Consumption and Copyright Intervention: Evidence from Spain before and after the 2015 "Link Tax". *Journal of Computer-Mediated Communication*, 22(5), 284–301. doi:10.1111/jcc4.12196
- Mukerjee, S., Majó-Vázquez, S., & González-Bailón, S. (2018a). Networks of Audience Overlap in the Consumption of Digital News. *Journal of Communication*, 68(1), 26-50. doi:<https://doi.org/10.1093/joc/jqx007>
- Mukerjee, S., Majó-Vázquez, S., & González-Bailón, S. (2018b). Response to Webster and Taneja's Response to "Networks of Audience Overlap in the Consumption of Digital News". *Journal of Communication*, 68(3), E15-E18. doi:10.1093/joc/jqy022
- Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D., & Nielsen, R. K. (2017). *Digital News Report*. Oxford, UK. Retrieved from <http://www.digitalnewsreport.org/>
- Nicholls, T., Shabbir, N., & Nielsen, R. K. (2016). *Digital-Born News Media in Europe*. Oxford, UK.
- Norris, P. (2000). *A virtuous circle: Political communications in postindustrial societies*. Cambridge University Press.
- Pons, P., & Latapy, M. (2006). Computing Communities in Large Networks Using Random Walks. *Journal of Graph Algorithms and Applications*, 10(2), 191-218.
- Prior, M. (2007). *Post-broadcast democracy: How media choice increases inequality in political involvement and polarizes elections*. Cambridge University Press.
- Prior, M. (2009). The Immensely Inflated News Audience: Assessing Bias in Self-Reported News Exposure. *Public Opinion Quarterly*, 73(1), 130–143.
- Scharkow, M., & Bachl, M. (2017). How measurement error in content analysis and self-reported media use leads to minimal media effect findings in linkage analyses: A simulation study. *Political Communication*, 34(3), 323–343.
- Serrano, M. Á., Boguñá, M., & Vespignani, A. (2009). Extracting the multiscale backbone of complex weighted networks. *Proceedings of the National Academy of Sciences*, 106(16), 6483-6488. doi:10.1073/pnas.0808904106
- Shehata, A. (2016). News Habits Among Adolescents: The Influence of Family Communication on Adolescents' News Media Use—Evidence From a Three-Wave Panel Study. *Mass Communication and Society*, 19(6), 758–781.
- Stroud, N. J. (2011). *Niche News: The Politics of News Choice*. Oxford: Oxford University Press.
- Sunstein, C. R. (2009). *Republic.com 2.0*. Princeton University Press.
- Taneja, H. (2016). Using commercial audience measurement data in academic research. *Communication Methods and Measures*, 10(2–3), 176–178.

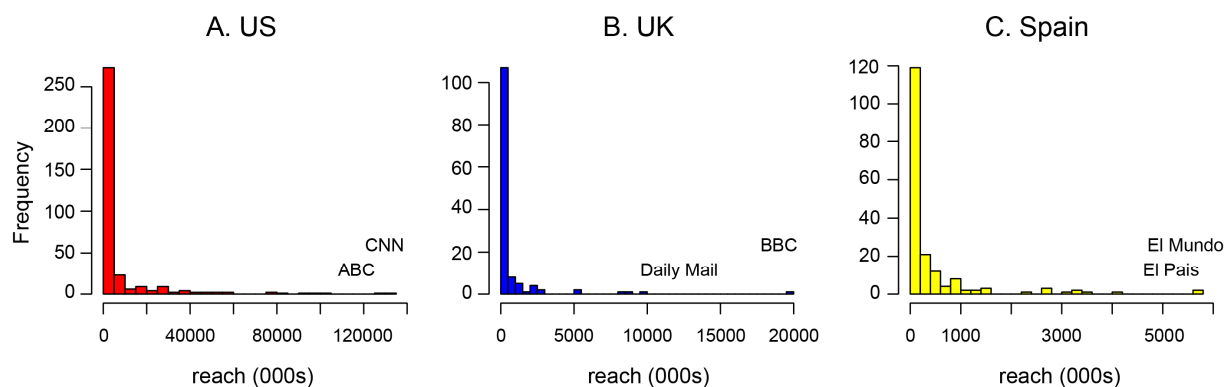
- Taneja, H. (2016). Mapping an audience-centric World Wide Web: A departure from hyperlink analysis. *New Media & Society*. doi:10.1177/1461444816642172
- Taneja, H., & Webster, J. G. (2016). How Do Global Audiences Take Shape? The Role of Institutions and Culture in Patterns of Web Use. *Journal of Communication*, 66(1), 161-182. doi:10.1111/jcom.12200
- Teixeira, F. (2018). SkyNet: Generates Networks from BTS Data. R package version 1.0.1, <https://CRAN.R-project.org/package=skynet>.
- 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. doi:10.1111/j.1460-2466.2011.01616.x
- Welbers, K., & van Atteveldt, W. (2016). SemNet: Semantic Network Analysis. R package version 0.102, <https://github.com/kasperwelbers/semnet>.

Figure 1. Main Sources of News in the US, UK, and Spain



Source: Reuters Digital News Reports. The question asked in the surveys was: “Which, if any, of the following have you used in the last week as a source of news?”

Figure 2. Total Audience Reach for News Sources in the US, UK, and Spain



Source: comScore. The histograms plot the total audience reach for the news sites classified by comScore under the category 'News/Information', which include both legacy and digital-born sites. The distribution of online visibility according to this measure is extremely skewed, with legacy news organizations at the right tail of the distribution.

Figure 3. Summary of the Audience Data Analyzed

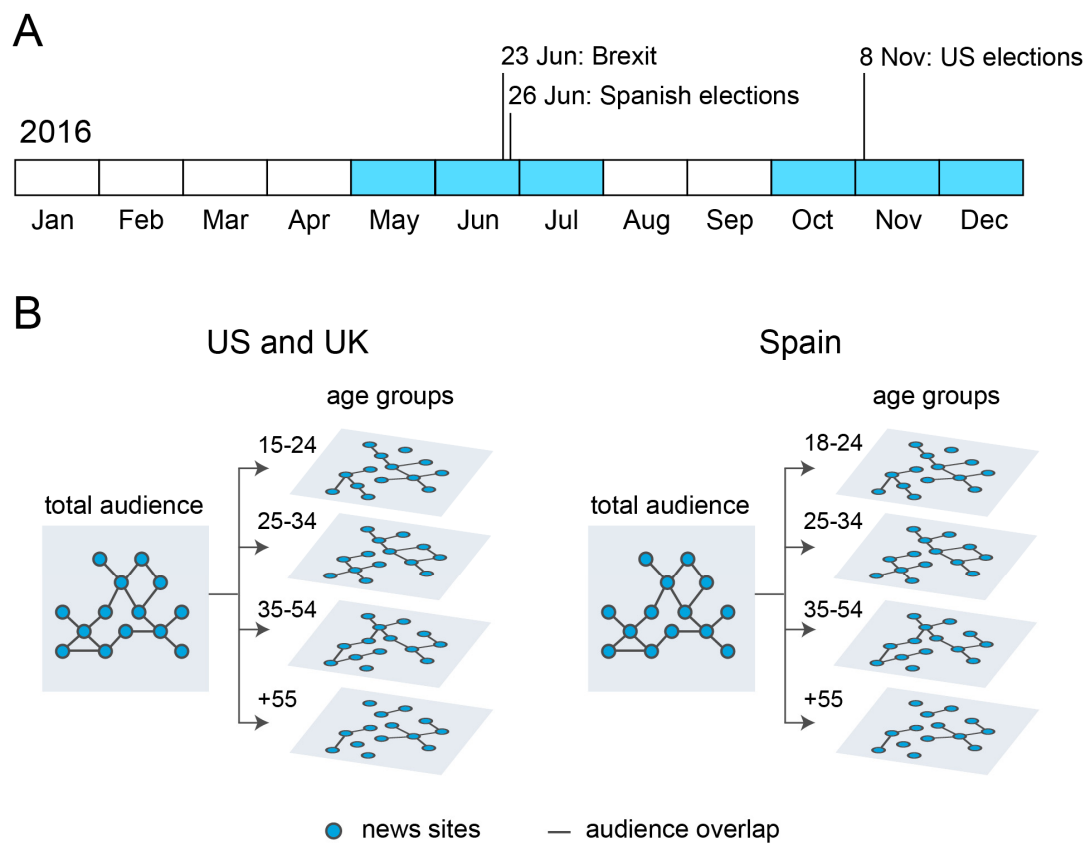


Figure 4. Schematic Representation of Backbone Extraction Technique

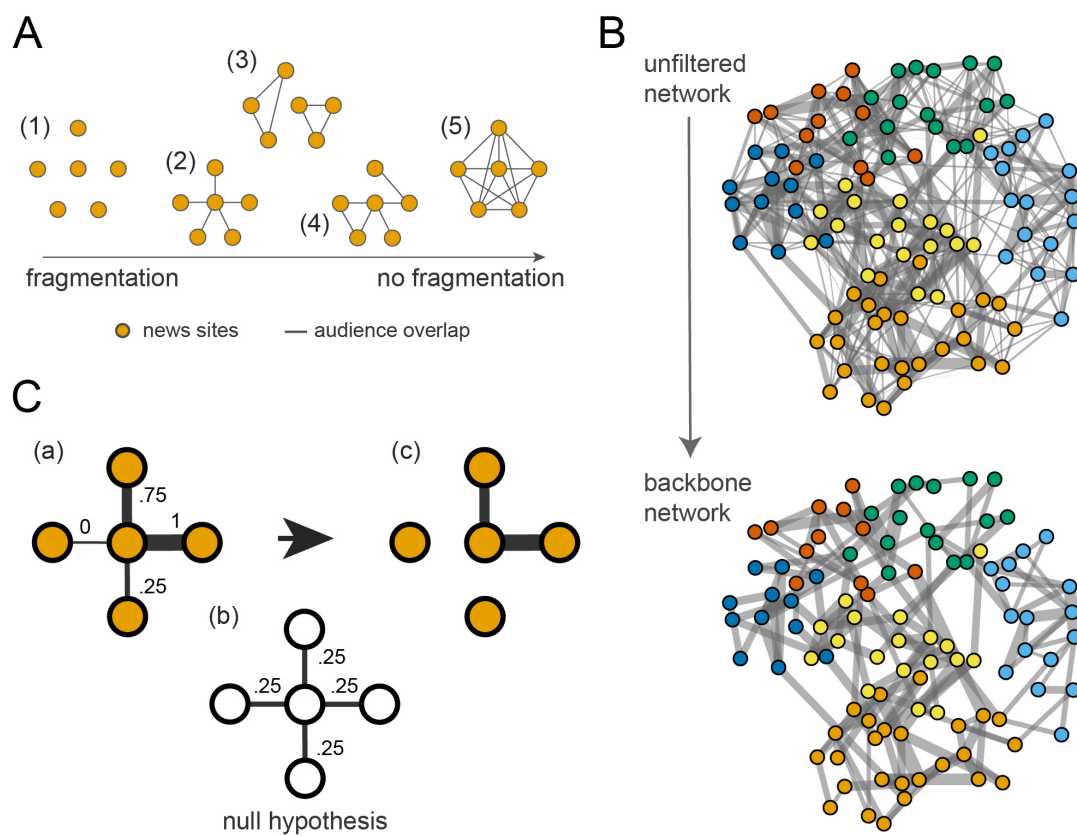
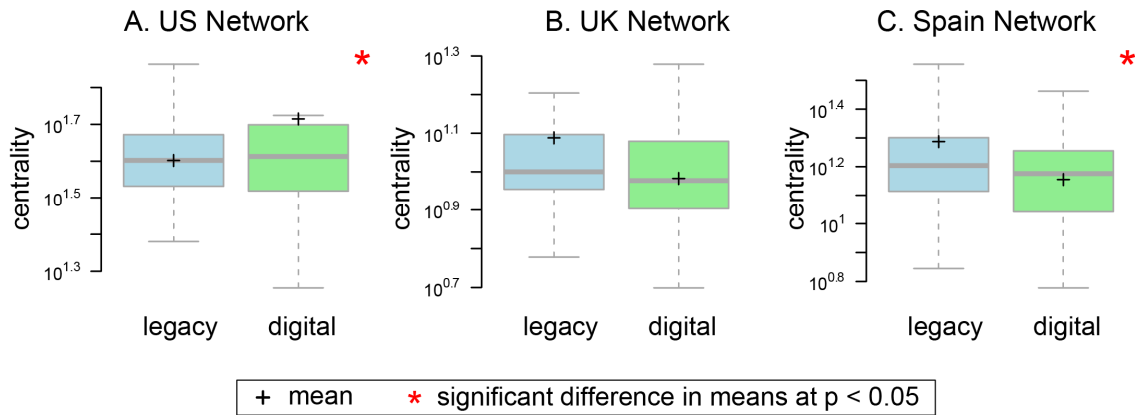
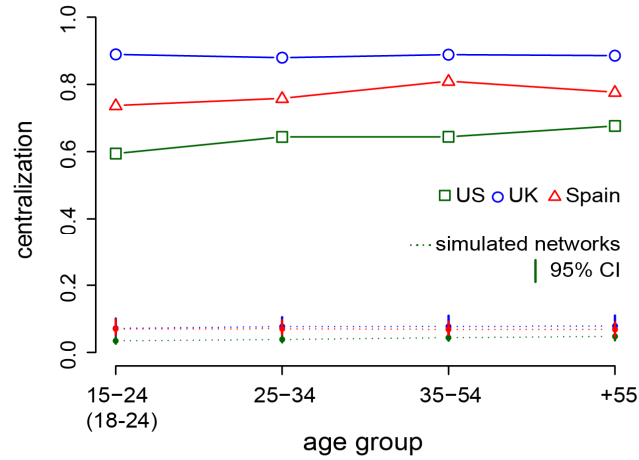


Figure 5. Differences in the Network Centrality of News Media Sites



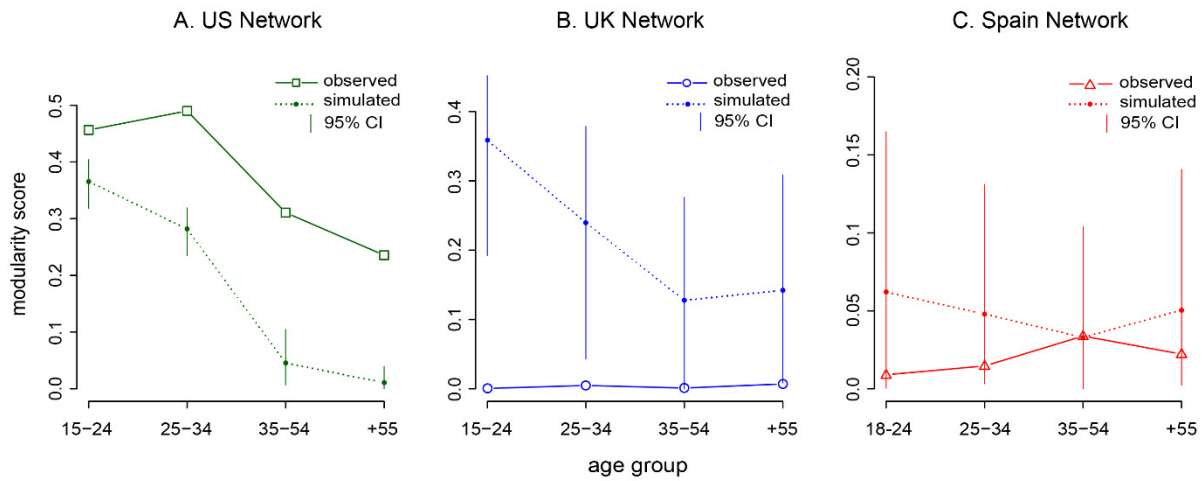
Note: outliers are not visualized; statistical significance is based on the Welch's t -test under the null hypotheses of no difference in means. Legacy media are significantly less central in the US network (the confidence interval in the log scale is CI: -0.22, -0.03) but more central in the Spain network (CI: 0.03, 0.23). Legacy media are also more central in the UK network, but in this case the difference is not statistically significant (CI: -0.03, 0.24). A bootstrapping test assuming unequal variance and applying the same probability threshold ($p < 0.05$) yields very similar results, with only slightly different confidence intervals.

Figure 6. Differences in the Network Centralization across Countries and Age Groups



Note: centralization measures the extent to which connections are concentrated around a few nodes in the network. This statistic can be interpreted as a measure of inequality or, in the context of our data, how spread audiences are in a media environment. The US network is the least centralized; the UK is the most centralized, signaling the influence of public broadcasting. There are no great differences across age groups but in all cases, centralization scores are substantially higher than those in random networks ($N = 1,000$), which preserve the same number of nodes and connections. The confidence intervals around simulated values (vertical bars) measure random variability, but they are so narrow that they are barely visible on this y-axis scale.

Figure 7. Differences in Network Modularity across Countries and Age Groups



Note: modularity measures the level of fragmentation in the networks (as defined by a random walk community detection method). These scores can be interpreted as proxies to audience self-selection. The US network is significantly fragmented: the modularity scores are substantially higher than those in random networks ($N = 1,000$, preserving the same size, density and degree sequence of the observed network). In the UK, the modularity scores are substantially lower than those expected by chance; they are in fact very close to zero, which means that there is no evidence of fragmentation in how audiences consume news. In the Spanish case, the modularity score is also very low but it is statistically insignificant.

Table 1. Statistics for Audience Overlap Networks before and after Backbone Extraction

	US		UK		Spain	
	before	after	before	after	before	after
Number of nodes	332	332	133	133	185	185
Legacy media		253		103		91
Digital-born media		79		30		94
Number of edges	53221	10979	6831	1200	13107	2390
Centralization	0.037	0.777	0.222	0.863	0.230	0.843
Degree correlation	-0.078	-0.724	-0.215	-0.682	-0.252	-0.636
Max degree	332	324	132	132	184	181
Min degree	234	0	21	1	46	2