Digital News in Spain: Characteristics and Effects of Online News Production and Consumption

Author: **Sílvia Majó-Vázquez**
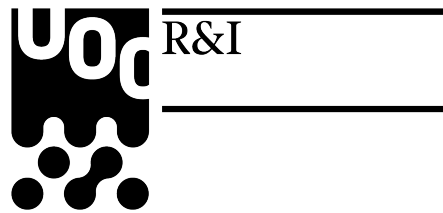Universitat Oberta de Catalunya (UOC)

Supervisor: **Ana S. Cardenal Izquierdo, PhD.**
Universitat Oberta de Catalunya (UOC)

A thesis submitted for the degree of
Doctor of Philosophy in Information and Knowledge Society

Barcelona
April 11, 2017

Universitat Oberta de Catalunya (UOC)

Doctoral Dissertation

Digital News in Spain: Characteristics and Effects of Online News Production and Consumption

Sílvia Majó-Vázquez

Supervisor: **Ana S. Cardenal Izquierdo, PhD.**
Universitat Oberta de Catalunya (UOC)

Thesis Committee: **Sandra González-Bailón, PhD.**
Annenberg School for Communication
University of Pennsylvania (UPenn)

Marta Cantijoch Cunill, PhD.
University of Manchester (UoM)

April 11, 2017

Sílvia Majó-Vázquez | majosilvia@uoc.edu

Digital News in Spain: Characteristics and Effects of Online News Production and Consumption

Doctoral Dissertation, April 11, 2017

Supervisor: Ana S. Cardenal Izquierdo, PhD.

Thesis Committee: Sandra González-Bailón, PhD. and Marta Cantijoch Cunill, PhD.

Universitat Oberta de Catalunya (UOC)

Rambla del Poblenou, 156

08018 - Barcelona - Spain

Abstract

This work presents a relational approach to the study of the structure of the online news domain from a macro- and micro-level perspective. Furthermore, it assesses the impact of the *digital news domain* on the *public agenda*. Firstly, it analyzes the online production and consumption of news by mapping two networks: the *network of links* connecting news media on the web; and the *network of audience flow* that users create when navigating web content. It determines that the provision of news is highly fragmented, with clear fault lines associated to media ownership structures. The audience network, which reveals no such fragmentation, is only slightly correlated to the media network. These findings have important implications for theoretical accounts of how digital technologies are reconfiguring the public sphere and bring novel empirical evidence that debunks claims of a fragmented online domain.

Secondly, it identifies the *news providers* that occupy the most power positions in the provision and consumption of news content. To this end, it measures the extent to which *new media* outlets -–those born on the web-– challenge the power monopoly held by traditional media brands in the offline sphere. By studying the network of news providers, as well as the patterns of audience navigation, it shows that legacy news media are still recognized as the most authoritative sources online, and only a few digital outlets occupy top authority positions in the flow of news content. Yet, when the analysis extends to identify *audience brokers*, i.e. those who have potential control over the *news audience* flow, the results show that digital outlets tend to occupy powerful positions among the young, even though legacy news media still retain much of the power in the audience network for the general population. These findings have implications for normative accounts of media function in democracy. Essentially, they suggest the roles of legacy media, e.g. *agenda setting* and *gatekeeping*, among others, have not yet been shared with new digital providers and, further, provide support for the idea that their offline reputation grants them strategic positions to have the control of audience attention online.

Finally, this study investigates whether the new digital news domain weakens citizens' consensus over the public agenda. While many extant studies of the role that digital

media plays on the public agenda contend that it undermines the public sphere, we argue that its impact depends on the type of media diet, and particularly on the level of concentration of online media diets. The type of news media diet online is also crucial to assess the association between the unlimited number of online media sources and two important dimensions of the public agenda: its *diversity* and *carrying capacity*. Using a combination of survey and browsing tracking data from a Spanish sample, we examine the relation between different types of digital news media diets and the diversity and capacity of the public agenda. We find that concentrated online news media diets are associated with more diverse agendas and a greater capacity to account for public problems. What is more important, though, is that, in contrast to various arguments, we do find common ground among digital audiences. Our results suggest that citizens agree on a set of common experiences when their online news diets are based on those outlets with higher audience shares in the news market. In line with our previous results, these types of *news providers*, i.e. legacy media, are central in most digital news diets. Only on the rare occasion that peripheral or small news providers are more predominant in news diets do people show lesser common ground and *agenda diversity*.

Overall the study contributes to the body of literature that debunks claims on the *fragmentation* of the digital domain, partly because as our results suggest legacy brands have not seen their power diminished when it comes to provide common ground for the public discussion. It also lends support to theories that highlight the positive role of the digital domain for enhancing the public debate.

Keywords: *digital news domain, news audience, news providers, fragmentation, media power, public agenda, agenda diversity, agenda capacity, network analysis.*

Resum

Aquesta tesi presenta un estudi relacional de la macro- i meso- estructura del domini digital de notícies. Així mateix, també analitza a nivell individual, l'impacte del domini digital de notícies en l'agenda pública. En primer lloc, s'estudien la producció i el consum de notícies en línia. Per fer-ho, es tracen dos tipus diferenciats d'estructures de xarxes: d'una banda, la xarxa de vincles que connecta els mitjans de comunicació a Internet; i de l'altra, es construeix un mapa que traça el flux de les audiències a la web i mostra com aquestes consumeixen el contingut que els mitjans de comunicació publiquen. L'anàlisi d'aquestes dues estructures mostra que la xarxa de proveïdors de notícies està altament fragmentada. Els nostres resultats suggereixen que el criteri que delimita els clústers que hi ha en aquesta xarxa responen als interessos i les estratègies de negoci de les empreses proveïdores de notícies. En la xarxa d'audiència però, les anàlisis no mostren que hi hagi fragmentació. Més important però, és el fet que els nostres resultats proven que les dues estructures, la xarxa d'audiències i la de proveïdors de notícies, només estan lleugerament correlacionades.

Les troballes d'aquest estudi tenen clares implicacions pel que fa a les anàlisis teòriques que fins ara, s'han fet sobre l'impacte de les tecnologies digitals en la reconfiguració de l'esfera pública. Cal subratllar, que els resultats d'aquest treball aporten evidències empíriques contràries a aquelles que sostenen que el domini públic digital està fragmentat.

En segon lloc, identifiquem els proveïdors de notícies que ocupen els llocs de poder més rellevants tan en la provisió de notícies com en el consum informatiu. Per assolir aquest objectiu, mesurem fins a quin punt, els mitjans de comunicació que en aquest estudi anomenem "nous mitjans" -–aquells que van néixer a la web i no tenen una versió fora d'ella– són un competidor dels mitjans tradicionals. La pregunta que ens proposem resoldre és: els nous mitjans de comunicació ocupen alguna de les posicions de més poder en el domini digital de notícies i per tan relleven als mitjans tradicionals dels seus rols com a elits del procés comunicatiu?

Per respondre a aquesta pregunta estudiem la xarxa de proveïdors de notícies i els patrons de consum de les mateixes. Quan analitzem les característiques estructurals

d'aquestes dues xarxes, els resultats mostren que els mitjans tradicionals encara mantenen el reconeixement com a fonts d'informació amb més autoritat a la web. De fet, els nostres resultats suggereixen que només existeix un petit número de nous mitjans que ocupen les posicions més altes en el rànquing de fonts d'informació amb més autoritat. Ara bé, quan analitzem el poder dels mitjans de comunicació en la xarxa d'audiència i hi identifiquem els *broker,* és a dir aquells actors que tenen el control potencials sobre el flux de l'audiència de notícies, els nostres resultats mostren que aquí sí, que els nous mitjans ocupen posicions de poder ens els segments joves d'audiència. No obstant això i en termes globals, els mitjans tradicionals són el que retenen el poder en la xarxa d'audiència que representa el consum de la població general. Aquests resultats tenen implicacions per les teories normatives sobre el funcionament de la democràcia. Bàsicament suggereixen que les funcions dels mitjans tradicionals i.e. *agenda setting* i *gatekeeping* entre d'altres, no són compartits amb els nous mitjans de comunicació. A més, aquests resultats també mostren que la reputació offline dels mitjans tradicionals és un element indispensable per explicar les posicions estratègiques que ocupen aquests en la xarxa d'audiència i que els permeten tenir control sobre ella.

Finalment, aquest estudi investiga si l'entorn digital debilita el consens ciutadà entorn a l'agenda pública. La recerca realitzada fins avui, sosté que l'entorn digital erosiona o fragmenta l'esfera pública. Nosaltres però, en aquest treball argumentem que el seu impacte dependrà del tipus de "dieta" mediàtica que cada individu tingui i més concretament, del nivell de concentració d'aquestes dietes. El tipus de dieta digital és també, un element cabdal per determinar la relació entre l'oferta il·limitada de fonts de notícies a Internet i dues dimensions bàsiques de l'agenda pública: la seva diversitat i la capacitat individual per identificar problemes públics. En aquesta part del nostre estudi fem servir una combinació de dues fonts de dades. D'una banda, utilitzem dades d'enquesta i de l'altra, dades procedents del monitoreig del consum de notícies digitals. Ambdues estan basades en una mostra d'individus espanyols i ens permeten analitzar la relació que hi ha entre els diferents tipus de dietes informatives digitals i la diversitat de l'agenda i la capacitat dels individus per esmentar qüestions importants en el debat públic.

Els nostres resultats mostren que les dietes digitals concentrades estan associades amb una agenda més diversa i amb una major capacitat per enumerar problemes públics. El més important però és que, contràriament als arguments existents en una part de la recerca acadèmica, les nostres anàlisis suggereixen que els ciutadans conflueixen alhora d'esmentar problemes comuns quan les seves dietes informatives digitals es basen en aquells mitjans que tenen un major control del mercat de notícies. En línia amb els nostres resultats previs, aquestes anàlisis identifiquen els mitjans tradicionals com a centrals en les dietes informatives digitals dels ciutadans. Tan sols en aquelles ocasions, poc freqüents, que nous mitjans digitals petits i perifèrics

són predominants en les dietes mediàtiques dels individus, aquests tendeixen a tenir menys capacitat per identificar els problemes públics i mostren una menor diversitat en l'agenda.

Cal destacar que aquest treball suposa una contribució rellevant respecte a la literatura existent perquè aporta noves evidencies per refutar les anàlisis que sostenen que el domini digital està fragmentat. Això es degut parcialment, al fet que el mitjans tradicionals, segons el nostre estudi, no han vist reduït el seu poder i continuen sent un element essencial per a la cohesió de l'espai públic. A més, les anàlisis que hem realitzat mostren que el domini digital contribueix a enriquir el debat públic quan els mitjans tradicionals són centrals en les dietes informatives.

Paraulas clau: *domini de notícies digital, mitjans de comunicació, audiència de notícies, proveïdors de notícies, fragmentació, poder mediàtic, agenda pública, diversitat de l'agenda, capacitat de l'agenda, anàlisi de xarxes.*

Resumen

Esta tesis presenta un estudio relacional de la macro- y meso- estructura del dominio digital de noticias. Además, analiza el impacto del dominio digital de noticias en la agenda pública. En primer lugar, se estudia la producción y el consumo de noticias digitales con el objetivo de medir la fragmentación en la esfera pública. Para ello, se construyen dos tipos diferenciados de estructuras de redes: por un lado, la red de hipervínculos que conectan los medios de comunicación en Internet; y por el otro, se reconstruye el mapa que traza el flujo de audiencias de noticias en la web y muestra como éstas consumen el contenido de los medios de comunicación. El análisis de estas dos estructuras prueba que la red de proveedores de noticias está altamente fragmentada. Los resultados sugieren que el criterio que delimita los clústeres que hay en esta red responden a los intereses y a las estrategias de negocio de las empresas proveedoras de noticias. Sin embargo, en la red de audiencia, nuestros resultados demuestran que no existe tal fragmentación.

El hallazgo más significativo del presente estudio es que las dos estructuras (i.e., la red de audiencia y la red de proveedores de noticias) están solo ligeramente correlacionadas. Estos resultados tienen claras implicaciones para los análisis teóricos realizados hasta el momento sobre el impacto de las tecnologías digitales en la reconfiguración de la esfera pública. Los resultados aquí expuestos aportan evidencias empíricas contrarias a aquellas que sostienen que el dominio público digital está fragmentado.

En segundo lugar, en este trabajo se identifican los proveedores de noticias que ocupan los espacios de poder más relevantes tanto en la provisión de noticias, como en el consumo de las mismas. Para alcanzar este objetivo, medimos hasta qué punto los medios de comunicación que denominamos "nuevos medios" -–aquellos nacidos en la web y que carecen de una versión fuera de ella– son competidores de los medios tradicionales. En resumen, pretendemos responder a la siguiente pregunta: ¿Los nuevos medios de comunicación ocupan las posiciones de poder en la esfera digital que venían ocupando los medios tradicionales en la esfera *offline*?

Para responder a esta pregunta, estudiamos de nuevo la red de proveedores de noticias y los patrones de consumo de la audiencia. Cuando analizamos las carac-

terísticas estructurales de estas dos redes, nuestros resultados nos permiten afirmar que los medios tradicionales todavía mantienen su reconocimiento como fuentes de información con más autoridad en la web. Más concretamente, nuestros resultados sugieren que sólo existe un pequeño número de nuevos medios que ocupan las posiciones más altas en el ranking de fuentes de información con más autoridad. Sin embargo, cuando analizamos el poder de los medios de comunicación en la red de audiencia e identificamos a los *brokers* dentro de dicha estructura (i.e., medimos qué actores tienen el control potencial sobre el flujo de audiencia de noticias) los resultados sí muestran que los nuevos medios ocupan posiciones de poder, aunque únicamente en los segmentos de audiencia jóvenes. En términos globales, la evidencia parece robusta al señalar que los medios tradicionales retienen el poder en la red de audiencia que representa a la población general.

Estos resultados tienen implicaciones para las teorías normativas sobre el funcionamiento de la democracia. Básicamente, sugieren que las funciones de los medios tradicionales (e.g., *agenda setting*, *gatekeeping*) no resultan en general compartidos con los nuevos medios de comunicación. Así mismo, los resultados también sugieren que la reputación *offline* de los medios tradicionales es un elemento importante para explicar las posiciones estratégicas que ocupan este tipo de medios en la red de audiencia y que les permiten ejercer el control sobre ella.

Finalmente, este estudio investiga si el entorno digital debilita el consenso ciudadano sobre la agenda pública. Las investigaciones realizadas hasta ahora sostienen que el entorno digital erosiona o fragmenta la esfera pública. No obstante, en este estudio argumentamos que su impacto depende del tipo de "dieta" mediática que caracteriza a cada individuo y más concretamente, del nivel de concentración de estas dietas. El tipo de dieta digital resulta también fundamental para determinar la relación entre la oferta ilimitada de fuentes de información en Internet y dos dimensiones básicas de la agenda pública: su diversidad y la capacidad de los ciudadanos para identificar problemas públicos. En esta fase de nuestro estudio combinamos dos tipos distintos de fuentes de datos: datos procedentes de encuestas, por un lado, y datos procedentes del monitoreo del consumo de noticias digitales, por el otro. Las dos fuentes de datos proceden de una muestra de individuos españoles y nos permiten analizar la relación existente entre los diferentes tipos de dietas informativas digitales, la diversidad de la agenda y la capacidad de los individuos para identificar temas importantes del debate público.

Nuestros resultados sugieren que las dietas digitales concentradas están asociadas con una agenda más diversa y con una mayor capacidad para nombrar problemas. Lo más relevante es que, según nuestros hallazgos, y en contra de los argumentos hasta ahora esgrimidos por una parte de la literatura académica, los ciudadanos confluyen en un conjunto de temas que identifican como problemas de la sociedad

cuando sus dietas se basan en los medios que tienen un mayor control del mercado de noticias. En consonancia con nuestros resultados previos, esta evidencia señala a los medios tradicionales como los actores centrales de las dietas informativas digitales. En los casos –menos frecuentes– en que los nuevos medios digitales, más pequeños y periféricos, son predominantes en una dieta, los individuos tienden a tener menor capacidad para identificar problemas públicos y su agenda es también menos diversa.

Cabe destacar la contribución relevante de este trabajo al aportar nuevas evidencias que parecen refutar los hallazgos recogidos en la literatura existente y que han permitido sostener que el dominio público digital está fragmentado. El hecho que los medios tradicionales, según nuestro estudio, no hayan visto mermado su poder y que continúen siendo un elemento fundamental para la cohesión del espacio público, explica la falta de fragmentación que revela nuestro análisis. Cabe destacar asimismo que el dominio de noticias digital, según nuestros análisis, contribuye a enriquecer el debate público cuando los medios de comunicación tradicionales son centrales en las dietas informativas.

Palabras clave: *dominio digital de noticias, audiencia, provedores de noticias, fragmentación, poder mediático, medios, agenda pública, diversidad de la agenda, capacidad de la agenda, análisis de redes.*

Acknowledgements

Treballaven a la fàbrica cosint jerseis, al taller passant cables de llum, al forn del vidre bufant bombetes, de lampista a Gibraltar i a casa. Treballaven i treballaven. Els primers, per poder menjar cada dia i els segons, per pagar-nos l'educació. Tossudament convençuda ella que així podria donar-nos un futur millor. A tots ells que durant dècades es van llevar cada dia, abans que el sol sortís i que tornaven negre nit. A tots ells, que van ser i són resistents com ningú altre que conegui, davant els reptes de la vida. A tots ells, els hi dedico aquest treball: la meva mare, el meu pare, els meus avis i les meves àvies. Gràcies per ensenyar-nos a les dues a fer camí. Gràcies per fer-nos-el infinitament més confortable que el vostre. Gràcies per ajudar-me a arribar fins aquestes línies que per mi tenen tant de sentit.

Perquè sempre ha vist què hi havia més enllà, perquè hi ha cregut quan jo ni tan sols hi pensava, perquè m'ha ensenyat a somiar realitats i ha sostingut els seus fonaments perquè creixessin fermes. Perquè m'ha ensenyat a estimar la recerca. Sense tu, no hauria estat possible. Gràcies Agustín Alemán González.

Gràcies per donar-me l'oportunitat de treballar plegades, per la generositat en compartir el teu temps i per la confiança que trauries el millor de mi per fer recerca, quan encara tot estava per fer. A la meva directora de tesi, la professora Ana S. Cardenal, moltes gràcies. Per acompanyar-me en les primeres passes en la recerca, potser les més ingrates, per ajudar-me sempre que t'ho he demanat al llarg d'aquests tres anys i escaig, i per perseverar fins a l'última línia d'aquest manuscrit perquè assolíssim l'objectiu marcat. Ha estat un privilegi haver fet aquest camí plegades.

Per ensenyar-me a pensar i pensar en gran. Pel teu valuós temps. Per perfilar amb una destresa única idees vagues i convertir-les en projectes de futur. Per compartir sempre el teu coneixement amb generositat i passió. Per transmetre la confiança que tantes vegades t'abandona en un camí com aquest. Gràcies a la professora Sandra González-Bailón, membre del comitè d'aquesta tesi i font d'inspiració.

I owe a debt of gratitude to the *Annenberg School for Communication* at the *University of Pennsylvania*, especially to the Dean, professor Michael Delli Carpini, the executive assistant to the Dean, Kelly Fernández, and the administrative assistant, Marcus

Mundy. *Annenberg* gave me two unique opportunities in my life to learn from the heart of the Communication research community. I am very grateful for your generosity in giving me those chances. This project has benefited immeasurably from these opportunities, and they have already had a great impact on my research career. Thank you very much.

Darrere d'aquestes pàgines hi ha tres anys i mig de treball en el que totes i cadascuna de les persones que hi ha a continuació, hi han jugat un rol, gran o petit, però important en algun moment. Moltes d'elles fins i tot, molts anys abans que el camí comencés. A totes, el meu profund agraïment. Amb vosaltres he après; m'heu inspirat; m'heu ajudat a fer quelcom que no sabia; heu compartit algun coneixement tècnic dels molts que he hagut d'aprendre; heu confiat en el projecte; heu volgut simplement, saber què era "això de la tesi" i així, per un instant, heu recorregut amb mi una part d'aquest llarg camí; heu finançat el projecte, o part d'ell, quelcom imprescindible perquè fos realitat; m'heu trucat quan els silencis eren molt llargs; heu eixugat alguna llàgrima; i m'heu arrencat rialles quan estava capficada.

A tots i a totes moltes gràcies (ordre alfabètic): Robert Ackland; Jordi Aguilar (la seva generositat i els seus coneixements tècnics han sigut cabdals en aquest projecte. L'agraïment és infinit) ; Sílvia Aguilar ; Elvira Altés; Lídia Arroyo; Anna Barniol; Joshua Becker; Biblioteca de la UOC ; Rosa Borge; Javier Borge-Holthoefer; Debon Brackbill; Hannah Brenneman; Marta Cantijoch (Gràcies per ajudar-me a començar i acabar aquest camí! Hi has estat en els moments imprescindibles i malgrat el poc temps que tenies. Moltíssimes gràcies); Nereida Carrillo; Laura Chaqués; Gabriel Colomer; Pol Colomer de Simón (Gràcies per confiar en una proposta molt vaga i dedicar-hi el vostre temps i la passió pel que feu); Companys i Companyes de la Cambra de Comerç de Barcelona (especialment, al seu president i la cap del Gabinet de Presidència); comScore ; Joana Conill; Alba Crespo Rubio; Camilo Cristancho (¡Cuánto he aprendido contigo! Gracias); Thomas Deneman; DiMeNeTTers; Escola d'Adults Maria Saus de Canet de Mar (especialment a la millor germana del món, Mònica Majó Vázquez i l'Emma Pazo); Escola de Doctorat de la UOC ; Agnès Esquerra; Estudis de Dret i Ciència Política de la UOC; Carol Galais; Aina Gallego; Joana Gallego; Albert Genís Juanola; Katerina Girginova; María del Carmen González; Idom (David Zaragoza & Xabier Ruiz); Internet Interdisciplinary Institute; Olga Llopis ; Susana López León (Per la paciència i el compromís. Per ensenyar-me quan això encara estava a les beceroles. Gràcies); Amèlia López Quiroga; Anna Mercadé; Laura Morató; Netquest; OSRT-UOC; Mònica Pérez Clausen; Melani Revilla; Ester Romeu Marín; David Rosell (ha escoltat en moltes ocasions, idees vagues sobre aquesta tesi. Sempre ha tingut la paciència i generositat d'ajudar-me a donar-hi la forma adequada perquè poguessin arribar a ser petites bones idees); Marta Rosés; Joan Ramon Rovira; Maria Rovira Montsant (Gràcies per haver-te tingut a prop durant aquests anys i per tot el que hem gaudit cada estona —massa poques– que hem passat plegades!); Oleguer Segarra (Gràcies pel

" *[. . .] he told him the four sentences that lead to wisdom. He'd said them only once, never repeating them. But once had been enough for Gamache.*

I'm sorry. I was wrong. I need help. I don't know.

— **Penny L.**
Bury Your Dead
A Chief Inspector Gamache Novel (2011)

Contents

Introduction

Is there fragmentation in the online news domain? And, if so, does it foster frag-
mented societies? To answer this research question, this thesis, first, offers a re-
lational approach to the study of the online news domain and, secondly, assesses
its impact on the public agenda. To determine whether the digital news domain
is or is not a fertile ground for a fragmented citizenry, we look at the relations
between information providers and news consumers at three different levels: at
the macro-level, where we identify clusters (or the lack of thereof) of news media
and digital audience; at the meso-level, where we consider the relocation of roles
amongst legacy and new media outlets and their control over the news domain; and
at the individual level, where we assess the extent to which people share common
ground for public debate by looking at and how different types of digital news diets
relate to their perceptions about salient issues.

We thus propose a comprehensive analysis of the fragmentation in the digital age that
focuses on the characteristics of the online news domain and its potential to make
people come together around public issues. Beyond the specific characteristics of our
case study, conducted in Spain, our empirical framework represents an important
contribution for the current research in that it can be applied regardless of the
specificities of national contexts to understand the structure and impact of digital
news domains universally. Further, and more importantly, our findings contribute to
reassessing theoretical debates about the pernicious effects of digital technologies
for democracies.

Fears about negative consequences of digital technologies on societies have recur-
rently emerged since the advent of the web (Chaffee & Metzger, 2001; Napoli, 2008;
Papacharissi, 2002; Pariser, 2011b; Sunstein, 2009; Turow & Tsui, 2009). In the
backdrop of these interpretations lays the argument that the fragmentation in the
digital news domain segregates societies into isolated enclaves that do not share
common ground for the discussion of public issues. Supporters of such stances agree
on the diagnostic but differ in their views on the actual mechanisms that relate the
current digital news domain to social fragmentation. Some scholars have equated
the emergence of the myriad of news sources on the web to the fragmentation of
audiences at the same level. The higher the number of news sources, the higher the
number of isolated clusters around which audiences will be organized, according

to their claims. Undoubtedly, the number of news sources available on the web is now unlimited. The digital domain has drastically lowered the cost for publishing and distributing news. Theoretically, it has put on equal footing legacy news organizations and new digital news providers (Castells, 2009), although we will bring new evidence that challenges these claims. What is also not so clear is whether the expansion of news sources has caused a segmentation of the society at the same level; we will elaborate on this in the following pages.

Parallel to the creation of hundreds of news sources, the web has also represented a great shift in the provision of the news because it has promoted the production of individually tailored information and the micro-specialization of news providers. Some scholars have interpreted these changes as another source of fragmentation (second view). Their underlying argument is that the specialization of news sources only attracts homogeneous groups of users (Sunstein, 2009; Tewksbury, 2005), and that the affordances of the web to filter content and match individual preferences fosters the emergence of social enclaves (Pariser, 2011a). Hence, they conclude, the web fragments societies.

Yet, it is not only the multiplication of sources and their digital features that have fed the arguments of those contending that the digital news domain causes social fragmentation, but the multiplication of issues in the public agenda has also nutured theses views. The very concept of public agenda would be in danger due to the current structure of the news domain, according to a related research line in this field. Because the number of sources has dramatically increased in the digital sphere, so has, or at least they argue, the number of salient issues that compete to draw public attention (Neuman, Guggenheim, Mo Jang, & Bae, 2014). The space and time for news content were preciously limited before the advent of the web. This constraint paved the way for journalists and news organizations to converge on a set of criteria that helped them to put order in the social life or to set the agenda, as McCombs and Shaw (1972) and many others have extensively proved. At that time, news media filtered and presented content in a similar fashion. As a result, there was a high redundancy in the agenda of those outlets predating the Internet era.

The digitalization of the news provision combined with the growing role of social platforms -—which we will discuss shortly-— has greatly extended the time and space for providing news content. What's more, they have shaken the well-established journalistic routines for news production and distribution. But returning to our main discussion, if time and space are now unlimited, there is no evident need for prioritizing a small set of popular issues. This is precisely what some scholars sustain to equate the unlimited digital space for news information with an unlimited number of newsworthy issues and, hence, the emergence of an endless public agenda. Should that be the case -—disregarding for the sake of their argument, the important

role that limited levels of public attention play in this process—- the digital news domain would make the concept of public agenda problematic (Boynton & Glenn W Richardson, 2016; Chaffee & Metzger, 2001). In other words, if the increase in salient issues has mirrored the growth of the number of digital sources, there would be no way to achieve and maintain a structured social debate around an ordered list of public priorities, i.e. a public agenda. Consequently, the new digital ecosystem would be pushing societies away from the common ground that holds them together and would make the public deliberation -—and, thus, the existence of democracies-— very difficult, if not impossible to sustain.

So far, though, the evidence in this respect -—that we will review in-depth in this thesis—- is not compelling. Contributions by those who have brought less pessimistic analyses about the consequences of the digital news domain over the public life work against the reviewed arguments above. On one hand, there is a growing body of literature that suggest that people still have a pretty similar news diet at the center of which branded news providers are located (Flaxman, Goel, & Rao, 2016; Gentzkow & Shapiro, 2011; Weber & Monge, 2011; Webster & Ksiazek, 2012). According to these scholars, legacy news media keep playing a leading role in the digital news domain and dominate people's attention. The multiplication of sources, thus, seems to have functioned more as a complement to than as a substitute of the news providers from the broadcast democracy (Prior, 2007), i.e. newspapers, commercial televisions and radios.

On the other hand, the dangers of the multiplication of issues in the public domain and its negative consequences over the very existence of a public agenda would also mitigate thanks to the fact that legacy outlets still lead the public debate (Coleman & McCombs, 2007; Nielsen & Schrøder, 2014). More precisely, the explosion of news content and public issues, due to the multiplication of channels of information, would vanish in favor of popular issues that legacy media put on the agenda (Lee, 2007; Messner & DiStaso, 2008; Park, Ko, Lee, Song, & others, 2013).

Although more optimistic, these latest arguments highlight an important conclusion: the fragmentation hypothesis is still open and unresolved. It is necessary to bring new evidence that harnesses more refined measures to understand the relationship between the digital news ecology and the public domain. There is an evident lack of consensus in this field that speaks for the importance of continuing to drill down on this relationship. Part of the literature summarized above presents a worrisome scenario for the future of democracy. Meanwhile, a more recent body of research provides counterarguments that give way to cautious optimism. Altogether, though, these claims prove that there is not enough evidence to get conclusive explanations about the extent to which the digital news domain fragments the public sphere. Our

work aims to contribute to this debate with novel measures, methods and data (from Spain) and by tackling fragmentation at three different levels.

At the macro-level (Chapter 2), fragmentation relates to both the supply and the demand side of the news domain. By simultaneously analyzing the two dimensions of the news domain, we overcome pitfalls of previous studies that have made assumptions on the levels of fragmentation of the news provision or the audience behavior by solely analyzing just one of these dimensions. To the best of our knowledge, no study so far has focused on comprehensively and simultaneously understanding the relation of users and producers of information in the online news domain.

We complement the study of the macro-structure of the news domain by identifying the power of news providers in the digital sphere (Chapter 3, meso-level analysis). Our interest is on the individual role of the two types of news providers that are present in the digital domain as defined in our study: legacy and new media. The former predates the Internet era and, according to some scholars, have seen their power diminished with the advent of the web. We test here those claims. The latter are born-of-the-web news sites that combine, to a greater or lesser extent, the aggregation of news content from several sources with their own production of information. The articulation of a clear definition of digital news media is still a challenge in the literature. We will propose a working definition to understand the difference that exists among brand sites and new media. It will take into account the advantages that new media have due to their deeper understanding of the digital technology and their greater engagement with news audiences. Yet, the question is whether these distinctive abilities grant them greater power than legacy news media. This question has long been posed, but it is still unresolved. Nowadays, it attracts growing interest due to the struggles that legacy brands face to successfully tackle the digital domain. Our research aims to shed light on this question because it helps us to reassess claims about the fragmentation of the public domain.

We use a definition of power that relates to both the level of authority that news media have and their potential control over the audience flow. Our approach contributes to unveiling the distinctive functions of news providers in the digital news ecology at the meso-level, and it offers a useful lens for understanding the extent to which legacy news media can guarantee common ground for democratic societies and, hence, preserve their traditional role in the digital sphere. Furthermore, this meso-level analysis contributes to set the boundaries of the impact that the new technologies have had on the reconfiguration of the media ecology. It disentangles the relationship between branded media outlets and born-of-the-web news providers and helps to revisit theoretical claims about the fragmentation of the digital domain.

Finally, at the individual level or micro-level (Chapter 4), we determine whether there is evidence of the so feared social fragmentation by measuring the extent to which people share common ground for public debate. In this analysis, fragmentation relates to whether or not people share a similar ranking of public concerns that enables them to participate in civic life. This approach serves as a proxy to measure the capacity of the digital news domain to strengthen or weaken democratic citizenry. The novelty of our analysis is that it hinges upon types of media diets. In other words, it measures the importance of legacy and digital news media on the overall individual news consumption. This is a necessary step for tackling the nuances of the relationship between digital news consumption and public fragmentation.

Furthermore, and since the web allows for tracking the patterns of news consumption, we do not only use reported information, i.e. survey data about news consumption, in our analysis, but also browsing tracking data that contributes to a more refined study. We rely on the best proxy available to map aggregated news consumption online and then we also track the behavior of a panel that previously agreed to share their browsing data. The richness of this data is an important asset of the present study. We argue that this is a complementary approach to the conventional and still very necessary analysis based on survey data which makes a fundamental contribution to the extant research in that it allows us to conduct a more nuanced analysis of the digital news consumption and the fragmentation hypothesis.

By looking at fragmentation at the micro or individual level, this thesis goes one step further than previous research in the field in relating the structural and organizational characteristics of the news domain to actual opinions, i.e. to the very existence of shared common ground for the functioning of democracies.

In concluding, a note is necessary: in this study we acknowledge the importance of social media for the news consumption and distribution, but we do not include them in our empirical analysis. Social media platforms, like *Twitter* or *Facebook*, are prominent actors in this changing news media environment (e.g., Bakshy, Messing, & Adamic, 2015). People increasingly rely on them to keep up with news: This is the case for 39% of the online population in Spain, a higher proportion than in Germany (21%) and the UK (25%), and lower than in Portugal (41%) and Turkey (48%) (Newman, Levy, & Nielsen, 2016). Yet our definition of news providers does not include them because they are not news organizations and do not abide by the same editorial guidelines as the type of news providers that we do study here.

Moreover, there are several other important reasons that justify our decision to solely focus our research efforts on the structure of the online news providers, i.e. legacy media and new media. First, newspapers are still more widely used and seen as more important sources of news online than social media, according to a study including

8 countries like the United Kingdom, France, Japan, Germany and Spain (Nielsen & Schrøder, 2014). These findings have been supported by most recent reports on news consumption in the online domain (Mitchell, Gottfried, Shearer, & Lu, 2017). Currently, the main news usage remains with brand media sites that "have a strong news heritage and have been able to build up a reputation over time" (Newman, Fletcher, Levy, & Nielsen, 2016:9). Secondly, predictions for 2017 show that the proliferation of fake and inaccurate news on social media platforms will strengthen the position of legacy news media (Newman, 2017). Several studies recognize an important future role of legacy news brands in the digital domain as a source of credibility (KantarMedia, 2016).

Thirdly, previous attempts by social media platforms to humanly edit trending stories, hence directly playing a traditional media role, were quickly ruled out to reapply purely algorithmic curation processes (Dewey, 2016; Facebook, 2016; Nunez, 2016) based on a combination of volume of attention, momentum, and, more importantly, personal preferences. Such filtering processes greatly differ from journalistic criteria, thus reaffirming our previous argument that social media are not news organizations. Furthermore, *Facebook* recently decided to use third-party journalistic organizations, mainly new media, and fact checkers to fight fake news (Mosseri, 2016; Sharockman, 2016; Snopes, 2016), a movement that clearly signals the distinctive role of the news media in the news production, which, so far, has not been replaced by social platforms.

Finally, the *European Commission* has recently announced a proposal to update the copyright law for the single market (European Commission, 2016). This proposal mirrors the Spanish "link tax" law passed in October 2014 that led to *Google* to shut down its News services in the country and remove Spanish media outlets from the service (Google, 2014; Rushe, 2014). The European proposal, which might not apply to social media platforms following the Spanish example (European Commission, 2017; Ministerio Educación, 2014), but to news organizations and aggregators -—see more on this discussion in our conclusions-— might increase the fragmentation of the web by raising the cost to send links. If European policymakers steer this policy course they might threaten the very existence of new media outlets (Méndez, 2017), whose main activity is the aggregation of news content. Previous evidence suggests that these types of news providers promote news consumption and cross-cutting views by linking content from several outlets and hence, from a normative point of view, play an important role for the functioning of democracy (Athey, Mobius, & Pal, 2012; KantarMedia, 2016).

All of these arguments justify again the decision to focus our research efforts on the impact that the digital news domain, defined as the structure where legacy and new media operate, play on social fragmentation. Finally, Table 1.1 and Table 1.2 provide

a summary of this thesis. The former shows the goals and the main findings and the latter presents the data and methods used in each empirical chapter.

Tab. 1.1.: Summary of the Research Goals and Findings

	Chapter 2	Chapter 3	Chapter 4
Goal	Analysis of the macro structure of the online news domain.	Analysis of the meso structure of the online news domain.	Analysis of the impact of the online news domain on the public agenda at the micro o individual level.
Description	In this chapter, the online news domain is operationalized by means of the network of news providers and the network of the audience for news. The study uses community network analysis as a method to identify fragmentation in the online news domain.	This chapter offers an examination of the structural position of legacy and new digital outlets in the online domain. It identifies authoritative sources and media brokers in the audience network using network science tools.	This chapter investigates whether the online news domain weakens citizens' consensus over the public agenda. It examines level of concentration of news media diets online and assesses the association between the unlimited number of online media sources and the public agenda. The diversity and capacity of the public agenda are studied too.
Research Question (RQ) or Hypothesis (H)	**RQ2.1:** *Is there fragmentation in the network of news providers?* **RQ2.2:** *Is there fragmentation in the network of the audience for news?* **RQ2.3:** *Does link activity by news providers shape audience navigation behavior?* **H2.1:** *The structure of the online news and audience networks will be significantly correlated if the audience responds to the connections created by media.*	**RQ3.1:** *To what extent are new digital outlets, by being recognized as authoritative sources of information, central actors in the flow of news content?* **RQ3.2:** *To what extent do new digital outlets and legacy organizations control brokerage relations in the audience network?* **RQ3.3:** *Is media brokerage power equally held across ages?*	**RQ4.1:** *Does the online domain erode the common ground for public debate?* **RQ4.2:** *Does digital media increase or decrease agenda diversity and capacity?* **H4.1:** *Online media diets do not erode public consensus over the issues that people believe are important.* **H4.2:** *Online news media diets increase the level of diversity of the public agenda, which, by definition, implies that there will be a higher carrying capacity regarding salient issues.*
Findings	The provision of news is highly fragmented with clear fault lines associated with media ownership structures. The audience network does not show fragmentation. The consumption of news is only slightly correlated to the network of news media sites.	Legacy media are regarded as the most authoritative sources. They also retain much of the control of the audience flow. Young audiences, who exhibit distinctive patterns for news consumption, confer brokerage power to native digital outlets.	Citizens agree on a set of common experiences when their online news diets are based on those outlets with higher shares in the news market, i.e. legacy media. Concentrated online news media diets are associated with more diverse agendas and a greater capacity to account for public problems.

Tab. 1.2.: Summary of the Data and Methods

	Chapter 2	Chapter 3	Chapter 4
Goal	Analysis of the macro structure of the online news domain.	Analysis of the meso structure of the online news domain.	Analysis of the impact of the online news domain on the public agenda at the micro o individual level.
Data Structure	**News Media Network (NMN)**: – Nodes: news media. – Ties: hyperlinks. – Network type: directed and weighted. **Audience Network (AN)**: – Nodes: news media. – Ties: audience overlapping. – Network type: directed and weighted.	**News Media Network (NMN)**: – Nodes: news media. – Ties: hyperlinks. – Network type: directed and weighted. **Audience Network (AN)**: – Nodes: news media. – Ties: audience overlapping. – Network type: directed and weighted. – 4 different audience networks by ages: 18-24; 25-34; 35-54; +55.	**Survey Data (SD)** **Browsing Tracking Behavior (BTB)**: – Data: Anonymized browsing history from an online sample of Spanish individuals.
Data Source	**NMN**: Supervised web crawling, *VOSON*. **AN**: Spanish official online audience meter, *comScore*.	**NMN**: Supervised web crawling, *VOSON*. **AN**: Spanish official online audience meter, *comScore*.	**SD**: Non-representative sample of the online Spanish population, Catalonia is not included (*Netquest's* opt-in panel) **BTB**: Sub-sample of individuals from the survey sample who explicitly agreed to be tracked long before the field work was launched.
Sample	**NMN**: Top 44 news media in Spain, audience rankings provided by *Alexa* and *comScore*. **AN**: Spanish Representative Online Panel ($N=30,000$).	**NMN**: Top 44 news media in Spain, audience rankings provided by *Alexa* and *comScore*. **AN**: Spanish Representative Online Panel ($N=30,000$).	**SD**: $N=725$ **BTB**: $N=372$
Period of Analysis	**NMN**: July, 2014 **AN**: September, 2014	**NMN**: February, 2015 **AN**: November, 2015	February to April 2015
Methods	**AN Tie Significance:** Standard correlation methods for binary variables are used to eliminate ties that result from random distribution. **Network Fragmentation Analysis:** Community detection analysis for network data i.e., Random Walker Algorithm. **Network Correlation:** – Quadratic Assignment Procedure or QAP Test. – Exponential Random Graph Model (ERGM) or Network Regression Model.	**News Media Authorities and Hubs:** – Hyperlink-induced topic search or HIT Algorithm applied to the NMN. – Significance analysis of the authorities and hubs scores by using Configuration Multi-Edge Model. **Audience Brokers:** – Betweenness centrality analysis for weighted networks. – Significance analysis of the betweenness scores by using Configuration Multi-Edge Model.	**Key Variables:** *Dependent Variables:* – Common Agenda: Top 3 Most Important Problems as mentioned by respondent. – Agenda Diversity: Shannon's *H* Measure. – Agenda Capacity: Number of items per respondent. *Independent Variable:* – Type of News Media Diet: Herfindahl-Hirschman Index (HHI). Statistical measure of concentration. It equals to the sum of the squared market share of each news outlet included in subjects' media diet. The same measure is applied to observed and reported data. **Analysis:** Ordinary Least Squares (OLS) Method.

Macro-level Analysis of the Online News Domain: Assessing Fragmentation in the Online News Network Structure and Audience Behavior[1]

2.1 Introduction

Among all the academic discussions about how digital technologies are affecting news consumption, one research strand has become highly influential. It focuses on the dangers of *personalized information* (Katz, 1996; Resnick, 1997; Sunstein, 2009; Turow, 1998), or on what Pariser (2011) named the effects of the "filter bubble". It highlights the potential threat that the digital domain poses on the society because it increasingly fragments digital audiences. As their argument goes, the unlimited fragmentation of online audiences has pernicious consequences because it erodes the necessary common ground for deliberative democracies. In the background of these discussions there is a model of democracy that stems from normative conceptions of the public domain (Berelson, 1952; Converse, 1964; Gutmann & Thompson, 2009; Habermas, 1994; Rawls, 2009). 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 much political action (Baum & Groeling, 2008; 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 or, in the words of Habermas (1994), on having a public sphere. Hence, according to these theories, holding people's attention is necessary for making communities part of a whole so that they can resolve shared problems and advance based on inclusiveness.

It is no wonder, then, that a lot of attention is being paid to how digital technologies are reshaping the digital news domain. With growing frequency, they have been equated with the end of the public sphere, yet it is our argument that the extent to which the web and related technologies preserve this ideal of a public space for democratic debate depends on the behavior of two actors: the *providers* and the *consumers* of information. Until now, scholars have studied the fragmentation of the web through the lens of one or the other (Adamic & Glance, 2005; Barabási, 2002; Benkler, Roberts, Faris, Solow-Niederman, & Etling, 2015; Hindman, Tsioutsiouliklis, & Johnson, 2003; Neuman, 2001; Tewksbury, 2003, 2005; Webster, 2005; Webster & Ksiazek, 2012). The relatively scarcer research assessing whether the structure of the web correlates with the information-seeking patterns of users has focused on global patterns of consumption (Barnett & Park, 2005; Taneja, 2016; Wu & Ackland, 2014) and has not addressed the structure of the digital news domain. This Chapter 2 aims to fill that gap.

We specifically analyze whether the structure created by news providers when they send links to each other correlates with the information-seeking patterns of users. Our approach allows us to determine whether the fragmentation in one dimension of the news domain, or the lack of thereof, is mirrored by the other dimension. We argue that it is essential to jointly study the supply and demand of information if we are to determine the extent to which the digital news domain is as highly fragmented as the scholars cited above have feared. Our goal is to offer novel evidence that helps us reassess these theoretical discussions on the effects of online information exposure. Our main contribution is that we offer an explicit analysis of the relationship that exists between the network of news providers and the navigational patterns of audience members.

We believe that this network approach, which has been applied before (Taneja & Wu, 2014; Webster & Ksiazek, 2012), offers a crucial improvement over studies analyzing media diets through surveys (Trilling & Schoenbach, 2013). Questionnaires rely on respondents' accurate recall of past behavior, and they limit the amount of information that can be gathered about online consumption. Frequently, respondents can only account for the sources that are fresh in their mind or, as Kahneman (2011) put it, more readily "available" (p.425), which means that respondents tend to forget media outlets that are less popular and visited less frequently; at the same time, respondents can also over-report news exposure, adding an additional source of noise to the measurement (Prior, 2009; Tewksbury, 2003). Here, we analyze audience behavior using browsing data, and we combine that information with a web crawl that reconstructs the hyperlinks amongst the most important news organizations in Spain.

Our data collection strategy thus relies on two sources: hyperlink networks as crawled from the web; and data on audience behavior, as provided by the Internet analytics company *comScore*. We used the top-read online news outlets as initial seeds in the web crawl, which resulted in a network of close to 9,000 nodes (or web pages). The audience data comes from a panel of 30,000 individuals that tracks browsing patterns for four months. By correlating the two networks, we evaluate the existence of fragmentation in news production (**RQ2.1**) and audience behaviour (**RQ2.2**) and determine the extent to which it shapes news consumption (**RQ2.3**).

The rest of this Chapter 2 proceeds as follows. First, we review prior work on how digital technologies are causing fragmentation in news production, paying special attention to the emergence of new actors and news organizations. Then we describe our data and methods, and we present and discuss the main findings. Our analyses show that the fragmentation in the network of news providers is only slightly associated to the information-seeking patterns of the audience. We conclude by offering a discussion on the future lines of work and an assessment of the effects that digital copyright policy regulation —especially stringent in Europe– might have in the configuration of an online public sphere.

2.2 News Consumption in the Digital Age

Digital technologies have brought fundamental changes to the way people consume political information. The web allows citizens to have greater control over news selection; it offers unlimited sources to keep up with political events; and it transforms agenda setting and gatekeeping processes, which according to some scholars, are no longer the monopoly of traditional media (Benkler et al., 2015; Farrell & Drezner, 2008)[2]. Early theoretical accounts of these changes were, for the most part, optimistic; but it did not take long for scholars to start paying attention to some of the pernicious consequences of digital technologies, identified with traditional conceptions of the public sphere in the backdrop (Gitlin, 2002; Sunstein, 2009; Turow, 1998; Chaffee & Metzger, 2001; Prior, 2008). These critical accounts pay special attention to the increasing fragmentation of the online domain and its negative effects on democracy.

Their argument relies mostly on the multiplication of sources and the specialization of new outlets: the assumption is that the fragmentation of news production provokes the fragmentation of the audiences and this, in turn, undermines a common space for the discussion of public affairs. However, the decisions of users on how to navigate

---

[2]Chapters 3 and 4 offers an in-depth discussion on the extent to what legacy news organizations might still hold the monopoly of the classic functions of media news organizations.

online content are, for the most part, disregarded: theoretical accounts on media fragmentation do not take into consideration the patterns of information-seeking revealed by users, especially since the web allows for greater control in exposure to information. Scholars who only assess the behavior of online audience, on the other hand, have also brought contending evidence that challenges theoretical accounts focused on fragmentation (Flaxman, Goel, & Rao, 2016; Gentzkow & Shapiro, 2011; Trilling & Schoenbach, 2013; James G. Webster & Ksiazek, 2012). These studies show that media diets are diverse and that they include the most prominent news media, i.e. mainstream news sources, which become the *de facto* common ground for news exposure. According to Flaxman et al., (2016), the vast majority of online news consumption is accounted for by individuals simply visiting the home page of typically mainstream media, which, as these authors conclude, tempers the consequences of recent technological changes brought by the advent of the web. To sum up, the debate on how digital technologies shape audience behavior and whether they cause a fragmented public domain is still open and unresolved.

2.2.1 Fragmentation in News Production and Consumption

In this study, fragmentation relates both to the supply and the demand sides of information. We jointly consider both dimensions to assess the level of fragmentation, or the lack of thereof, in the digital news domain. Prior work has studied these dimensions separately and mostly making assumptions about audience behavior based on the structure and distribution of online news content. Authors studying the digital domain 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; Pariser, 2011; Sunstein, 2009). For example, Tewksbury (2005) sustains that the specialization of online media outlets leads to fragmentation because they attract homogeneous groups of users. Similarly, Papacharissi (2002) warns that fragmentation of the online sphere derives from the proliferation of smaller and specialized groups. However, Lee (2007, 2009) looks at the reception of news information and refutes those claims by showing that a fairly similar agenda across news outlets undermines the actual fragmentation levels.

Parallel to these studies, there is a related area of work studying the online supply of news. This line of research considers whether audience fragmentation depends on the informational structure built by news providers through their web sites by citing content from one another using hyperlink (Ackland & Gibson, 2004; Williams, Trammell, Postelnicu, Landreville, & Martin, 2005). Theoretically, the provision of a hyperlink should be a journalistic activity: through hyperlinks, readers can trace sources of information, expand on the context that gives meaning to data,

elaborate on interpretations, and improve their knowledge about the reported news. But due to the crisis affecting the media industry (Pew Research Center, 2015), business criteria tend to dominate journalistic decisions. As a consequence, news sites rarely link to external competitors: hyperlinks are not treated as journalistic objects that add value to the stories, but as economic assets or functional devices that can, theoretically, keep audiences within corporate boundaries (Dimitrova, Connolly-Ahern, Williams, Kaid, & Reid, 2003; Karlsson, Clerwall, & Örnebring, 2014). According to De Maeyer (2012), the tendency to avoid linking to outside sources shapes the network of providers as "walled gardens" (Napoli, 2008:63). Drawing on this metaphor, advocates of the fragmented public sphere assume that consumers of information will not venture beyond the walls established by hyperlinks and, instead, will remain within the confines defined by news organizations, which, ultimately, fosters audience fragmentation. But again, this argument is not supported by evidence explicitly measuring the extent to which the audience responds to the connections, i.e. hyperlinks created by media organizations.

An additional research line of the demand side of online information runs, in fact, counter to many of these arguments. This area of study borrows tools from network science to assess the structure of the supply dimension of digital news domain. Their results suggest that the web does not act as a driving force towards fragmentation. Webster and Ksiazek (2012), for instance, find no evidence of the predominance of web fragmentation; moreover, they challenge the arguments highlighted above by stating that "the sheer number of digital outlets in competition does not determine the extent of fragmentation" (2012:44). The authors focusing on global patterns of information consumption —who study any type of information consumption, not only news content— do not find evidence of fragmentation in the digital domain either (Barnett & Park, 2005; Taneja & Webster, 2016; Wu & Ackland, 2014). In fact, what is more relevant for our study is that they find that patterns of audience formation are less determined by the structure of news providers than they are by cultural and linguistic factors. In other words, according to their evidence, geographic similarities are more powerful predictors of the audience behavior than hyperlinks. On the basis of these findings one might expect that the pernicious consequences of the unlimited sources of information that some authors, quoted above, have feared should be, at the very least, moderate.

Other recent studies applying survey methods support this statement: they show that a significant share of consumers gets news from multiple outlets (Garrett & Resnick, 2011; Gentzkow & Shapiro, 2011; Trilling & Schoenbach, 2013). Overall, following these latest findings, we should expect that even if the web is fragmented, the provision of hyperlinks is not the main driver of online news consumption, which is contrary to the assumption made in much prior work, and hence we should not find fragmentation in the consumption side of the news domain.

2.2.2 The Emergence of New Players

These contradictory scenarios demand attention to the role that new media actors play in shaping the online public sphere. These new players are increasingly important on the web (and related mobile technologies), especially as a source of news information and as mediators of the audience engagement with news providers. According to our definition, new media are news organizations that combine a deep understanding of journalism with a high technological know-how to curate news from several sources and tailor them to match individual preferences; new media also typically engage audiences to collaborate in the production of information[3]. More importantly, they are content discoverers and, through hyperlinks, they direct attention to a diversity of ideas and other news providers (Athey, Mobius, & Pal, 2012; Turow & Tsui, 2009). They do not only mediate the distribution of news —as social platforms do– by aggregating links from several sources into a single page, but some of these actors also produce a small portion of their content. They are native digital outlets, blogs or news aggregators like *BuzzFeed*, *Feedly*, *Snapchat*[4], the *Huffington Post* or *Politico* in the U.S., and *Eldiario.es*, *Elespanol.es* and *Meneame.net* in Spain.

Thus, unsurprisingly, while some major newspapers barely link to external sources at all (Tsui, 2008), most of the new media outlets link heavily to other web sites. Precisely for this reason, Garrett and Resnick (2011) defend the role that they can play in diversifying news diets and disregard the presumption that personalization makes fragmentation of the news market inevitable. Their argument is backed by the latest findings on the role of one of the major news aggregators in the worldwide news market, *Google News*. By means of a natural experiment, Athey et al. (2012) show that this type of new media fosters news reading and increases the number of news outlets that people visit when keeping up with current events. They used the case of *Google News* in Spain as a natural experiment. This service shut down in 2014, when the government introduced a non-waivable copyright fee to be paid by online new media for linking to content created by legacy news organizations (Posada de la Concha, Gutiérrez, & Hernández, 2015; Rushe, 2014; Xalabarder, 2014). According to the aforementioned study, after the *Google News* aggregator service went away, there was a very substantial reduction in the number of news outlets that Spanish users visited. The magnitude of the effects that the study unveils is greater for small news outlets. Further, and according to their work, news aggregators reinforce people's knowledge about the most important news events and help people to go deeper into a news topic, which we can interpret as a mechanism to anchor people to the necessary common ground for the functioning of democracies.

---

[3]See an in-depth discussion about the characteristics of the new media outlets on Chapter 3.

[4]*Snapchat* includes fifteen major media brands in its *Discovery* tab including *ESPN*, *CNN* and *Vice* (Tepper, 2015).

In spite of the evidence reviewed above that signals positive consequences of the new media, however, new media actors are more often than not identified as the main promoters of fragmentation on the web (Chaffee & Metzger, 2001; Sunstein, 2009). But whether this fragmentation shapes audience behavior is, again, a matter for empirical analysis. In this regard this Chapter provides evidence on two fronts, illustrated in Figure 2.1.

Fig. 2.1.: Schematic Representation of the Fragmentation Analysis
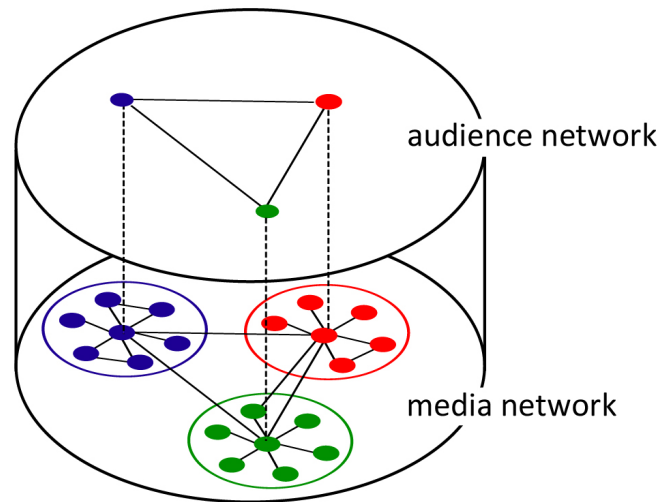
Note: The media network represents hyperlinks connecting news media on the web and the audience network represents people behaviour when navigating web content.

First, we determine the extent to which the news or hyperlink network is organized around clustered communities, as depicted by the colored circles in Figure 2.1, which we would interpret as evidence of fragmentation; and second, we want to assess whether audiences respond to that organization by navigating online content in a similarly clustered fashion. This leads to our main hypothesis in this Chapter:

Hypothesis 2.1 (H2.1): *The structure of the news (hyperlink) and audience networks will be significantly correlated if the audience responds to the connections created by media organizations.*

To test this hypothesis, we first examine the structure of both networks and determine whether there is evidence of fragmentation. Then we assess whether the existence of links in the audience network is positively correlated with the existence of links in the news network or hyperlink structure. Our study departs from previous research, in that we define fragmentation as related to both sides of the digital news domain: the supply and demand side of information. We do not operationalize fragmentation in terms of the aggregate number of media options available and how audiences distribute across them; instead, we take a structural approach that looks at media networks and at how audiences navigate those structures. More formally, we measure

fragmentation as the existence of network clusters within which information and audiences flow more frequently. If these clusters exist, either because media outlets link to each other strategically or because audiences navigate only certain media circuits, then we claim we have evidence of fragmentation.

2.3 Data and Methods

2.3.1 News Media Data

The network of news media was collected in July 2014. We used a web crawl that followed hyperlinks starting from a list of 44 seed sites and resulted in a network of close to 9,000 nodes. The list of initial seeds corresponds to the top-read online news outlets in Spain, and were selected using the publicly available rankings produced by *Alexa* (Alexa Internet, 2014). *Alexa* limits the publicly available information to the top 500 sites by country; from these, we selected the sites published by Spanish news outlets, which lead to the 44 sites we used as seeds. Several prior studies have relied on these rankings to obtain traffic information (e.g. Ennew, Lockett, Blackman, & Holland, 2005; Wu & Ackland, 2014). However, to further test the accuracy of the *Alexa* data, we compared the ordering with the list provided by *comScore*, one of the most referenced audience trackers in Spain. The correlation of the two rankings was 0.906.

Once the initial sample seeds were identified, we crawled their links using commercial software that allows mapping web networks (Ackland, 2010). We designed a supervised web crawling process to build the hyperlink network, and we used the top-read online news outlets to do so, following previous research that shows that selecting seeds sites according to well established prominence (in our case, size of online audience) minimizes the bias of rendering peripheral sites as central (Weber & Monge, 2011). In addition, the crawler we used combines data collected through snowballing from the seed sites (e.g. links going out) with search engine information of incoming connections to those sites (e.g. links coming in). This, again, minimizes the bias introduced by using small set of seeds in the process of data collection.

The network we obtained is summarized in Table 2.1. In total, we collected hyperlinks connecting close to 9,000 unique domains. The network is very sparse, with a few outlying hubs in the degree distribution (i.e. the allocation of content links), and a clear core-periphery structure, as indicated by the negative degree correlation coefficient, which suggests that peripheral sites tend to connect to very central sites. Since the web is a constantly evolving information environment, we conducted another crawl with the same seed sites in February 2015, applying the same method

but using more exhaustive search parameters (which determine how many incoming or outgoing hyperlinks the crawl should collect by site). The descriptive statistics for this second network[5] are also summarized in Table 2.1. The network is twice as large as the network collected in July 2014, which means that the crawling collected a larger number of peripheral sites; the range of the degree distribution is also broader, but overall the network exhibits the same structural properties.

Tab. 2.1.: Descriptive Statistics for the Hyperlink Networks Crawled from the Web

	July 2014	February 2015
Size	8,708	16,659
Number of edges	14,941	32,628
Density	0.0002	0.0001
Reciprocity	0.05	0.05
Clustering	0.007	0.006
Mean incoming links	1.7	2
Max incoming links	522	1,223
Mean outgoing links	1.7	2
Max outgoing links	183	704
Degree correlation	-0.687	-0.615

Note: Density is calculated by counting the number of realized links relative to potential links. Reciprocity measures the proportion of mutual links, and clustering summarizes the number of closed triangles in the network, which serves as a measure of local density in link formation.

Figure 2.2 further supports the claim that the two networks are very similar. More precisely, Panel 2.2-1 shows that centrality (measured as the number of incoming links) is highly correlated in the two web crawls. Panel 2.2-2 zooms into the most central websites in the network (identified in the grey area of Panel 2.2-1). Unsurprisingly, these sites are published by the main news providers in Spain, with the two main daily newspapers at the head of the distribution, which speaks for the representativeness of our network. In all, the scatterplots suggest that the basic composition of the network, and especially its core, does not change substantially over time. In other words, we can assume a high level of stability of the news network. From here on, the analyses focus on the data collected in July 2014.

2.3.2 Audience Data

The audience network is also formed by media sites, but the ties measure something different. Instead of tracking linking activity on the web, the network maps the total overlap in the audience of web sites, following prior work (Ksiazek, 2011; Webster & Ksiazek, 2012). In particular, a link from media site i to media site j indicates the total audience visiting site i that also goes to site j. If two sites do not share

---

[5]See Chapter 3 for more information about the micro-structure of this news network.

Fig. 2.2.: Correlation of the Online Centrality Measures over Time (Hyperlink Network)
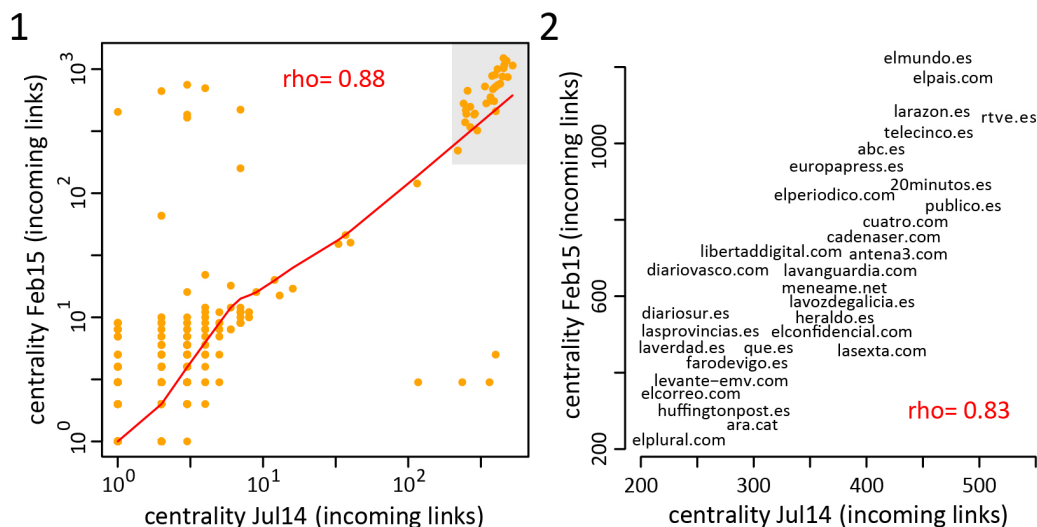
any audience, they are disconnected. The network is directed because the ties are expressed as portions and the base total changes from site to site. In other words, dyads in this network can be considered as frequency tables with different row totals: the percentages used to assess the strength of the audience overlap depend on who initiates the tie. The audience network is built with data gathered in September 2014, one month after the web network was collected, a necessary time lag given the causality implied in our hypothesis (i.e. that media connections shape audience behavior). Although our data do not allow us to strictly disentangle causality, we need a temporal lag to test the claim that fragmentation in the hyperlink network, or that a fragmented media structure, leads to audience fragmentation. Given that audience data is aggregated on a monthly basis, we decided to use audience data for a month posterior to the collection of our web data. The audience network captures the aggregated behavior of users seeking news, and it is the best available proxy to the dynamics of online news consumption.

We drew these data from a representative panel of 30,000 individuals tracking their browsing behavior, collected and provided by *comScore*. These data are unique in several ways. First, *comScore* is, since 2011, the official source for audience information and online news analytics in Spain (Interactive Advertising Bureau, 2011). Second, the data are based on observed behavior instead of self-reported exposure, which offers a more accurate representation of the media diet of users (the company uses biometric identification systems to disregard data from non-panel members). Third, the online data is combined with other indicators drawn with audience metrics collected by a separate agency (AIMC, 2014) to check the accuracy of their measurements[6]. And finally, *comScore* data is a highly renowned and trusted

---

[6]Panel recruitment is done entirely online and non-cash incentives are provided. Upon joining the panel, users download software that routes their Internet traffic via *comScore*'s proxy servers

source of audience data for news organizations. To the best of our knowledge, this is the first time that *comScore* has provided these data for academic purposes in Spain.

In order to calculate basic statistics, the data provider requires a minimum of 16 panelists visiting a site within a given month. Additionally, to build the metric of audience overlap between two sites —the main statistic we use to build the audience network– our data providers require there to be a minimum of thirty minutes between two visits to the same site, as well as that users remain on one site at least three seconds to count that as a single visit. Because of this, and our focus on news providers (which are in general less visited than, say, entertainment sites), only a small fraction of the domains collected during the web crawl are part of the provider dataset —which are, nevertheless, the most relevant in terms of actual audience reach. In total, we collected information on audience reach and audience duplication for 114 sites. The network is summarized in Table 2.2. Two features characterize this network: first, it is very dense, which means that most sites share an audience with most other sites; and second, it exhibits high levels of reciprocity, which means that audience flows in all directions.

Not all the ties in this network signal a significant amount of overlap; as prior research has suggested, some level of audience duplication is likely to occur just by chance (Ronen et al., 2014). We determine the strength and significance of the ties using standard correlation methods for binary variables, an empirical approach used previously in the analysis of similar networks (Ronen et al., 2014). In particular, we use the phi correlation coefficient, which measures the departure of observed overlapping audience from the expected value if the audience moved randomly; and we use the t-statistic to measure the significance of that departure.

The phi coefficient is positive if audience members of site i visit site j more often than expected by random chance; we use $t > 2.6$ as the critical value to count that difference as significant. Following this criterion, we eliminated all links that fell short of the statistical test, as well as those based on a total audience overlap $A_{ij} < 10$ (i.e. small numbers make significance statistics unreliable). As Table 2.2 shows, the network that results from this filtering exercise is smaller (once non-significant links are deleted, some nodes become isolates, which we removed from the network) and less dense; reciprocity, however, is still high, and there is a clearer core-periphery structure, as indicated by the negative degree correlation.

---

allowing passive measurement of online activity. Finally, the data are weighted to capture the size and shape of the overall online population, which ensures representativeness. More information about the socio-demographic characteristics of the *comScore* panel can be found here https://goo.gl/FtFG4e.

Tab. 2.2.: Descriptive Statistics for the Audience Network (September 2014)

	Full Network	Only Significant Links
Size	114	111
Number of edges	12,493	6,471
Density	0.97	0.53
Reciprocity	0.98	0.96
Clustering	0.99	0.74
Mean incoming links	110	58
Max. incoming links	112	100
Mean outgoing links	110	58
Max. outgoing links	113	106
Degree correlation	-0.07	-0.33

We also assess the level of fluctuation in our audience data, as we did with the web data, by comparing data for the months of September to December of 2014. Figure 2.3 shows a scatterplot matrix with the correlation of those rankings[7]. Websites are ranked according to their reach or audience size during this four-month period. The correlation coefficients are very strong, which means that the relative prominence of websites in terms of reach barely changes regardless of the timing of measurement.

Fig. 2.3.: Correlation of Rank Position on Audience Reach over Time (Audience Network)

---
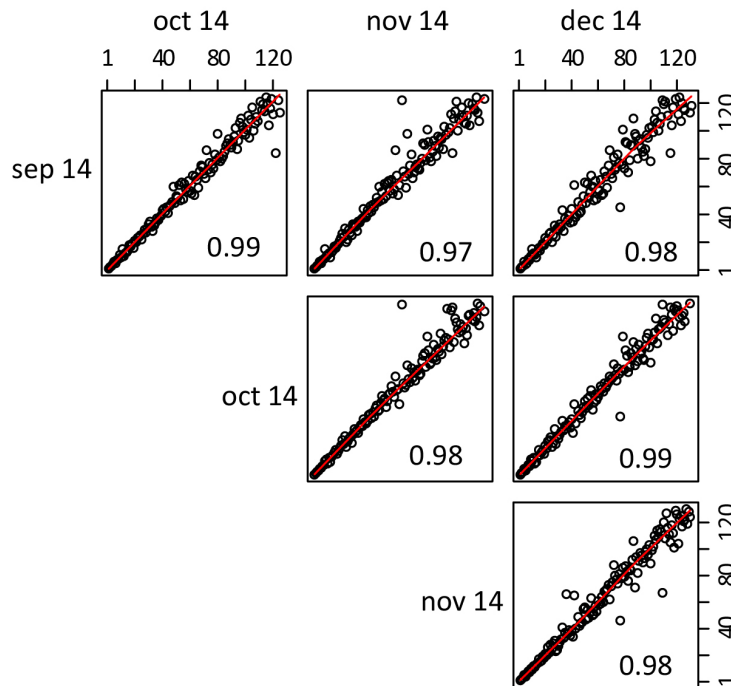
[7]An expanded analysis of the stability of the audience reach is provided in Figure 3.1. The results lend further support to our assumption about the stability of the audience network.

2.3.3 Mapping Fragmentation and its Effects

As the literature reviewed above suggests, the irruption of digital technologies has generated much discussion about the rise of media fragmentation —and its impact on information exposure. In order to operationalize the concept of media fragmentation, we use a community detection method based on simulating random walks in networks (Rovall & Bergstrom, 2008). *Community detection* is a technique for the reduction of networks that classifies nodes into modules according to the density of connections: nodes in the same module have more connections to each other than to nodes in other modules (Girvan & Newman, 2002; Newman, 2012). This approach helps characterize the organizational logic of a network by delineating areas where the network is denser and, therefore, more likely to channel information. The method we use makes this idea explicit by allowing a *random walker* to follow the links (i.e. this walker is an algorithmic representation of a user that would wander the network randomly), and then drawing communities on the basis of how long the walker stays in each area of the network. This approach offers a statistically robust method to determine if a network can be split into groups, which, in our context, signals the fragmentation of the news media landscape, or what some have called "balkanization" of the online domain (Sunstein, 2009; Van Alstyne & Brynjolfsson, 1996). The advantage of the method we employ here is that it was explicitly designed to model information flows in directed and weighted networks, matching the nature of our data and our research question more closely than other available community detection methods. The traditional way of identifying community structure simply disregards the direction and strength of the links. The approach we use here does not discard this information, which has been shown to offer crucial insights into the structure of networks, especially when that structure channels information flows (Rovall & Bergstrom, 2008:1121)[8].

Ultimately, our analyses aim to provide evidence on two fronts: first, the extent to which the news media network is organized around communities, which we would interpret as fragmentation and community detection, helps us to test this claim; and second, the extent to which the audience responds to that organization. Our hypothesis is that, if the audience responds to the connections created by media organizations, the structure of the two networks will be significantly correlated; in other words, the flow of audience from one site to another (i.e. the existence of links in the audience network) will be positively affected by the creation of hyperlinks on the web connecting those sites. We identify and analyze these correlation patterns using the QAP procedure (Krackhardt, 1987, 1988) and exponential random graphs models, or ERGMs, which are regression models for network data (Koehly, Goodreau,

---

[8]For more information on community detection methods see Borge-Holthoefer & González-Bailón, 2015; Ziv, Middendorf, & Wiggins, 2005.

& Morris, 2004; Lusher, Koskinen, & Robins, 2012; Snijders, 2011; Snijders, de Bunt, & Steglich, 2010).

2.4 Results

2.4.1 Fragmentation in News Production

The community analysis of the web network yielded a partition with 66 groups and a modularity score of $Q = 0.68$. This score measures the fraction of links that fall within a group minus the expected fraction if links were distributed randomly (as explained above, groups are identified using a random walker). The score is positive if the number of links within groups exceeds the random expectation and, as it gets closer to 1, the modularity structure of the network (i.e. its fragmentation) is considered to be stronger (Newman & Girvan, 2004). What our results mean is that the approximately 9,000 sites in the network summarized in Table 2.1 can be grouped in 66 modules that have significantly more internal links than external. About half of these modules are, in fact, formed by single sites: because of their lack of connections and peripheral nature, they are difficult to reach from the other sites in the network —hence their classification as single-node communities. The largest communities are represented in Figure 2.4, which describes how different they are in terms of centrality and size. The most populated community is number 1, formed by more than 800 sites (Panel 2.4-2(a)). The hub in this community (i.e. the site that receives more links from other sites) is the online newspaper *Que.es*; however, this is not the most central community in terms of hyperlinks arriving from other communities (in-degree centrality): this position is occupied by community 2, whose hub is the website of a commercial television channel, *Telecinco.es*; this site is followed in the ranking by *Cuatro.com*, a website of a television channel that also pertains to the same media business group *Mediaset*. The size of the nodes in Panel 2.4-1 is proportional to the centrality of the communities they represent; Panels 2.4-2(b) and 2.4-2(c) give a summary of the degree distribution within the communities. What these descriptive statistics suggest is that most communities are organized around hubs or web sites that are disproportionately more central within their respective groups. These hubs are the websites identified in Figure 2.1, Panel 2.1-2, which are also the sites that have larger audiences, according to *Alexa* and *comScore* data.

Nodes are colored depending on whether the hubs are legacy media types (i.e. commercial television channels, radio, traditional newspapers that precede the digital revolution) or new media types (i.e. news aggregators, born-of-the-web news providers). Ties are colored depending on whether they link communities led by old

Fig. 2.4.: Communities in the News Media Network
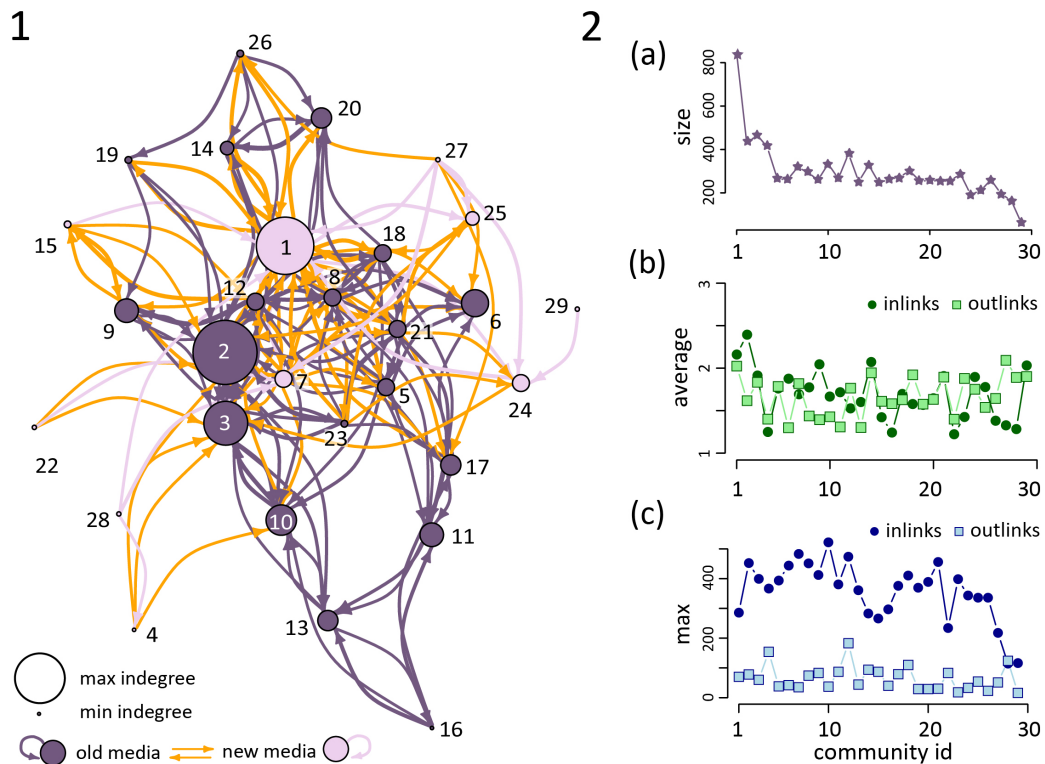
Note: Edges with low weight ($w < 100$) have been filtered to improve visualization. The size of nodes is proportional to the centrality of communities in the network; the color of nodes indicates if the hub in each community is a new or an old media type. The color of edges identifies links between those two categories (orange for links that connect new and old media).

and new media (in orange), or remain internal to each of the categories (i.e. dark purple if they connect old media; light purple if they connect new media). Figure 2.4 shows a visible prevalence of orange ties and suggests that hyperlinks between legacy media and new media are prevalent. In general, though, communities organized around new media sites are peripheral compared to those formed by more traditional news organizations. For instance, the hubs in communities 27, 28, and 4 are the two native digital news sites *Elplural.com* and *Eldiario.es* (which are left-leaning outlets) and the social news aggregator *Meneame.net*. Peripheral communities formed by traditional media (i.e. 13, 16, 11 and 14, 19, 26, 29) have regional news providers as hubs (i.e. *Tv3.cat*, *Ara.cat*, and *Lavanguardia.com*, which share the same language, Catalan, and *Farodevigo.es*, *Heraldo.es*, *Lavozdegalicia.es*, and *Canarias7.es*). The two most important legacy media, *Elpais.com* and *Elmundo.es*, are the hubs of communities 12 and 8, respectively.

On the basis of these analyses, we can claim there is evidence of fragmentation in how news providers create their content. The divides seem to respond mainly to the structure of media ownership and conglomerates, a finding that lends support

to previous research (Dimitrova et al., 2003; Karlsson et al., 2014). For instance, *Lasexta.es*, *Antena3.com* and *Ondacero.es*, placed on community 3, share the same owner, *Atresmedia*; in community 12, we find *Cincodias.es*, a business media outlet that belongs to *Grupo Prisa*, the same group that owns the hub of the community, *Elpais.com*. The overall allocation of online visibility is still largely dominated by big traditional media organizations, that is, richer organizations publish more central sites (González-Bailón, 2009). Overall, digital news organizations have peripheral positions in the web network. But there are two important exceptions: the hubs of communities 1 and 7. These are the digital version of newspapers that, due the economic crisis of 2008, had to discontinue their printed editions in 2012 (Fernández-Santos, 2012). The first was a popular free newspaper founded by one of the main media groups in Spain, *Grupo Recoletos* in 2005; and the second, *Publico.es*, is a well-known left leaning outlet promoted by two of the more influential media businessmen in Spain (Casanueva, 2007). Although less evident, the left-right ideological dimension might also play a role in the actual structure of the web network: community 4, for instance, hosts several left leaning news sites e.g. *Infolibre.com* or *Lalamentable.org*[9].

2.4.2 Fragmentation in News Consumption

The network formed by audience overlap, summarized in Table 2.2, is so dense that all nodes belong to the same community (i.e. the modularity score is close to 0). This means that Internet users, on the aggregate, navigate widely when searching for news. However, we can use the tie strength (i.e. the extent of audience overlap) to identify parts of the network where connections are stronger and where the most significant amounts of audience share take place. We use the t-values as a measure of tie strength: they offer a standardized metric to assess how significant the audience overlap between two sites is compared to the overlap we could expect by chance; the larger the t-values are, the more significant the overlap. Figure 2.5, Panel 2.5-1 shows what happens as the weakest ties are removed from the audience network. As mentioned above, all the ties in the network are statistically significant, but here we progressively remove the weakest amongst those (t-values have been rescaled to a 0-1 interval). The number of components in the network starts growing soon after the first ties are filtered out, up to the point when all nodes become isolates. Panel 2.5-2 zooms into one of the networks that result from this process (isolates are removed from the visualization). The specific threshold used for this illustration is arbitrary in the sense that there is no theoretical guidance that dictates the right value of tie strength. In fact, there cannot be such guidance other than pointing out that there is

---

[9]Community 27 and 28 seem to follow the same ideological pattern, too. Both communities are organized around a left-leaning hub, *Elplural.es* and *Eldiario.es* respectively, and include several blogs that share the same ideological slant.

a trade-off between how stringent we are in the definition of tie thresholds and how meaningful the resulting network is for the purposes of illuminating audience flow patterns. Here, we considered the full possibility space, from all ties to none, and picked one intermediate example for illustrative purposes.

Fig. 2.5.: Panel 2.5-1 and 2.5-2 show the Audience Network at different Tie Thresholds
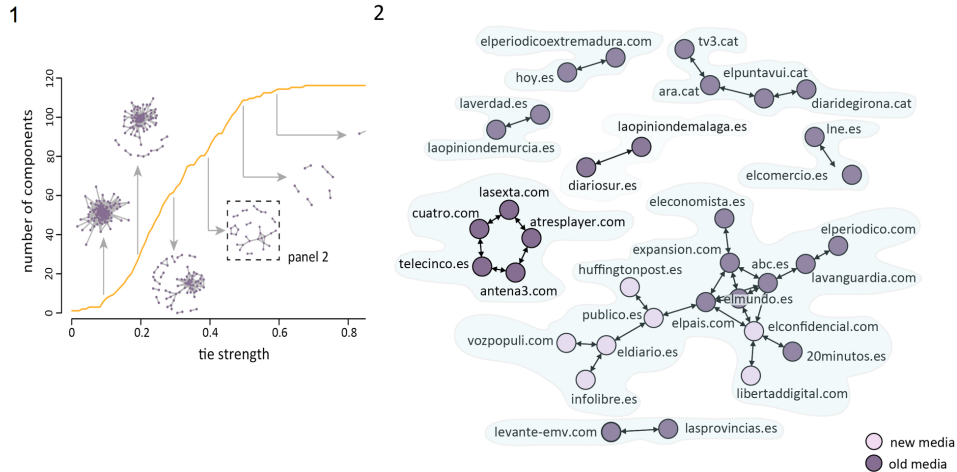
Note: Isolated nodes are removed from the visualization (t-values have been rescaled to a 0-1 interval).

There are eight components in the filtered network, each of them formed by sites that are more strongly connected to each other than to other sites (i.e. they share a relatively greater audience). The smaller of these components are formed by the web sites of regional news media, which, intuitively, share more readers than other news providers. Commercial televisions (e.g. *Lasexta.com*, *Antena3.com*, *Telecinco.es*) form an isolated cluster. The largest component, at the center of the visualization, is formed by a combination of new and old media: the most prominent newspapers (e.g. *Elpais.com*, *Elmundo.es*, *Abc.es*) are at the center of the component; native digital newspapers form a clear cluster of audience overlap, suggesting that they share an audience with similar demographics, probably younger, urban, and educated, and perhaps also with similar political attitudes (but we can only speculate about this; *comScore* did not give us access to demographic information at this level of analysis)[10].

According to the modularity score, the audience network does not exhibit fragmentation patterns as clearly as the web network, but when we refine our analysis and focus on the actual strength of the audience overlap, it still reveals certain dynamics of selective exposure. First, and intuitively, news media that cover information of regional interest share a larger fraction of audience members. Second, and more

---

[10]The cluster of native digital newspapers includes *Infolibre.es, Vozpopuli.com, Eldiario.es, Publico.es* or *Huffingtonpost.es*. All of them are left leaning news media, with the exception of *Vozpopuli.com*, which is owned by the former president of right wing newspaper *Elmundo.es*, Alfonso de Salas.

interestingly, new media outlets seem to be creating an audience niche separate from that of traditional media. However, the main conclusion we can draw from our analysis of the audience network is that there is a significant overlap of readers amongst most news sources. This suggests that the sector of the population that seeks political news performs, on average, broad searches. These patterns stand in opposition to theoretical accounts about fragmentation of online audiences (Chaffee & Metzger, 2001; Hindman et al., 2003; Sunstein, 2009) and fall in line with evidence on audience media diets which shows, both in the U.S. and in Europe, that people consume a sample of diverse ranges of outlets instead of a steady media diet (Garrett & Resnick, 2011; Gentzkow & Shapiro, 2011; Trilling & Schoenbach, 2013; Webster & Ksiazek, 2012).

2.4.3   Association between the Web and the Audience Networks

The previous two sections assessed the fragmentation on the supply and demand sides of news consumption. This section test the main hypothesis of the Chapter 2, namely whether there is evidence that linking practices are the main driver of audience behavior. We use network correlation techniques to evaluate the extent to which audience behavior responds to the fragmented structure of news supply. For this, we only retain nodes that are included in both the audience and the media networks. Then, we calculate the correlation between the adjacency matrices of the two networks, and we assess the significance of that correlation using the non-parametric QAP test (Krackhardt, 1987). Figure 2.6, Panel 2.6-1 shows that, although small, the positive coefficient is significantly higher than random; Panel 2.6-2 shows how that coefficient changes when different versions of the audience network are used: the figure shows that the correlation remains positive and significant until a certain tie threshold level is reached (i.e. when the audience network starts breaking up in too many small components). In other words, the analysis suggests that stronger connections increase the correlation among networks. These two panels show that hyperlinks are significantly associated to audience overlap, although to a very small extent. The nature of our data, though, does not allow us to disentangle the actual causal mechanism or the pathway followed by ties in the news and audience network.

Panels 2.6-3 and 2.6-4 display the estimated coefficients and goodness of fit diagnostics for the best fitting ERGM explaining the structure of the audience network. Table 2.3 gives more information to evaluate our model specification process. Because this is a very dense network, many statistics (in particular, those related to clustering and transitivity) created collinearity problems. The four structural parameters in the best-fitting model capture basic network configurations that manage to reproduce

Fig. 2.6.: Correlation between the Hyperlink and Audience Networks
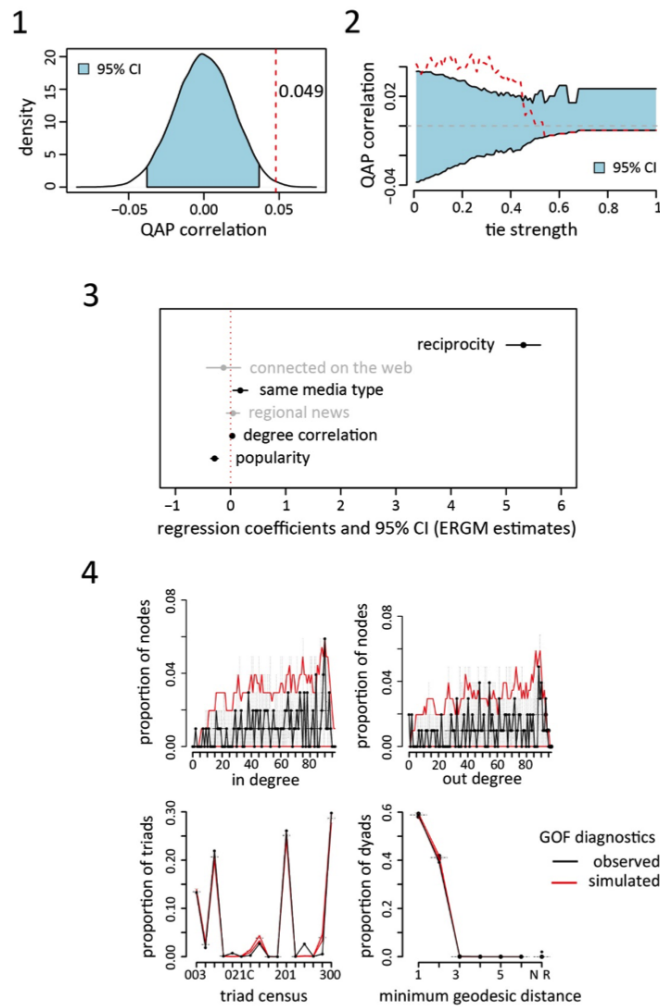
Note: Panels 2.6-1 and 2.6-2 show the results of the QAP test; Panel 2.6-3 shows the coefficients of the best-fitting ERGM explaining the structure of the audience network; Panel 2.6-4 shows GOF diagnostics.

the observed degree distribution, triad census and geodesic distance between pairs of sites: as Panel 2.6-4 shows, the observed values (black lines) are well reproduced by networks simulated using the estimated coefficients (the red lines serve as confidence intervals); we use this agreement as an indication of the goodness of fit of the model. Controlling for these network effects, a tie between two web sites is not significantly associated with audience overlap. In other words, the association identified with the QAP test vanishes once we control for the autocorrelation typical of network data. The model confirms, however, that audience overlap is more likely if two sites are published by similar types of news organizations: readers visit two news sites more frequently if both are either new media or old media. This confirms the significance of the clustering of new media sites identified during the threshold analysis of the audience network (Figure 2.5).

Tab. 2.3.: Network Regression Models Explaining the Formation of Audience Ties

	M1		M2		M3		M4		M5		M6		M7	
	est	se	est	se	est	se	est	se	est	se	est	se	est	se
Structural Effects														
edges	**0.36**	0.02	**-2.92**	0.07	**-2.95**	0.08	**-3.01**	0.07	**-3.01**	0.08			**-6.71**	0.25
reciprocity			**6.23**	0.14	**6.24**	0.14	**6.21**	0.14	**6.23**	0.15	**5.71**	0.21	**5.32**	0.16
transitivity											**-2.05**	0.25		0.16
popularity											**-0.56**	0.05	**-0.29**	0.03
degree correlation											**0.03**	0.00	**0.03**	0.00
Node Attribute Effects														
connected on the web					**0.47**	0.12	**0.43**	0.13	**0.48**	0.11	-0.03	0.29	-0.13	0.16
same media type							**0.14**	0.04	0.00	0.03	-0.02	0.05	**0.18**	0.07
regional news									**0.10**	0.04	-0.03	0.08	0.05	0.06
AIC	14284		8923		8885		8873		8878		5282		4889	
BIC	14291		8938		8907		8902		8915		5333		4940	

Note: Estimates can be interpreted as conditional log-odds ratios. Coefficients in bold are significant at the 95% level. Structural effects aim to control for network autocorrelation; we could think of these parameters as the building blocks that help assemble the observed network. Node attribute effects capture the influence that web links and type of media (e.g. new or legacy, regional or national) have on audience flow ties.

In short, the analyses show that the audience does not just follow the structure of links to seek political information; rather, our findings suggest the image of a more active citizen that navigates across a variety of sources and is motivated by factors that transcend the structure of information supply (Garrett, 2009). This more active citizen would consume information at her own pace and this, as Picard (2008) put it, is a radical change from the relatively passive way with which people have been consuming forms of mass media predating the Internet era.

2.5 Discussion

According to our results the fragmentation in the news domain is not mirrored by the audience behavior to the extent that some have feared. After jointly studying the supply and demand of information, this Chapter 2 offers novel evidence that helps us to reassess theoretical discussions on the effects of online information exposure. More precisely, in light of our findings, the high levels of audience overlap among different types of news providers —legacy and new media-– suggests that there is still a solid common ground to anchor informed citizenship and to provide public space for discussion of political events. Of course, that common ground is still likely to be organized around polarized fronts, two argumentative sides opposing each other rather than encouraging a deliberative, rational assessment of ideas. The actual content of news and their effect on political beliefs and behavior is a question that lies beyond the scopes of research on fragmentation; it is, nonetheless, the next logical step in a better understanding of how news exposure (as mediated by online technologies) affects the quality of democratic life.

The weaker association between the news and audience network suggests that the audience's seeking patterns respond to individual motivations that cannot only be directed with linking practices. Guiding the attention of audiences is, in light of our results, a more sophisticated endeavor than simply sending a hyperlink between news outlets. Links provide a path to browse the web and add relevance to content that otherwise would be less visible; and new media, especially aggregators, bring content together that would otherwise be scattered across the Internet and would require a more active search to find. But neither of them offers the full picture to explain navigation patterns. Our results show that the audience widely explores the available online news content. We offer evidence that audience overlap is stronger if two sites are new media or legacy media outlets, and that connections among left-leaning news sites are also denser. These findings suggest that sociological characteristics and political attitudes should also be taken into consideration in future studies to offer a more complete explanation of differential browsing patterns.

Moreover, according to our results, far from following journalistic criteria, hyperlinks are based on business strategy that theoretically aims to retain audiences for as long as possible within online boundaries that are drawn by media ownership structures. Traditional media barely link to external sites (Tsui, 2008) and, as our results show, when they do so, they refer to sites that are owned by the same parent corporations; these linking practices also help enhance the network position of the news outlets, which brings more traffic in the long run through search engines (Karlsson et al., 2014).

The results of the macro-analysis of the online news domain presented in this Chapter also has implications for international policy attempts to regulate the linking activity of the news sector. The most prominent example along these lines is Spain where the Parliament, as explained before, passed a law to impose a tax on linking activity in 2014 (Posada de la Concha et al., 2015; Rushe, 2014; Xalabarder, 2014). In Belgium and France, there were previous attempts to impose similar taxes in 2013, although at that time publishers and the main news aggregator in Europe, *Google News* reached an agreement to avoid the "link tax" law (Google, 2013)[11]. Recently, however, in 2016, the European Commission has presented a proposal for updating the copyright law in the digital age. If finally passed, the initiative will affect the 28 members of the Union. This proposal does not only mirrors the new Spanish legislation, but includes more stringent regulations[12] (European Commission, 2016, 2017; Google, 2016; Xalabarder, 2016) to regulate hyperlinks directed to legacy media organizations. The argument of traditional media that are lobbying to get this type of regulation approved is that aggregators add no value to the media business sector and capture advertisement market share without producing original content (Dellarocas, Katona, & Rand, 2013). Yet, in its current form, both the Spanish law and the European proposal could affect any type of new media that send links to legacy news organizations (European Commission, 2017; Froman, 2016; Xalabarder, 2014, 2016). As previous research has suggested, new media have the potential to promote cross-cutting exposure to thematically and ideologically diverse content by aggregating news content from several sources and increasing news consumption. In light of our results, we might therefore expect that this policy course that raises the cost of linking to otherwise disperse content, by imposing taxes on hyperlinks, should have pernicious consequences[13]. On the one hand, it might reinforce the

---

[11]The settlement included a fund for supporting publisher's digital initiatives in those countries (Google, 2013).

[12]The European Commission's proposal in its current form slightly differs from the Spanish version in that it gives exclusivity of the online copyright holders. This regulation could result in giving exclusive licensing to very few or even worse, to just one single platform (Xalabarder, 2016).

[13]Needless to say that there are alternative models to support news creators such as ad driven business. Some examples, along these lines, are *Facebook Instant Articles* (Goel & Ravi, 2015) and the *Local Media Consortium* which includes and alliance with *Yahoo* since 2006, to share ad revenues for news content aggregation from hundreds of U.S. local dailies (Helft & Lohr, 2006; Local Media Consortium, 2016) However, this is not the case of the *Google News*. The leading search engine is against ad driven models for news aggregation and supports news organizations through the

already existing fragmentation within the network of news providers, making it more difficult to explore news content for those who do follow hyperlinks. On the other, these types of regulations might increase the overall fragmentation of the news domain because they discourage the creation of emerging business models, i.e new media that promote exposure to diverse content. In fact, they threaten the very existence of these new media outlets[14] (Méndez, 2017) and, ultimately, can have a negative impact on the maintenance of a shared online news domain; this is a long-term effect that should be studied in future research using longitudinal data.

Digital News Initiative, a €150 million program to foster news innovation in Europe (Google Digital News Initiative, 2015).

[14]Since the "link tax" law was passed, in October 2014, the most important Spanish news aggregator, *Meneame.net* has lost over 55% of its total audience. According to *comScore* data the aggregator received 437,657 total unique visitors in September 2014, one month before the approval of the law. This figure has sharply dropped two years later and in July 2016 —the last data we have available– it received 193,000 unique visitors.

3

Meso-level Analysis of the Online News Domain: Identification of Digital News Authorities and Audience Brokers[1]

3.1 Introduction

Great progress has been made in theoretically discussing the impact of the Internet on the news media ecology. Less research attention, however, has been given to empirically testing the changes in the power relations between new and legacy media outlets. This chapter aims to fill this gap by identifying the role of different types of news outlets in the provision of information. It articulates a definition of media power based on the relationship between news media and their potential ability to control the audience flow. The overall goal is to understand the extent to what media's network position confers power to engage wider audiences and to broaden their journalism in the digital domain.

The growth of the popularity of digital news providers, or born-to-the-web media, and the increasing role of the Internet as a main access for news consumption have motived scholars to contend that there has been a reconfiguration of media power. As a result, digital media should be challenging the power monopoly of legacy news brands and new elites should be emerging [2] (Castells, 2009; Chadwick, 2013; Couldry & Curran, 2003; Gurevitch, Coleman, & Blumler, 2009; Hermida & Thurman, 2008; Jarvis, 2016; Pavlik, 2001). Yet, whether this is the case or not

---

[1]This Chapter is based on Majo-Vázquez, S., Cardenal, A.S., Sagarra, O., Colomer, P., (2016) "Challenging Power in the Flow of Digital News: Emergent Authorities and Audience Brokers" (submitted). Previous versions of this Chapter were presented at the 9th *CIS-Harvard Summer Seminar on Sociological and Political Research*, 2016, Harvard University; *Social Media & Society Conference 2016*, London; *Internet, Law and Politics Congress UOC*, 2016, Barcelona. The authors thank discussants at these occasions for their comments.

[2]We have explicitly excluded social media platforms like *Twitter* and *Facebook* from this study, as we did in Chapter 2, because they do not respond to the same editorial guidelines as news organizations. They are platforms of news distribution; they do not produce news content. In Spain, traditional media still dominate the access to news. Nonetheless, it is important to highlight the increasing role of social media sites (39%) as an entry point of those who seek news information online (N. Newman, Levy, & Nielsen, 2016). The Introduction to this thesis provides an in-depth discussion about the differences between social media platforms and news organizations.

remains largely untested. There are a few empirical studies to support this line of work (Meraz, 2009), and they are contested by evidence that shows that important media functions like gatekeeping, agenda setting and framing are still largely in the hands of traditional media (Lee, 2007; Messner & DiStaso, 2008; Vergeer & Franses, 2016).

We know that legacy news organizations are struggling to deal with the endless technological upheaval and to secure their positions as leading information providers in the online domain (Althaus & Tewksbury, 2000; Cagé, 2016; Downie & Schudson, 2009; Fenton, 2010; McChesney & Pickard, 2011). Their paper editions' readerships shrink and so do their revenues (Anderson, Bell, & Shirky, 2012). At the same time, unlimited online competitors make it more difficult for traditional news media to engage a new loyal public, especially when it comes to paying for content and attracting a young target audience (Cage, Viaud, & Herve, 2015; Pew Research Center, 2015).

Several recent examples speak to the efforts that legacy media are carrying on to protect their role as elites (Castells, 2009; Chadwick, 2013; Cook, 1998; Garnham, 1995; Meraz & Papacharissi, 2013) in the digital age. In the UK, for instance, the *Independent* has announced an unprecedented digital-only format movement (Independent, 2016). It has ceased its print edition to focus on attracting more readers to its online platforms. Similarly, one of the top Spanish news brands, *El País*, informed its newsroom about the decision to terminate its paper edition in the short-term in favor of adopting a digital-only strategy too (Caño, 2016). Not too long ago, *The New York Times* (Times, 2014) and more recently, *Der Spiegel* (SWR, 2016), commissioned innovation reports that contain detailed self-criticism analyses for failing to successfully embrace the digital sphere. *The New York Times* put it in the following words: "We have always cared about the reach and impact of our work, but we haven't done enough to crack that code in the digital era" (2014:3). Finally, *The Times* has allegedly attempted to better serve their digital readers by implementing a new web-updates strategy (Witherow & Ivens, 2016). With no other precedent among leading news brands, the British newspaper has decided it will only release news at peak-traffic time. It will result in three fixed-time digital editions that, according to the editors, aim to give more insightful and reliable reporting online (Witherow & Ivens, 2016).

These recent decisions speak to the importance that legacy media give to the online domain. Nonetheless, citizens increasingly turn to the Internet—including social media—as their main access to political information. In Finland (46%), Australia (44%) and USA (43%), online outlets are already the preferred sources for news (Newman, Levy, & Nielsen, 2015).

Underlying the movement of the legacy brands described above—all aimed at concentrating efforts on the web—there is a broad question related to the function that their main competitors, native digital news outlets, are playing in the provision of information. The last U.S. electoral campaign brought some examples of this new competition. For instance, in an interview with the Democratic candidate Hillary Clinton, published on the front page of *The New York Times* (Leibovich, 2016), a reporter wrote that he was waiting for her to finish a previous encounter with *Snapchat*, the American-based new media brand that combines instant message service with news provision and aggregation. This example illustrates that in the current hybrid media system (Chadwick, 2013), there is a confluence of older and newer media and a new allocation of roles among them. Yet, whether this implies a decline in power of legacy media and the rise of new elites, i.e. digital media, must be put to an empirical test.

Therefore, the main question guiding the meso-level analysis of Chapter 3 is: Are legacy media still the most powerful actors in the online news domain as they have been in the offline sphere? Their audiences' rates might prove they continue to be. Overall, traditional media still retain higher portions of attention, measured in number of unique visitors. According to *Alexa* ranking of top visited sites, legacy media in the US, the UK, France, or Germany, to name just a few cases, are ahead of the new media (Alexa Internet, 2014). The same is true in the case of Spain (Alexa Internet, 2015), which is our case of study. However, we argue that the capacity for generating traffic is merely one factor that determines the ability of a news source to hold power in the online domain (Osborne, 2015). As our argument goes, it is also necessary to be a gatekeeper in control over information (Shoemaker & Vos, 2009; White, 1950) and, even more important, to be considered a provider of unique content by digital audiences and by your peers, i.e. news providers. In other words, powerful news sources are social institutions that operate with some authority (Robinson, 2007:306) within the online domain.

We propose an observational approach to evaluate the actual power of legacy organizations. In this chapter we also use network science's tools to disentangle the allocation of roles between legacy and new media. We draw on the methodological framework by Kleinberg (1999) and partially build on previous work by Weber and Monge (2011) to measure the power of news sources. We reproduce the hyperlink network connections among news providers and measure the authority (Kleinberg, 1999) of the actors involved in the online journalistic practices. We bring empirical evidence that legacy media retain control of the most powerful positions in the online news domain. In other words, legacy outlets are by far endorsed as authorities or sources of reference. To further substantiate their power, we also examine the network of audience among news publishers. As we did in Chapter 2, we gathered online navigation data from a representative panel of 30,000 people

in Spain and reproduce their patterns of news consumption. Our results show that legacy news media hold the most powerful positions in controlling the audience flow. Interestingly, though, the analysis focused on patterns of young audiences reveals that new media also occupy broker positions (Fernandez & Gould, 1994) and, hence, are able to control the flow of audience navigation across different news providers.

In what follows, we review the previous work on the impact of the Internet in the news media ecology, paying special attention to the role of new and legacy outlets. We offer a detailed analysis of the Spanish media sector and the growing number of new digital providers recently created. We then introduce the data and methods and discuss their potential to tackle the relationships among different types of news organizations and measure media capacity to operate with power. In the results section of Chapter 3, we present the outcomes of the meso-level analysis of the online news domain. Finally, we discuss our results, focusing on the changing patterns of news consumption among the youngest audiences.

3.2 The Power of News Organizations in the Digital Age

The interaction between new and legacy media organizations and the new allocation of power among them have lately spurred much interest. In the early years of the 21st century, Dutton, Gillett, McKnight, and Peltu (2004) already advanced the disruptive impact of digital technologies on the reconfiguration of power among several types of actors and organizations. Media institutions have not been an exception. Some scholars have argued that the web has created new spaces for power (Bennett, 2003), new hierarchies have emerged in the online domain (Mansell, 2004), and, ultimately, the web has recast the roles of the actors in the media ecology (Castells, 2009; Gurevitch et al., 2009).

To secure their positions as elites in the political communication process, legacy media took advantage of scarcity, exclusivity and control of information for much of the 20th century (Lewis, 2012:311). Today, as then, journalism continues to expose corruption, draw attention to injustice, and hold politicians and businesses accountable (Anderson et al., 2012). Within these lines, and to name just one example, one may take the *Panama Papers* (ICIJ, 2016), a coordinated effort of more than a hundred journalistic organizations. In reporting such information, and from a normative point of view, legacy organizations accomplish an essential function for the democracy. They foster political knowledge (Curran, Iyengar, Lund, & Salovaara-Moring, 2009) and individual deliberation, they guarantee a plurality of points of views and the public debate of ideas. They attach importance to the

range of issues that fill the public agenda (McCombs & Shaw, 1972) and, in doing so, put order to the social reality. These functions are deeply rooted in the ideal of public sphere that Habermas (1994) envisioned for a democracy. Likewise, by controlling information dissemination in a structured environment, legacy media shape the social order. This is what ultimately has allowed them to operate with some authority (Robinson, 2007:307).

3.2.1 New Actors in the Media Ecology

The digital technologies have, however, shaken up the traditional role of the legacy media in the society. The web gives audiences greater access to a broader range of facts, data, and opinions and changes the traditional journalistic standards for the provision of news (Fortunati et al., 2009; Riordan, 2014). According to this research strand, communicating what matters to the public is no longer a lineal process from sources to journalist to be finally transmitted to mass audiences. Instead, the Internet has created a new distinctive scenario where providing information and news is not an exclusive task of legacy organizations anymore (Croteau, 2006). The mass-self-communication system (Castells, 2007) or the many-to-many model of information theoretically puts the smallest news providers on an equal footing with the transnational conglomerates (Castells, 2009; Fenton, 2010). Consequently, people do not just rely on traditional media to make sense of the myriad of information around them. Information control or gatekeeping, according to this strand of research, are now shared either with audiences that perform as content producers or with a wide range of new actors, some of which act as journalists, providing newsworthy content too. All of them are what we indistinctively categorize here under the labels of new media or digital news outlets, e.g. blogs, aggregators, niche online outlets or fact-check sites.

There is no way to look at these new actors and see anything like coherence, claim Anderson, Bell and Shirky (2012). If there is communality among them, it is that they are born-of-the-web and that –similarly to legacy organization– can reach a massive audience. Even a blog by one single author can reach thousands of people today, although they have much smaller structures. This is the case, for instance, for *talkingpointsmemo.com*, a political blog founded by Josh Marshall during the U.S. elections in 2000. It reached 400,000 viewers per day in 2008 (Aldred, 2008). No wonder the web has radically lowered the cost of distribution (Cage et al., 2015).

Likewise, digital outlets have diversified standard journalistic narratives, formats and rituals. They provide easily sharable content (Riordan, 2014), multimedia news products, and take advantage of new forms of collaboration, analytic tools, and sources of data (Anderson et al., 2012). Equally important, digital outlets

have quickly adapted their standards to fit the increasing trend of mobile news consumption, which is paired with a decline in desktop consumption and growth in video news consumption online. News accessed from smartphones jumped significantly in 2015, particularly in the UK (42%), US (44%), and Japan (43%). On average, the weekly usage, in twelve countries studied by *Reuters Institute* (Newman et al., 2015), grew from 37 to 46%. In Spain, the growth in mobile news consumption was 48% that year.

Among all the changes brought about by the advent of the Internet into the media sector, though, Livingstone considers that the potentially most radical change is the shift from one-way mass communication towards more interactive communication (1999:61). New digital outlets have given an active role to the audiences in the production of information. They have blurred the distinction between producers of news content and the audiences themselves (Croteau, 2006), a relationship well established for decades by legacy news media. They engage diverse actors in horizontal and conversational practices to filter and promote content, blending broadcasting with social conventions (Meraz & Papacharissi, 2013:142). Similar to *Digg* and *Reddit*, new digital outlets in Spain are "crowdsourcing the collective intelligence" (Meraz & Papacharissi, 2013). They go beyond the comments' section in legacy news media webs pages or the "have to say" initiative by the *BBC* (BBC, 2016). *Ctxt.es*, a small Spanish digital news site—it had 62,000 unique visitors in December 2015, according to *comScore*—promotes this type of new interactive process. Its audience can become owners of the organization and participate in editorial meetings. It follows the steps of previous new media outlets in Spain, such as *Lamarea.com* and *Critic.com*, that also invited their online readers and patrons to become members of their editorial board. In doing that, they share and transform journalistic routines, once considered exclusive to legacy news media, like gatekeeping and agenda setting, with a broader set of actors and the audience too.

Alternative sources of funding have also become a growing mechanism to differentiate new and legacy outlets and to interact and engage with the public. Beyond the classical sources of media funding, such as advertisement or even government subsidies (Brogi, Ginsborg, Ostling, Parcu, & Simunjak, 2015; Colino, 2013), new media have also sought alternative sources of income to sustain their journalistic practices, highlight their independence, and increase audience participation. In 2008, 17 journalism projects were launched in one of the largest crowdfunding platforms in the world, *Kickstarter*. This number grew to 173 in 2015 (Vogt & Mitchell, 2016). In all, those projects raised nearly $6.3 million across sixty countries. The figure trails any other category in that platform for the same period of time.

Some Spanish examples along these lines include the web-based newspaper *El-diario.es*, which reached 1.7 million unique visitors in December 2015, according

to *comScore*. It is owned by the same journalists who work there and has more than thirty thousand subscribers (Minder, 2015). Recently, *Ctxt.es*, the digital outlet mentioned above, also launched a crowdfunding initiative to raise €77,000 (Ctxt.es, 2016). So too did the Catalan outlet *Critic.com*, dedicated to investigative journalism (Verkami.com, 2015) and obtained €45,000 from 865 patrons and *Lamarea.com*, raising €33,000 in 2012 (Maspublico.org, 2012).

Elconfidencial.es, the leading digital-born site in Spain, which reached almost 3 million unique visitors in September 2015, is owned by journalists and individual investors, but any private corporation can take part of its ownership. Additionally, it raises funding from organizing informative events (Elconfidencial, 2015). Certainly, though, the most outstanding case among the big digital outlets is *Elespanol.com*. It raised €2.2 million in a crowdfunding process in 2015 (Elespanol.com, 2015). This new digital media has been founded by the former director of the second largest legacy outlet in Spain, *El Mundo*, and it combines a mixed model of financing its operations, which includes subscribers and crowdfunding resources from over four thousand people and private investors (Eldiario.es, 2015). Beyond the Spanish examples though, one can also take the case of *ProPublica* in the United States. This new media ended 2015 with $450,000. In 2016, that amount had grown to $2.9 million (Mullin, 2017).

3.2.2 The Spanish News Media Landscape

In Spain, the popularity of fundraising campaigns among new outlets is directly related to two facts. First, only three companies currently control 58% of the Spanish media market and the media system is considered to be at medium risk due to the lack of transparency of the media ownership structures (Brogi et al., 2015). In other words, the audience cannot keep track of the identity of the private sponsors of a great portion of legacy media outlets. And second, the search for alternative funding sources—a growing trend among new digital outlets in Europe—is related to the low levels of trust towards journalism across countries. According to Edelman Trust Barometer, only 47% of the population, on average, trusts news media across 27 countries (Edelman, 2016). This figure accords with low levels of trust shown towards journalists in the last quarter of the century in Spain and other countries, like the UK (Davis, 2009). There, journalism is the fourth least-trusted profession out of 24 categories (Mori, 2015). On average, 49% of the Spanish population trusted news media in 2016. This figure represents an increase of 7 points with respect to the previous year, but it is among the lowest in the Edelman study.

The Spanish media are overtly partisan (Hallin & Mancini, 2004), and the overall system cannot be understood without taking into account the significant impact

of the economic crisis that hit the country from 2008 to 2012. Since then, legacy media have struggled to overcome both the media crisis and the economic crisis. Between 2008 and 2013, the unemployment rate for Spanish journalists rose 132%, the communication sector lost 11,151 jobs and 284 news media shut down during those years, 73 of which did so only in 2013 (APM, 2013). On the upside of this adverse scenario though, we find that 458 new outlets were founded since 2008, the vast majority of them online (APM, 2015). Most of them are lead by journalists who used to work for legacy organizations (Minder, 2015; Schoepp, 2016). Among several other cases, one can find the aforementioned *Elespanol.com* and *Elnacional.cat*, founded by the former director of the fifth top-read Spanish newspaper, *La Vanguardia*, after being laid off.

In sum, the technological changes brought by the web have paved the way for a new type of news providers—digital news media—whose distinguishing characteristics have been outlined above. They have proved to be more keen on embracing the constant digital innovations and responding to increasing demands for more participative and transparent journalistic practices. However, their power in the overall flow of information is yet to be empirically assessed. Theoretical accounts of the reallocation of power in the digital domain contend that new media have shown their ability to transition from volume to value, and that authority has been conferred to them by building relationships with people based on relevance (Jarvis, 2016). If that were the case, new elites would be emerging in the media ecology. Yet, we still lack empirical evidence to prove such stances, partly because, as some scholars have already pointed out, it is necessary to make explicit a clear conception of power in the media ecology (Mansell, 2004). To assess media power in the online domain, we take into account the relationship among news providers and the potential control of audience behavior. Hence, we first identify authorities—i.e. those media outlets that are recognized as reputable sources of information by their peers—within the network of news providers. This process delves into the following research question:

Research Question 3.1: *To what extent are new digital outlets, by being recognized as authoritative sources of information by other news providers, central actors in the flow of news content?*

Yet, as we argued before, power in the online domain is multidimensional, and this articulation leads us, secondly, to assess whether authoritative sources of information also have central positions in the control of the audience flow. In the digital news domain, power holders should not only be regarded as valuable sources of information, but also should have potential control over the way audience navigates the news domain. This idea is illustrated in network theory by measuring betweenness centrality of nodes. Media outlets with high betweenness centrality lie

on the audience shortest path between other news sources and therefore, may have considerable influence over the overall audience behaviour. To have a comprehensive understanding of the elites or power holders in the digital news domain, we also assess brokerage power of news media outlets mapping the network that represents how people navigates news content online, i.e. the audience network. This leads to our following sub-question:

Research Question 3.2: *To what extent do new digital outlets and legacy organizations control brokerage relations in the audience network?*

Finally, relying on the previous literature, we expect that the underlying power structure in the audience network might differ if we consider specific patterns of news consumption across demographics. Studies that look at the news uses of young generations unveil distinctive media diets for this target (Mitchell, Gottfried, Shearer, & Lu, 2017; Tewksbury, 2005; The Pew Research Center, 2015). For instance, young people are more keen on using social networks and mobile devices to keep up with current affairs (Newman, Levy, & Nielsen, 2015a). They are also less likely to remember news sources names (Mitchell et al., 2017), which might imply that brand awareness, which is especially important for legacy outlets, is less effective in attracting and engaging with young public. Further, news publishers are setting up new channels to reach young audiences, which, in turn, increases the competition against legacy brands. In light of this specific scenario, our final research question in this chapter examines the power of news providers across different age groups:

Research Question 3.3: *Is media brokerage power held equally across ages in the news domain?*

3.3 Data and Methods

3.3.1 News Network

In this Chapter 3 we use two different techniques (more on this to follow) to analyze two types of networks: the news network and the audience network. We study five networks in total–four audience networks representing news consumption patterns of different age groups and one news network.

The news network represents the total number of hyperlinks that news media send to each other and is weighted and directed. Here, as we have explained in the previous Chapter 2, nodes are news media outlets and an edge between two nodes exists if there is at least one hyperlink sent from one outlet to an external source.

Two media outlets are disconnected if none of them has cited content from the other. The weight of the edges equals the total number of hyperlinks among each pair of nodes.

We argue that a hyperlink is a conservative measure of the level of autority that news media confer to competitors. Previous research has shown that linking has a very high cost barrier in the media ecology. News media rarely link competitors to avoid guiding audience attention outside its boundaries (Chung, Nam, & Stefanone, 2012; Larsson, 2013; Napoli, 2008; Tsui, 2008). Therefore, and following previous research in the field (Kleinberg, 1999; Weber & Monge, 2011), we assume that when they do send a hyperlink to another news outlet, it is to acknowledge the unique value of its content.

To build the news network, we start by selecting the top 44 news outlets in Spain in February 2015. We used *Alexa* rankings (Alexa Internet, 2015) to choose these first seeds of the crawling. *Alexa* limits the publicly available information to the top 500 sites by country. From this list, we only selected sites published by Spanish news organizations. Several prior studies have relied on *Alexa* rankings to obtain traffic information (Ennew, Lockett, Blackman, & Holland, 2005; Price & Grann, 2012; Wu & Ackland, 2014). However, to further test the accuracy of *Alexa* data, we compared the initial ordering with the list provided by comScore, the official digital audience meter in Spain. The correlation of the two rankings was 0.906, which speaks for the accuracy of the initial list. Then, as our main interest was in comparing roles between new and legacy news media, we added top digital outlets in Spain using the *comScore* ranking. These outlets were not included in our initial list due to their total audience fall short of the threshold for publicly available data provided by *Alexa*. We ended up with a list of 100 news media that we input as seeds in a commercial sofware for crawlings (Ackland, 2005). The web crawl was supervised to ensure that we extracted the maximum number of incoming and outgoing links from the deepest level. The process involved snowballing from seed sites to get outgoing hyperlinks, but also querying a search engine (*Bing*) via its application program interface (API) to collect incoming hyperlinks. Our approach is consistent with previous research that shows that selecting seed sites according to well established prominence (in our case, size of online audience) minimizes the bias of rendering peripheral sites as central (Weber & Monge, 2011).

The first resulting network had close to 17,000 sites. Yet, many of them belonged to the same original parent webs. The cleaning process included pagegrouping subdomains with their parent sites and pruning those nodes that were not news providers. The final network has 100 nodes and almost 800 edges or connections.

Tab. 3.1.: Descriptive Statistics of the News and Audience Networks

Statistic	Web Network feb-15		Audience Network nov-15	
	Full	News Network	Total Audience	Young Audience
Size	16659	100	108	93
Edges	32628	799	10387	8464
Density	0.000	0.081	0.890	0.989
Reciprocity	0.046	0.471	0.980	0.989
Clustering	0.006	0.367	0.930	1.000
Max. Inlinks	1223	32	38	92
Min. Inlinks	0	0	106	91
Max. Outlinks	704	49	0	92
Min. Outlinks	0	0	106	0
Max. Instrenght	2793	1173	44150	5199
Min. Instrenght	0	0	0	0
Max. Outstrenght	1282	447	42120	2946
Min. Outstrenght	0	0	0	0
Assortativity	-0.615	-0.166	-0.11	0

Note: Density is calculated by counting the number of realized links relative to potential links. Reciprocity measures the proportion of mutual links and clustering summarizes the number of closed triangles in the network, which serves as a measure of local density in link formation. In and out strength scores add up the tie weights coming in or out of each nodes. Assortativity degree equals 1 when edges only connect vertices of the same category and a negative coefficient signals that there is a tendency of peripherals nodes to connect to very central sites.

Table 3.1 summarizes its main descriptive statistics. The negative assortativity indicates that nodes with a large number of links tend to connect to those that have a small number of links. To better know the structure of the news network we also analyse its in-strength distribution (see Figure A.1 in the Appendix), and it reveals that most of the vertices have low in-strength (i.e. number of hyperlinks received from news outlets) and that there are few vertices which are receiving a higher number of hyperlinks, which results in a right-skewed strength distribution. The node with the highest in-strength is the online aggregator *Meneame.net* which receives 1,173 hyperlinks from the other news sources on our network.

3.3.2 Audience Network

Finally, we built our four audience networks in November 2015, drawing on the previous work by Webster & Ksiazek, (2012). Ties here represent shared audience among news media outlets. In other words, ties stand for the total audience overlap between media i and media j. The audience networks are directed and weighted too. Because our data collection process for mapping the hyperlink and the audience networks had to take place at different points in time, we tested the volatility of the

audience data by comparing media reach across one year period. Figure 3.1 shows that we can assume a pretty stable audience behaviour during the period of time that our study spans with no major changes in the patterns of news consumption between February 2015 and November 2015. Additionally, we refer to Figure 2.2 to further show that centrality scores of media outlets in our study are pretty stable too.

Fig. 3.1.: Correlation of Media Audience's Reach Over Time
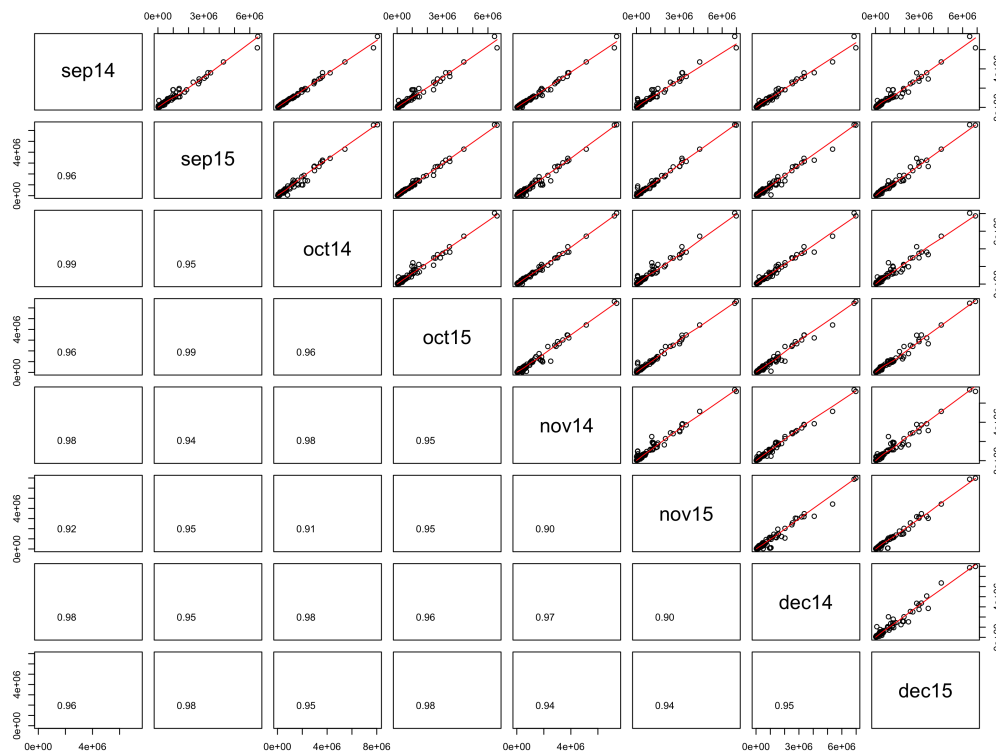
We gathered the audience data from the official digital audience meter for Spain, *comScore* (Interactive Adversiting Bureau, 2011). As explained in Chapter 2, the company tracks the browsing behaviour of a representative panel of 30,000 Spanish people and combines these observations with other indicators, drawn with direct audience metrics from news media collected by a separate agency (AIMC, 2014), to check the accuracy of the measurements. The audience networks represent observed data instead of reported news media diets, which have been proved to be overstated and result in less accurate analyses (Guess, 2015; Prior, 2009). More importantly, and to the best of our knowledge, *comScore* data is the best proxy available to study aggregated patterns of news consumption online. The company only provides audience overlapping data for those sites that have been visited by a minimum of 16 panelists in a given month. As a consequence, the total number of nodes of the four networks varies. We reproduced the online audience navigation patterns for the general population and people aged between 18–24, 25–34, 35–54 and

more than 55 years old. Table 3.1 summarizes the basic statistics for the network representing the Spanish general population and the young group, the two principal targets of interest in this chapter. They vary in size because not all sites in our sample attract young public; they do, however, share many characteristics, like a high level of reciprocity and density, which means that, in line with our previous results, audiences widely navigate the media landscape. The audience networks also exhibit a core periphery structure and a right-skew strength distribution (see Figure A.2).

3.3.3 Methods

Our main goal is to identify power holders in the news domain. To this end, we use two different techniques borrowed from network science. We first measure the level of authority of news media using the *hyperlink-induced topic search* or HIT algorithm developed by Kleinberg (1999). Following the previous work in the area of Link Analysis Rank algorithm, we assume that a hyperlink from node i to node j denotes an endorsement for the quality of the page j (Borodin, Roberts, Rosenthal, & Tsaparas, 2005; Kleinberg, Kumar, Raghavan, Rajagopalan, & Tomkins, 1999; Weber & Monge, 2011) and that the authority score is a proxy to identify the most relevant news content online. HIT algorithm provides a two-level propagation scheme, where endorsement is conferred on authorities through hubs. In other words, every page has two identities: the *hub identity* captures the quality of the page as a pointer of useful news sources, and the *authority identity*, our main concept of interest, captures the quality of the page as a resource itself (Borodin et al., 2005 :235). Therefore, we identify news providers' prominence in two roles: authorities and hubs. The former are news media, which provided content highly cited by their peers, and the latter are news media that aggregate the worthiest sources of news information. Formally, the authority centrality of a vertex is defined to be proportional to the sum of the hubs centralities of the vertices that point to it:

$$x_i = \alpha \sum_j A_{ij} y_j, \tag{3.1}$$

where A_{ij} is an element of the adjacency matrix, y_j stands for the hub centrality and α is a constant. Consistently, the hub centrality of a vertex is proportional to the sum of the authority centralities of the vertices it points to:

$$y_i = \beta \sum_j A_{ji} x_j, \tag{3.2}$$

where β is also a constant (for more information see Kleinberg, 1999; M. Newman, 2010; Weber & Monge, 2011). Additionally, we use a random benchmark known as the *configuration multi-edge model* in the network literature (Sagarra, Font-Clos, Pérez-Vicente, & Díaz-Guilera, 2014). It is specifically designed for weighted networks and matches the strength distribution of our news and audience networks. This benchmark aims to assess the significance of the authority and hub scores previously calculated and its departure from what we could expect if the news links were sent randomly. The null model is based on the reconstruction of 1,000 random networks with exactly the same strength distribution as our observed network and maximally random to all other respects. Because our network is directed, the null model is defined separately for incoming and outgoing ties (for more information see Sagarra et al., 2014).

Moreover, because we operationalize power on the online news domain as a multidimensional concept, along with the authority analysis, we also assessed the potential control of news media over the audience flow. To this end, we measured the betweenness centrality score of all the news sites in the four audience networks that represent patterns of news consumption across ages. This approach allows bringing evidence to the role of the news providers as audience brokers. Formally, the brokerage scores are roughly equal to the number of the shortest paths between others that pass through a node. More interestingly, though, the resulting ranking of brokers for each network will identify those news providers that have higher potential control over the audience flow in each age group and can thereby engage wider audiences.

Since Freeman (1979) conceptualized it, in the sociological literature there is a rich amount of work devoted to the study of the brokerage in social networks (Burt, 2000; Coleman, 1988; Fernandez & Gould, 1994; González-Bailón & Wang, 2013). We employed here the formalization of betweenness centrality by Opsahl, Agneessens, and Skvoretz (2010), which generalizes the shortest path calculations for weighted networks matching the nature of our data. Their measure also takes into account the number of ties for the calculation of the brokerage score (for more information see Opsahl et al., 2010). In other words, we did not only take into accounts shortest paths between nodes in our audience network but also the amount of audience that flow through them. As a result, our measure not only accounts for those news sources that more frequently mediate visits to other news sources, but also for those that receive more visits themselves. In order to determine whether the brokerage scores are statistically significant (not due to random fluctuations of the audience flow), we use the configuration multi-edge model too.

3.4 Results

The identification of the media authorities, or the most reputable sources of news content in the digital domain, yields evidence that legacy news outlets broadly control this strategic position. Figure 3.2 shows the authority scores of the two types of news providers under study: new media and legacy media. The distribution of the scores suggests that legacy brands are more frequently considered sources of unique content than new media are. There are only two digital media exceptions among the top ten authorities identified in the hyperlink network: the aggregator *Meneame.net* and the business outlet *Finanzas.es* . One plausible explanation for the outstanding position of *Meneame.net* is that it is pointed to by the biggest hubs in the network, like the broadcasters *Antena3.com* and *Atresplayer.com*, amongst others. These results allow us to answer our first research question.

Fig. 3.2.: Association between the Authority Scores in the News Network and Audience Reach
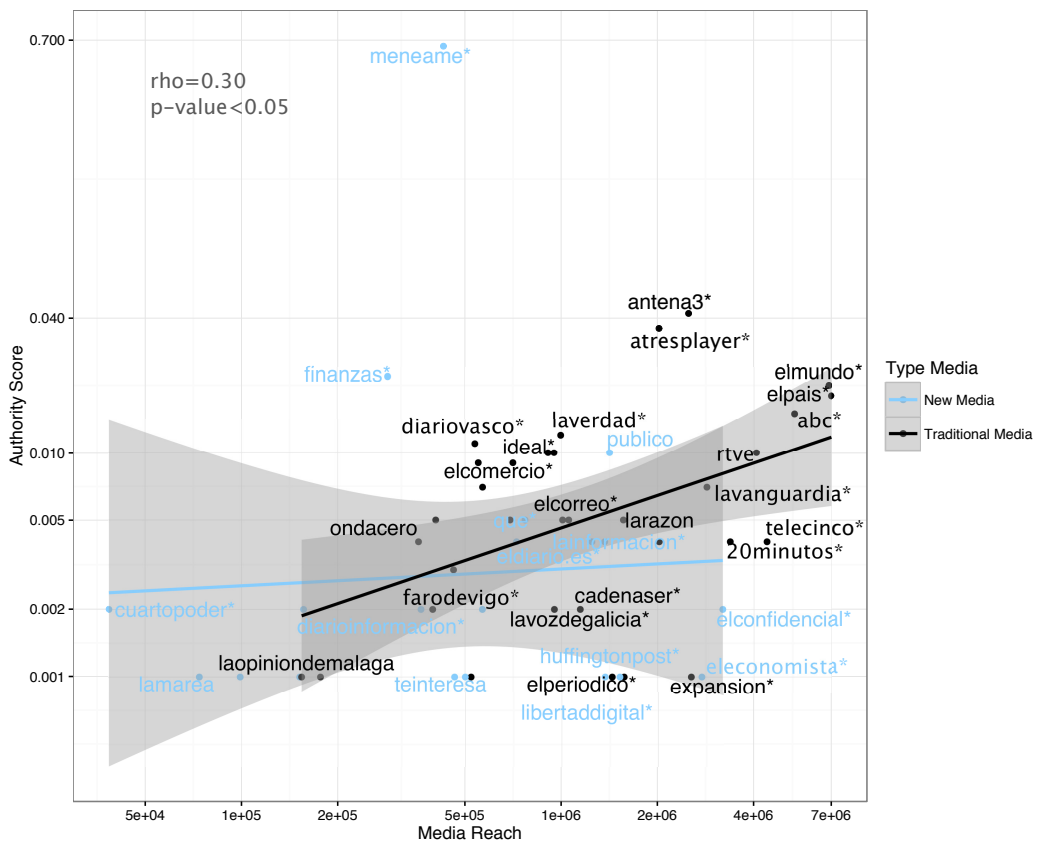
Note: A base 10 *log* scale is used for the *x* and *y* axis. Nodes are identified where space permits. Authority scores are considered significant (*) when the values are within 1.96 standard errors of the average of the estimates scores of authority. We have used linear regression to draw the trend line and the grey bars are the standard errors. Authority scores that equal zero are not included in the analysis. Correlation between authority scores and reach are calculated using Spearman's coefficient and setting the statistical significance at 0.05.

As mentioned before, we have also analysed the extent to which the observed authority scores are different from what we would expect if the hyperlinks of the news network were sent randomly—significant scores are labelled (*) in Figure 3.2. The threshold to determine statistical significance is set at 2σ, or within 1.96 standard errors of the average of the estimated authority scores.

Interestingly, though, we have also assessed the significance of the difference in means of authority scores of new and legacy news outlets. As Figure 3.3 suggests, the type of news provider is not significant to explain levels of authority in the online domain. Taking into account that our observations are not independent, to obtain confidence limits for mean difference, we have used bootstrapping techniques. The coefficients have credibility intervals reported at the 95% and 99% intervals. They are deemed credible at these levels as long as the upper and lower boundaries do not include the value of zero (Muthén & Asparouhov, 2012). These intervals are black for the 95% level and blue for the 99% level. Because the confidence intervals do cross the zero line, we can conclude that the difference is not significant.

Additionally, in Figure 3.2 we can also see that the reach of the news outlets, that is, the total amount of visits media receive, is only slightly associated with authority scores ($rho = 0.30$ p-value < 0.05). In other words, the number of visits that one outlet receives does not offer a complete explanation for its media authority score or the level of recognition of its news content. Rather, we argue, it is the underlying structure of connections that confer legacy media with authority. Their ability to generate valuable content attracts other media's attention and grants them the most powerful positions in the news network.

By means of HIT algorithm, we have also identified hubs in the news network, i.e. pointers to useful information. This analysis reveals that there are more new media acting as pointers to useful news sources, i.e. hubs, than as authorities (see Figure 3.4). Among them, we find *Libertadigital.com, Elconfidencial.com*, and *Diarioinformacion.com*, all born-to-the-web sites that tend to link to sources with high valuable news content. Figure 3.4 shows that some news outlets, mainly broadcasters like *Antena3.com, Atresplayer.com*, and *Lasexta.com*, play a double role in the news domain, they are not only relevant authorities, but they also specialize in identifying the most important news sources on the web and send links to them that increase their overall visibility.

To answer our second and third research question, that is, to find out which type of news provider is in control of the audience flow, we have measured the betweenness centrality of the outlets included in our study. More precisely, we have determined the differences in the brokerage power of new and legacy media across demographics. Figure 3.5 shows the results of the analysis. The main conclusion and answer to

Fig. 3.3.: Difference in the Authority Means by Type of Media and Confidence Intervals
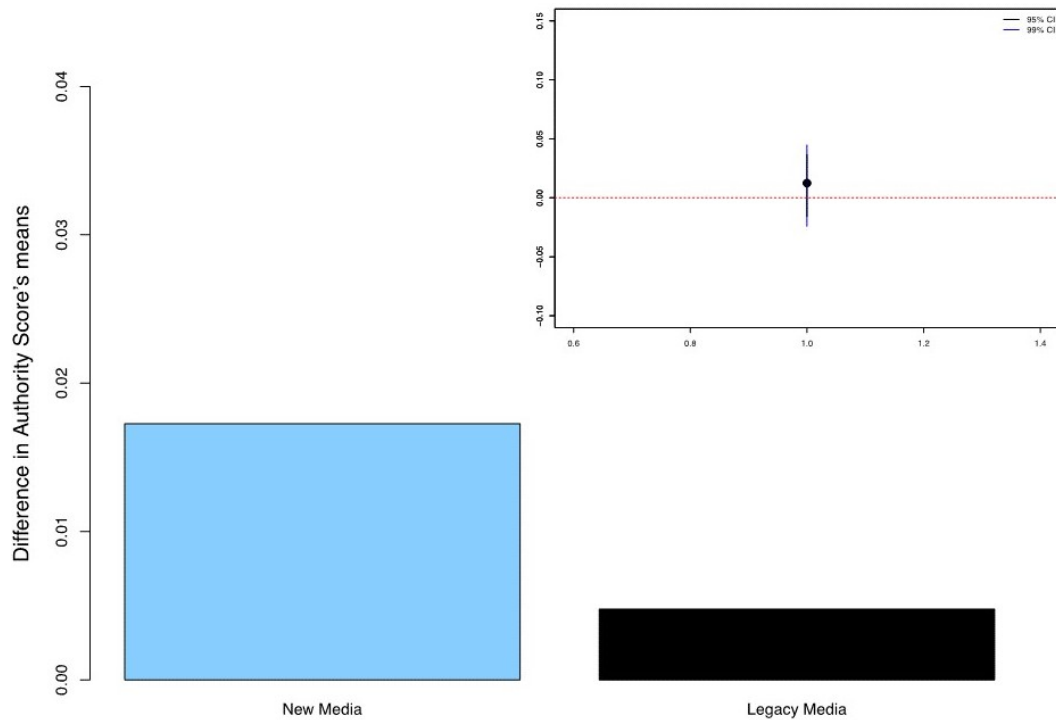

Note: The main graph plots the difference in authority means of the new media and traditional media in the news network. The graph on the top shows the confidence intervals around difference in means of authority scores. According to the result, the difference in the authority scores of new and legacy media is not significant.

question 2.2, is that the control of the flow of audiences is almost a monopoly of the legacy organizations. They hold top brokerage positions in each age group.

Yet, when we drill down to patterns of navigations of young audiences, those aged 18 to 24, the analysis reveals that new media obtain, on average, a higher betweenness centrality score than in other age groups. According to our results, three digital media are playing as the top brokers for young audiences: *Elconfidencial.com*, *Eldiario.es*, and *Eleconomista.es*. Figure 3.6 maps their positions within the strongest connected component of the young audience network. The figure shows the most important brokers. Here, nodes represent news media outlets and their size is proportional to their brokerage power. To obtain the strongest connected component of the structure, we have iteratively removed the weakest connections of the young audience network, a process known as *percolation* in network science (Abbar, Zanouda, & Borge-Holthoefer, 2016; Albert, Jeong, & Barabási, 2000; Borge-Holthoefer & González-Bailón, 2015). The advantage of this method is that it reveals the areas with the highest levels of audience overlapping and helps us to better understand patterns of news navigation between new and legacy media.

Fig. 3.4.: Association between the Hub Scores in the News Network and Audience Reach
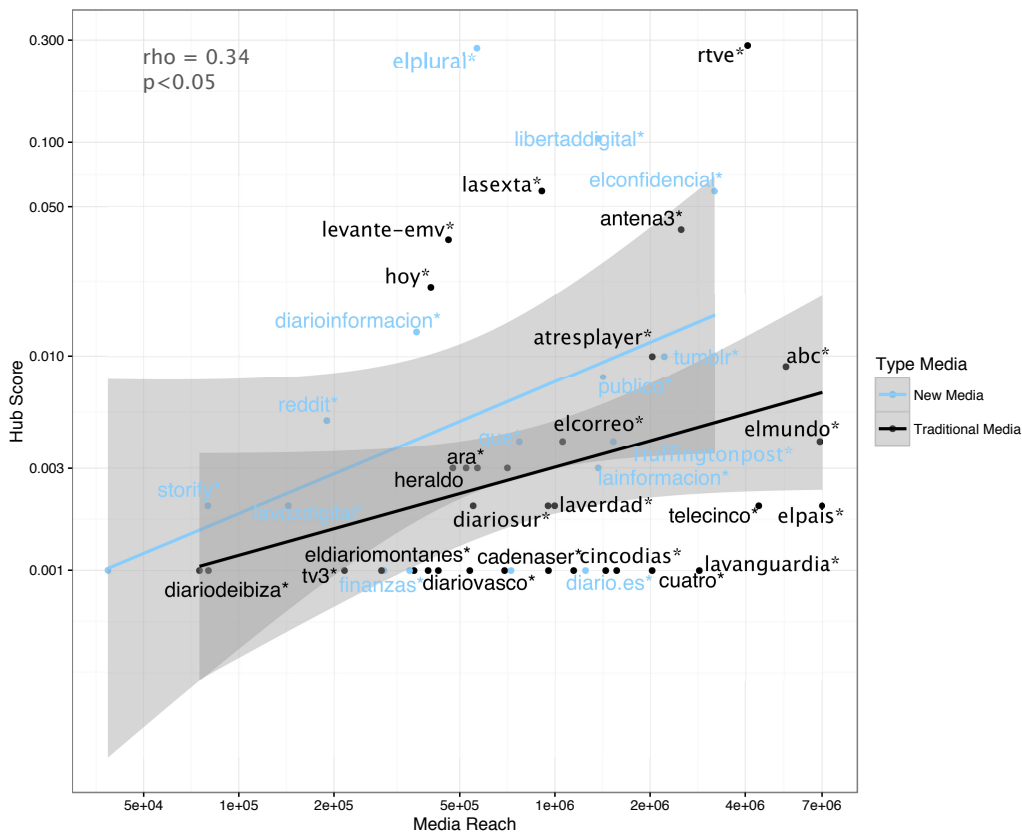

Note: We have used a base 10 log scale for the x and y axis too. Nodes are identified where space permits. Hubs scores are considered significant (*) when the values are within 1.96 standard errors of the average of the estimates´ hub scores. We have used linear regression to draw the trendline and the grey bars are the standard errors. Hub scores that equal zero are not included in the analysis.

Overall, and in line with previous research, we can see that patterns of young audience navigation differ from those of the general public (see Figure 3.7) in that young people include more born-to-the-web sites in their media diets. These types of news sources compete for the control of the audience flow with legacy brands. They occupy core positions within the network and have major roles in the mediation of news attention. The findings allow us to answer the final research question in the current chapter.

After removing the weakest connections of the general audience, i.e applying the percolation process, we obtain the network map on Figure 3.7. The resulting structure shows the distribution of brokerage power in the general audience network and is congruent with our previous findings: The map also highlights the greater role of legacy outlets–see the greater size of dark-blue coloured nodes–as mediators of the audience behaviour in the online news domain. *Elmundo.es* and *Elpais.com* are the top brokers in the general audience network. They are also the most important legacy brands in Spain in terms of audience reach and power in the

Fig. 3.5.: Mean of Betweenness Centrality Scores in the Audience Networks by Different Age Groups
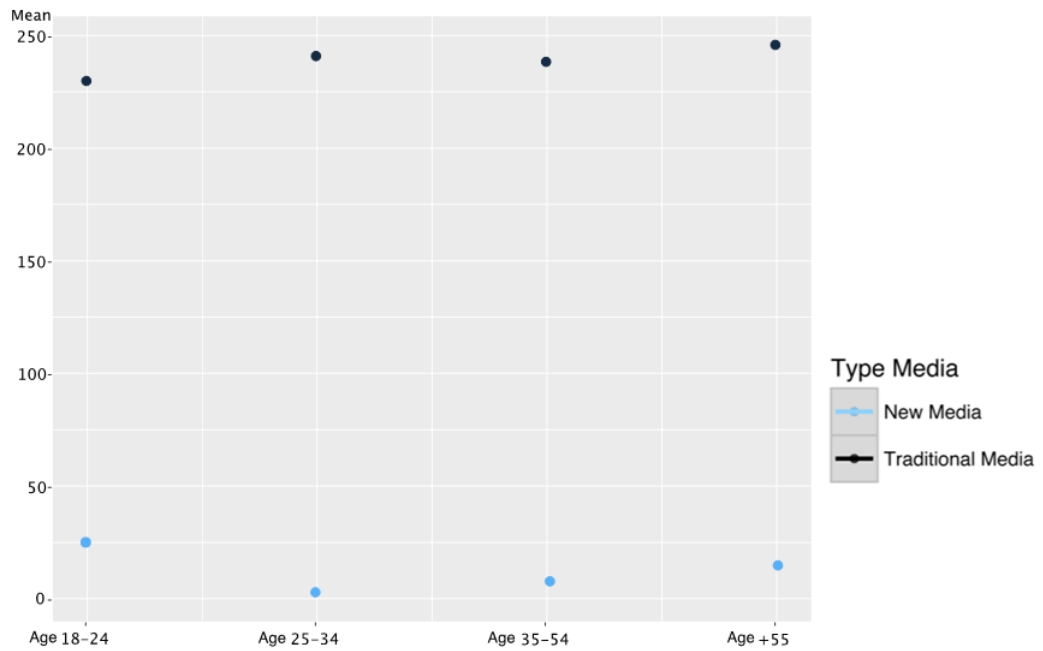

Note: The figure shows the mean of betweenness centrality scores of new and legacy media after analyzing the audience networks representing four different age groups. We disaggregate news consumption patterns by age to determine the differences in the brokerage power of new and legacy media across demographics.

Fig. 3.6.: Brokers of the Young Audience Network
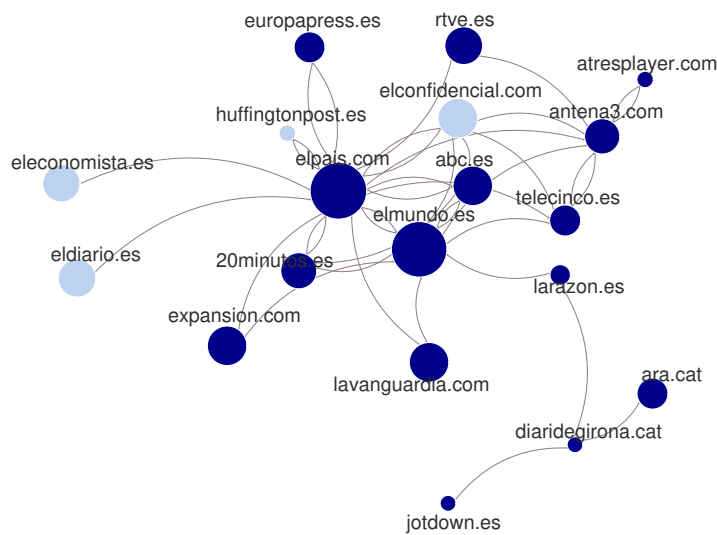
Note: This figure maps the strongest connected component of the young audience network, those aged between 18 and 24 years. Nodes represent news media outlets and their size is proportional to their brokerage power i.e., betweenness score. Ties represent audience overlapping among news providers. A tie is send from site i to site j when a share of the audience of the former also visits the latter. Hence, the network is directed and weighted.

Fig. 3.7.: Brokers of the General Audience Network
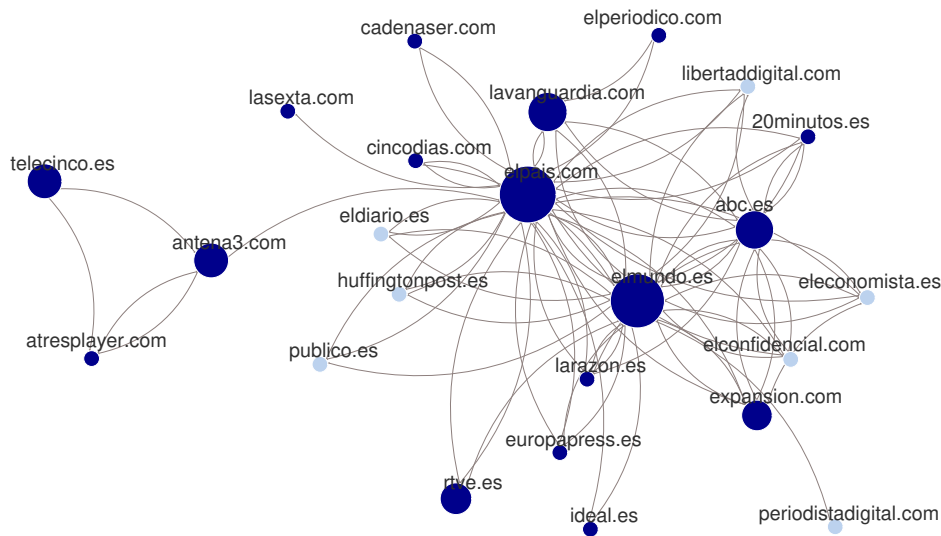
Note: This figure maps the strongest connected component of the general audience network. Nodes' size is proportional to their brokerage power, i.e. betweenness score, and they are colored according to the type of news providers: dark blue are legacy media and light blue are digital news sources.

political communication process. Following them, we find *Lavanguardia.com*, *abc.es*, and the Spanish broadcasters. All are news sites that predate the Internet era. In light of this evidence, we argue that offline heritage is still a key element to understanding media power in the news domain.

3.5 Discussion

In this study, we have argued that the power in the online domain hinges on the dominance of authoritative positions in the news network and brokerage power in the audience flow. We have mapped the network structure between news providers, which has offered a useful lens to interpret the actual underlying relationships among news outlets based on their hyperlink structure. Our results prove that legacy media are regarded as the most authoritative sources and that they retain much of the control of the audience flow too. However, young audiences, who exhibit distinctive patterns for news consumption, confer brokerage power to native digital outlets.

The contribution of this chapter is twofold. First, it offers an empirical framework that can be applied regardless of the media context to unveil power relationships in the media ecology. In doing so, it helps to advance our understanding of how the Internet impacts the role of legacy media as elites in the political communication process. Second, this chapter has implications for the literature on the reconfiguration of the

media ecosystem by providing directed evidence that, to the best of our knowledge, had not yet been available on the actual role of the legacy news media as elites in the online domain. Our results suggest that new media outlets are still far from displacing traditional brand outlets, whose stronger offline reputation may be securing their role as elites in the political communication process.

Finally, some limitations deserve consideration. Our data represents snapshots of the news digital domain, and a temporal analysis might be necessary to account for variations in the power positions of the news media across time. Additionally, future studies should also consider more refined measures of media engagement than total audience visits, such as attention measured by minutes when exploring the structural mechanisms for the authority and hubs scores that our study unveils.

Individual Analysis of the Online News Domain: the Role of Media Diets on the Public Agenda[1]

4

4.1 Introduction

Under ideal democracy, people should share a set of common experiences. They should agree on the issues that deserve attention and careful consideration to improve their communities. Raising the expectations for an ideal democracy, they should also display an increasing carrying capacity for the problems they see as politically relevant. And being even more idealistic, in a perfect democracy this wider range of issues should be diverse in terms of the types of thematic categories to which they belong (Maxwell McCombs & Zhu, 1995). In other words, to preserve the ideal of democracy that normative theorists envisioned (Berelson, 1952; Converse, 1964; Habermas, 1994; Rawls, 2009), people should share a public agenda, and this should be diverse, which by definition implies that they should have greater agenda capacity, too.

A predominant consensus among researchers studying the public agenda is that the structure of the media environment is a key aspect for assessing the relation between news diets and the extent to which people share common ground. The current media environment is characterized by the endless number of news sources and the increasing role of platforms of news distribution, e.g. *Twitter*, *Facebook*, *Snapchat* and other content aggregators. The provision and consumption of the news have recently seen a profound transformation. The growing move to distributed content (Newman, Fletcher, Levy, & Nielsen, 2016) and use of mobile technologies, is connected with people being less willing to access news on branded sites. Instead, they increasingly rely on their social platform's newsfeed to keep up with current political events (Nielsen & Schrøder, 2014). Some fear that the affordances of these sites to provide personalized information might have lead to *filter bubbles* (Pariser,

2011) or *echo-chambers* (Sunstein, 2009), where people refrain from confronting counter political views or even accessing news content at all (Prior, 2007; Trilling & Schoenbach, 2012). Pre-existing journalistic routines for sorting out newsworthy content and setting the media agenda are mixed with new professional practices for curating brand media stories and, to a lesser extent, creating new ones (Messner & DiStaso, 2008). Further, audiences are not only consumers of information, but they are now able to create content and distribute it, in theory, on equal footing with those well-known brand sites. This process of shifting from one-way mass communication (Livingstone, 1999) or two-step flow of communication (Lazarsfeld, Berelson, & Gaudet, 1944) towards mass-self communication (Castells, 2009) has resulted in a new hybrid media system (Chadwick, 2013) where legacy, new media outlets and social platforms compete for the audience's limited attention, their need for orientation (N. Y. Lee, 2016; Matthes, 2006; Valentino, Hutchings, Banks, & Davis, 2008), and the power to set the public agenda (Meraz, 2009).

In light of these changes in the media ecology there are pressing questions that remain unanswered regarding the impact of media on the public agenda. Theorists expected, and some empirical studies confirm, that the web has brought the end of the public sphere (Baum & Groeling, 2008; Katz, 1996; Papacharissi, 2002; Sunstein, 2009; Turow, 1998). They conclude that the structure of the digital domain does not favor the existence of public consensus over the most important problems. As their argument goes, the idea of a unified media agenda becomes problematic (Chaffee & Metzger, 2001) in the digital domain, or even impossible, as reverse agenda-setting processes potentially arise (R. W. Neuman, Guggenheim, Mo Jang, & Bae, 2014). Their underlying logic is that the traditional one-way pattern, from mass media to audience, has been transformed into complex interactions that make less evident to identify who set the agenda. Furthermore, competing or fragmented public agendas are bound to emerge in the face of increasing availability of digital news sources (Park, Ko, Lee, Song, & others, 2013; Shaw & Hamm, 1997) subsequently limiting the extent to which people share a set of common experiences.

However, whether or not the web preserves common ground for the discussion of public concerns and challenges the agenda-setting function of news media is still unresolved. Recent efforts to shed light on patterns of news consumption agree on the fact that people widely navigate news content online, and that mainstream outlets or legacy brands are central to their browsing activity (Flaxman, Goel, & Rao, 2016; Gentzkow & Shapiro, 2011; Majó-Vázquez, Cardenal, & González-Bailón, 2015; Webster & Ksiazek, 2012). All these contributions have taken advantage of the traces that people leave behind when navigating the web to identify media diets and overcome consistent distortion from self-reported media exposure data (Guess, 2015; Prior, 2009). But to the best of our knowledge, observed patterns of digital news consumption have not been used before to tackle the impact of digital media

exposure on citizens' consensus over the public agenda. We aim at filling this gap by relating both observed and reported patterns of digital news consumption to perceptions about the public agenda.

We go beyond existing studies by also analyzing the role of digital media not only on public consensus but on agenda diversity and capacity too. One cannot fully understand how the digital domain relates to the public agenda without considering also its impact over the diversity of issues regarded as important problems. Agenda diversity has not been a matter of much research after the emergence of broadcasting television with some significant exceptions (more on this later). Although, in one of the few contributions to the subject, Chafee and Wilson (1977) proved that richer media environments increase the capacity of individuals to regard certain issues as important, there is scarce or no evidence at all on the relationship between online media diets and the number of issues that subjects regard as salient and their thematic diversity. Filling this gap is a second goal of this Chapter 4.

To study how online media diets affect the public agenda, we resort to a combination of both observed and reported data from a sample and a sub-sample of 725 and 372 people, respectively. We take advantage of the structure of our data to first assess differences between reported and observed media diets. We then identify different types of digital media diets by measuring their level of concentration, and analyze the extent to which consumers of online information share public concerns. Our ultimate goal is to assess whether the digital media sphere can preserve the common ground and whether that hinges or not on the type of patterns of news consumption. Then we tackle the relation between type of media diets and the diversity of the public agenda. Our study shows that that online news diets based on outlets with higher shares in the news market tend to preserve the public agenda and are associated with a higher diversity of topics for debate in the public realm.

In sum, the contributions of this Chapter 4 are threefold. First, it provides more direct evidence than is currently available on the types of media diets that people have online by complementing survey with observed browsing behavior data. Using observed data, we find that concentrated media diets tend to predominate online since most people actually converge on the outlets with higher shares in the market of news. In line with previous research, our results also provide strong evidence that recalling past media habits is largely inaccurate. People tend to report higher news consumption activity than they actually carry on and to cite more visits to peripheral news outlets than they actually perform.

Secondly, our study also brings evidence that helps us to set the boundaries of the impact of the unlimited number of sources online on the public agenda. Focusing on types of media diets online, we find -–in contrast to the various arguments

related above— that the unlimited supply of sources and news outlets online does not erode common ground among digital audiences. Furthermore, and somehow counterintuitively, our results suggest that concentrated online news media diets, where mainstream outlets are central, are associated with more diverse agendas and a greater capacity to account for public problems. Congruent with findings of previous research, this study also demonstrates the importance that legacy brands still have on the overall consumption of news content in the online domain. Our results thread through the recent audience studies that have produced empirical evidence supporting the predominant role of news providers predating the Internet era in shaping online news diets (Flaxman et al., 2016; Gentzkow & Shapiro, 2011; Nielsen & Schrøder, 2014; Weber & Monge, 2011; Webster, 2014; Webster & Ksiazek, 2012) and, consequently, as we expect, the agenda-setting process.

Finally, this study goes one step further than existing studies in assessing the role that online media plays on the shared experiences that hold democracies together. While an abundant literature has examined the impact of digital media on -–aggregated and individual-– news consumption behavior, few studies have related observed media diets to actual opinions, particularly to perceptions about the public agenda and its diversity.

In what follows, we first review the previous work on the agenda-setting function of news media. We pay special attention to the latest studies examining the impact of the Internet and fragmented modes of exposure to online news on the long-established relation between the media agenda and citizens' ability to regard an issue as important. We also consider related works that focus beyond legacy and new outlets and unveil the role of social media platforms in the public agenda. Secondly, we describe our data and methods highlighting the advantages of the web to trace actual people's behavior. Finally, we present and discuss our main findings.

4.2 Theoretical Framework

Since the seminal book of Lippmann (1922), communication scholars have investigated whether the news media influence our views of the world. But it was not until the study in Chapel Hill by McCombs and Shaw (1972) that the process of agenda setting was identified and conceptualized as the idea that issues emphasized by the media become the issues that the public thinks are important (Coleman, McCombs, Shaw, & Weaver, 2009). Scholars in several disciplines have largely demonstrated this relationship -–that what public view as currently important corresponds to what media choose to report (Kahneman, 2011; Wanta & Ghanem, 2007). A limited amount of news sources characterized the media environments that were under

study in those works. During the '90s the multiplication of television channels fueled the first fears that nations no longer gather together around public issues, thereby depriving societies of common ground (Katz, 1996). However, before the advent of the web newspapers, magazines and radios still had limited amount of space and time. Only a limited number of news stories could be included in their editions, exclusively offline at that time. "Editing processes were essential and journalists agreed upon newsworthiness criteria to make a small selection from all the events that occurred in one day. As a consequence, there was a high degree of correspondence among the media agenda of different outlets" (Coleman et al., 2009:143), a process studied under the name of *intermediate agenda setting*.

That media environment, predating the Internet era, still granted the consensus over the public agenda. With the advent of the web, though, the number of news sources has tremendously increased. The post-broadcast media environment (Prior, 2007) has predominantly attracted the interest of scholars to assess its impact on aggregated and individual news consumption behavior. The bulk of the research has focused on the opportunities the digital domain offers for cross-cutting news consumption and avoiding ideological segregation (Barberá, 2014; Gentzkow & Shapiro, 2011; Mutz & Young, 2011; Stroud, 2010; Wojcieszak & Mutz, 2009). Yet, much less attention has been paid to empirically assess how media diets, i.e news consumption behavior, relate to actual opinions and, particularly, to one dimension of the public opinion: the perceptions about the public agenda and whether the new media ecology dampens or, on the contrary, strengthens the public agenda. Theoretical discussions underscore the threats that the Internet poses for shared common experiences. An influential body of literature claims that the unlimited multiplication of news sources prompts audiences to be more fragmented and isolated, and thus curtails the function of media to hold people together through sharing experiences (Chaffee & Metzger, 2001; Papacharissi, 2002; Sunstein, 2001; Turow, 1998).

Does digital media weaken citizens' consensus over the public agenda, though? (**RQ4.1**) Previous empirical studies suggest that the agenda-setting effect between the media agenda and the public agenda has become weaker over time and that, as the public agenda diversity increases, the ability of legacy media to set the agenda decreases (Tan & Weaver, 2013). Social platforms have also added a new dimension to agenda setting by giving voice to people (Boynton & Glenn W Richardson, 2016), which calls into question the monopoly of news outlets in setting the agenda and increased exposure to fragmented agendas. These findings must be considered alongside the sharp move to social platforms for getting news content. Both in the United States and the European Union, 46% of people report using social media as a source of news. The percentage rises to 60% in Spain (Newman et al., 2016). Additionally, overall, people are less prone to access news directly from branded sites.

Undoubtedly, new patterns of media consumption, the ability of people to generate news content, and the personalization of consumption processes increase the space for political issues in the online domain, while simultaneously providing the affordances to totally avoid them. But whether it does affect the existence of a public agenda has not been resolved. Coleman & McCombs (2007) using survey data concluded that while agenda-setting effects were somewhat weaker for both heavy Internet users and younger people, they still were significant. Additionally, a high redundancy on the media agenda would counterbalance the explosion of new media sources (McCombs, 2005). In this regard, after analyzing the top 120 english-language blogs in a study about the intermedia agenda-setting, Messner and DiStaso (2008) concluded that 43% of blogs' sources were traditional news media. Their findings were in line with previous research that also found that the blogs' agenda is similar to that of mainstream media and that people were likely exposed to a fairly redundant agenda (Lee, 2007).

Despite the diversification of channels of news provision, television still leads over social networks as the main source of news in a study of seven countries, including Spain and the United Kingdom (Nielsen & Schrøder, 2014). Scholars addressing the relationship between issue salience on television and social networks provide compelling evidence that the former affects the issue salience on *Twitter* (Vergeer & Franses, 2016), which, contrary to what some expected, contributes to reducing the number of problems in the public realm. Hence, social networks might not have transformed news consumption to the degree many have hoped or feared (Flaxman et al., 2016; Flaxman, Goel, & Rao, 2013). Not only is most online news consumption still driven by individuals directly accessing the websites of legacy organizations as Flaxman et al. (2016) shows, but the unlimited diversity of issues in the online domain vanishes in favor of a tendency toward popular issues (Park et al., 2013) that are put on the agenda, mostly by outlets that predate the Internet era. These findings come out from studies using digital data to learn about how people get informed. Ceron, Curini and Iacus (2016) also draw on digital footprints to determine that online news sources anticipate the attention paid to issues on *Twitter*. They used time stamps of almost 200,000 tweets during major political events in Italy and conclude that legacy media still control the agenda-setting process. From a structural perspective, researchers that apply a relational approach to news consumption behavior -–mostly borrowing tools from network science-– show that legacy news outlets are located in the most important positions in the flow of news audience (Majó-Vázquez et al., 2015; Majó-Vázquez, Cardenal, Sagarra, & Colomer, 2016; Webster & Ksiazek, 2012). Thanks to these locations on the "audience networks", they still have the potential to dominate the agenda-setting process and, hence, to promote a media environment where a limited number of issues are predominant.

Following the previous discussion, we would expect legacy media to continue to play a central role in peoples' online media diets ensuring a fairly stable media agenda with limited number of popular issues. This would lead to our first hypothesis:

Hypothesis 4.1 (H4.1): *Online media diets do not erode public consensus over the issues that people believe are important.*

However, the new media environment has still increased the time and space that news media had before the advent of the web. In the '70s, Chaffee and Wilson (1977), who coined the term of *agenda diversity*, already proved that media richness (i.e, communities with local daily newspapers) was positively related to the ability to list a higher number of more diverse problems. Decades later, already in the digital age, Tan & Weaver, (2013) used survey data to add new insights to the original longitudinal study by McCombs & Zhu (1995) on agenda diversity. They brought novel evidence that proved that agenda diversity has increased over time too. However, the aforementioned study exclusively focused on legacy news media to study the agenda setting process and the increasing diversity of public issues. The major contribution of this research is that we directly study the impact of the new digital news environment on the agenda diversity and to this end we use not only survey data —as previous study have already done– but trace observed patterns of digital news consumption too.

There are several reasons to expect the new information environment to increase agenda diversity. In the new media environment, there is more space for reporting on diverse political issues that cannot be included in the offline editions. Also, competition from native-born digital news media outlets and social media has been proved to influence agenda setting by expanding the number and diversity of issues, even if legacy media continue to play a central role in setting the agenda. Furthermore, to compete for audience attention, news media have had to differentiate from each other either by means of their editorial lines and the thematic categories of their pieces of news. At the same time, from a demand perspective, the cost of accessing a wider range of news outlets for the public has decreased enormously, increasing the chances that people will visit more than one outlet online. Therefore, even if people have concentrated media diets online —i.e., they mostly consume legacy media-–, we expect the online news domain to increase the number of issues that individuals consider salient (agenda capacity) and how diverse they are (agenda diversity). This leads to our second hypothesis:

Hypothesis 4.2 (H4.2): *Online news media diets increase the level of diversity of the public agenda.*

4.3 Data and Measures

To investigate how the digital media affect the public agenda this study uses two interdependent sources of data. First, we use self-reported measures of online media exposure ($N = 725$), and we also track individual digital news consumption activity of a sub-sample ($N = 372$) coming from our main pool of subjects. We compare the outcomes of the analysis of two sources of data in order to identify differences between observed and reported media diets and to understand the impact of digital news consumption on the study of the public agenda.

4.3.1 Sample

Participants in our study are part of an opt-in panel of a spanish market research firm, which works with us on all aspects of the sample and the implementation of the survey[2]. Recruitment was done using online contacts and offering incentives for completing structured questionnaires on their personal electronic devices (i.e. home computers, tablets, or cell phones)[3].

We targeted a sample of 1000 people[4] and the final response rate was 75%, which is a satisfactory figure and in line with the tendencies reported in the academic literature (Cook, Heath, & Thompson, 2000; Krosnick, 1999). We retained 37.2% of the subjects for our tracked sample. A total of 372 individuals thus, explicitly agree to share their anonymized browsing history for our study. The figure accords with previous research analyzing individual observed digital news consumption (Guess, 2016; Guess, 2015). All our participants answered a first questionnaire from January to February 2015. A second survey was administered from February to April 2015 and prompted them to recall past news consumption activity that we had already tracked[5].

As for the sociodemographic characteristics of our sample, their age ranges between 18 and 74 ($M = 36$, $SD = 13.73$). In total, 49% are female. Education levels vary

---

[2]The questionnaires were pretested to identify questions that respondents might have difficulty understating or interpret differently than we intended. In line with the previous survey research (Krosnick, 1999) we conducted 10 tests in December 2014 consisting of individually observing participants while they answered the online surveys. We then modified them to increase the likelihood that the meaning of each item was clear and that the participant could self-administer the survey smoothly.

[3]See in the online Appendix B, Figure B.1 to B.4, several examples of the graphic interface of our survey for desktop as well as mobile version.

[4]The recruitment process started with set of screening questions from which the initial pool of 1000 individuals was selected. Appendix C shows the full questionnaire including the screening questions to target a sample of online news users.

[5]The second survey wave took place right after we finished the first wave to increase the likelihood that people could recall their past media diets.

between primary studies and college. The largest group completed a college degree (57%), closely followed by those with secondary studies (45%). These figures match the spanish online population (Robles, Molina, & De Marco, 2012), which is precisely the target that we aim to study. Table D.1 in Appendix D compares the characteristics of our sample with that of the overall Spanish population. It shows that our sample, just as the online Spanish population, skews younger, more educated and politically interested, and that they are more left-leaning than the overall population. Despite these similarities, we must refrain from making overgeneralizations from our final sample. Foremost, although it closely resembles the Spanish online population, people who voluntarily accept being tracked are generally less concerned about privacy. Yet we can see this attitude as an advantage and assume that they will not modify their news consumption routines as a result of our study. Notably, our subjects agreed to being tracked long before we started the study, which may have also helped to mitigate any initial change in their regular behavior.

Despite acknowledging the challenge of establishing representativeness of this sample, we follow previous studies (Flaxman et al., 2016) to establish a measure of representativeness, and we compare the list of the top 20 most visited news sites by our tracked sample with that provided by *Alexa*[6] (Alexa Internet, 2014) for the Spanish online population. We obtained a strong correlation score, which equals 0.81. This figure speaks for the representativeness of the news consumption behavior of our pool of participants.

Our sample targets all individuals of the Spanish online population, except for those living in Catalonia. At the time we launched our study, one issue clearly marked the public agenda in Catalonia, in stark contrast to other Spanish regions: the prospects of becoming an independent state. In a very short period of time, people in support of independence in Catalonia almost tripled, moving from 14% in 2006 (Muñoz & Tormos, 2015) to around 41% in 2015[7] whereas the Spanish population did not perceived this issue as important. Due to this exceptional situation, we did not include this region in our sample. The decision helps us to assess the impact of the digital media on the public agenda in everyday life instead of during major political events, which tend to more easily draw people together around short-lived problems (see the discussion on Soviet Russia relations or Vietnam War in McCombs & Zhu, 1995)[8].

---

[6]Several prior studies have relied on these rankings to obtain traffic information (e.g. (Ennew, Lockett, Blackman, & Holland, 2005; Flaxman et al., 2016; Price & Grann, 2012; Wu & Ackland, 2014).

[7]We use data from the quarterly public opinion barometer of the Catalan government (http://ceo.gencat.cat) to get the last figure available.

[8]See in Appendix E a broader discussion about this decision.

4.3.2 Dependent Variable

To measure the public agenda we borrowed the standard open-ended questions for the most important problem, largely used in this research field (Maxwell McCombs & Zhu, 1995). Our basic measure of the public agenda includes four questions. The first question was worded as follows: "What do you think is the most important political or social problem in Spain?" Respondents could skip this question if they wanted to. Then we repeated the same question for the second and third most important problems. Finally, they were asked: "Could you tell us whether there are any other important political or social problems in Spain?" They were also instructed to leave these three last open-ended questions blank if they preferred it.

Two coders were trained to categorize each responder's answers using the codebook of the *Policy Agenda Project*[9]. Overall, subjects could provide up to 13 most important problems; but again, they were also prompted to leave the items blank. In total, 9,789 answers were coded, and the percentage of agreement ranged from 94.82% to 77.02%. We average the individual intercoder reliability coefficient for each answer and obtained the overall 85% agreement. Table D.2 in Appendix D shows individuals percentages of agreement per each answer.

Common Agenda

We measure the extent people share a set of common experiences by identifying the top 3 most important problems in Spain as mentioned by our sample. Then, we code in a continuum how many of these problems each individual has mentioned among the total ones in her list. Literature offers little basis for operationalizing this construct at the individual level. Yet, in assessing its validity, one has to consider that, on average, people can offer no more than 4 to 5 issues (Zhu, 1992) and there has not been an increase in people's carrying capacity across time (Maxwell McCombs & Zhu, 1995). As Table 4.1 and Table 4.2 show, our samples reported, on average, 5.9 issues (full sample) and 6 issues (tracked sample). Furthermore, most of the studies that measure the public agenda have only asked about the most important problem in an open-ended question (Chaffee & Wilson, 1977; W. R. Neuman, 1990). Hence, we assume that our threshold, set at the top 3 most important problems, offers a valid conservative measure.

---

[9]The codebook was developed following the methodology of the *Comparative Agendas Project* (www.comparativeagendas.net), which is an international network of scholars from 11 countries including the United States, Canada and the United Kingdom that analyses political, public and media agenda across time. We decided to use this codebook to allow future cross-national comparative studies.

When we launch our study, the top three most important problems according to the Spanish population (see Figure 4.1) were *corruption*, *unemployment* and *the economy*. This ranking closely resembles the list of the three most important problems as mentioned by our sample: *corruption*, *unemployment* and *politicians and politics*. The third problem mentioned by our sample is the fourth problem in the Spanish ranking. This similarity speaks again for the representativeness of our sample.

Fig. 4.1.: Most Important Problems for Spanish Population Over Time
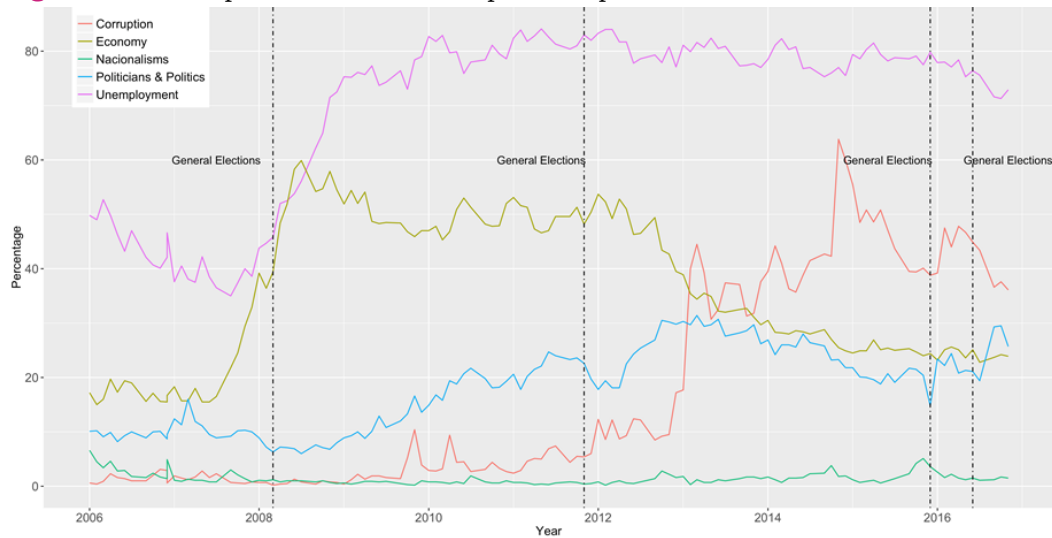
Note: Data was collected from *Centro de Investigaciones Sociológicas* which carries out the most important Spanish poll and it is funded by the Government. See on Appendix E a further explanation for the evolution of the "nationalisms" problem.

Agenda Diversity

Following seminal work on this field by McCombs & Zhu, (1995) we operationalize agenda diversity using the Shannon's H measure. It is a widely used statistic in the field of agenda diversity and communication studies (Chaffee & Wilson, 1977; Culbertson, 1992). H is known as a dual-concept measure of diversity (McDonald & Dimmick, 2003) because it accounts for the number of categories in which responses fall, the popularity of those categories among the overall sample, and the extent to which individual items are spread evenly across those categories rather than falling primarily in one or a few of them. In other words, as McCombs and Zhu point (1972), more issues and a more even distribution among existent categories both indicate greater diversity. Larger H values thus, are an indication of agenda diversity. Formally the H statistic is defined as follows:

$$H = -\sum_{i=1}^{k} A_{ij} P_i (log_2 P_i),\qquad(4.1)$$

where P_i is the percentage naming the i-th issue (i ranges from 1 to k) as the most important problem, and log_2 is the logarithm with 2 as base. We measure the H statistic for each individual in our sample. For the sake of comparison, we have standardized H to range from 0 to 1, where 1 indicates the maximum level of diversity of each individual's agenda.

Agenda Capacity

To operationalize agenda capacity in our study, we use the measure introduced by previous works (Allen & Izcaray, 1988; McCombs & Zhu, 1995; Park et al., 2013), which consist of measuring the total number of items that each respondent provides when asked for the most important problems. In contrast with agenda diversity, this is a raw measure to assess the carrying capacity of each individual, regardless of the category to which the issues belong and their distribution among categories.

4.3.3 Independent Variables

Reported News Consumption

We draw on the *European Social Survey* questions to measure frequency of media use in our study (Newton, 2001)[10]. Additionally, to prompt individuals to accurately report their news diets, we included check-all questions[11] with a list of 20 television channels, 35 newspapers brands, 11 radio stations, and 42 Internet news outlets[12]. We also included an open-ended question to add other news providers' names that were not listed previously. Respondents were allowed to check off any entry or none from randomized lists of the aforementioned news sources[13]. This approach, known

---

[10]See the questionnaire in Appendix C.

[11]This is a common format used also by the *National Annenberg Election Survey* in its 2008 and in 2012 survey for the *Evaluations of Government and the Society Study* and pre-election questionnaire (Guess, 2015).

[12]See the full lists in the questionnaire in Appendix C. Lists of offline media i.e. tv channels, newspapers and radio stations were elaborated based on the *Estudio General de Medios* which is the main survey for media audiences in Spain and it is broadly used by the advertisement industry too. As for the digital outlets list, it is based on *Alexa* ranking of top visited news sites in Spain. We compare the ranking provided by *Alexa* with that of the comScore, the official online audience meter for the Spanish digital market. The strong correlation (0.906) between media's position in both these lists proves the accuracy of our digital media list. Participants could also report sites not included in the list, through an open-ended question. Yet, the provided list already included 99.85% of all reported visits and 85% of all observed visits, which represent an important percentage of the total audience for digital news consumption.

[13]To avoid influencing respondents selection we randomize the lists of news outlets for each respondent so that they do not had news sourced ranked according to the number of visits. Lists of news media outlets were individually tailored according to the geographic location of each subject. Thus they only included local news providers of their area.

as *program list technique,* decreases the cognitive demands placed on respondents and can result in better measures of news exposure (for more on this debate see Dilliplane, Goldman, & Mutz, 2013).

Observed News Consumption

By means of a passive tracking software, we collected navigation data from our sub-sample during a period extending from February 23 to April 20, 2015, when the second survey was completed. Then, we identify how many news sites have each individual actually visited and which they are. This process serves to create the main independent variable (i.e. type of online news media diet or HHI online observed) included in the models using browsing behavior data (see Table 4.4). Our data set includes 624,811 URLs, which correspond to 56,223 unique domains, and 8.11% of which are news sites. We identify news outlets from the overall navigation dataset using the final list of news sources that our sample reported to have visited, which amount to a total of 80 outlets.

A descriptive analysis shows that the bulk of online news consumption, as revealed by the tracking data, is not represented by news media sites. This evidence is consistent with navigation data for the overall Spanish population[14]. We identify news sites at the domain-level, the typical level of analysis in studies tracking online behavior[15]. However, this criteria imposes a limitation because we cannot identify which type of news contents users finally accessed once they are inside each news outlet.

Type of News Media Diet

Our main independent variable measures the level of concentration of the reported and observed online news diet, as well as the offline media diet. Following the strategy introduced in prior work (Athey, Mobius, & Pal, 2012; Trilling & Schoenbach, 2013), we assess the level of concentration of media diets at the individual level by using the *Herfindahl-Hirschman Index* (HHI). It is a statistical measure of concentration used in a variety of contexts, especially for the analysis of markets' activity (for more on this see McDonald & Dimmick, 2003; Rhoades, 1993). Formally,

---

[14]Out of the top 500 most visited sites during the period of study, only 23 overlap with the list of 80 outlets reported by the surveyed sample. The search engine *Google* and the social platform *Facebook* lead the ranking of top most visited sites by our sample. The two first news media in our ranking are on the 13th and 14th position and they are the legacy media, *Elmundo.es* and *Elpais.com*, respectively. This order closely resembles the ranking of top visited sites for the Spanish general population provided by *Alexa* (Alexa Internet, 2015)

[15]URLs corresponding to social platforms profiles are sources of personal data. Therefore we do not have access to visits to news media outlets through social platforms like *Facebook* and *Twitter*.

the HHI index is defined to be the sum of the squared market share of each news outlet included in subjects' media diets. The index ranges from 0 to 1, and the higher it is, the more concentrated the media diet. In other words, scores close to 1 identify news diets containing one or several outlets with high shares in the news market. Conversely, scores close to 0 identify news media diets that mainly include peripheral outlets with small shares in the news market[16]. We calculate three types of individual HHI using the market share of each news outlet that participants visited or they reported to have visited: HHI online observed; HHI online reported; HHI offline reported.

Political Interest

This is a very relevant control in our study because it has been long established the relationship between media use and political interest (Althaus & Tewksbury, 2002; Holt, Shehata, Strömbäck, & Ljungberg, 2013; Newton, 2001; Strömbäck & Shehata, 2010). Those who show a tendency to be highly interested in politics are more likely to develop richer information repertoires (Wolfsfeld, Yarchi, & Samuel-Azran, 2015). They are also more likely to learn about politics (Carpini & Keeter, 1997) and to remember or to acquire political information (Bode, 2016) and, hence, to develop an expansive public agenda. Political interest is assessed by asking, "How much you would say you are interested in politics. Very much, quite interested, hardly interested, or not at all?".

Frequency of News Consumption

Our measure of types of online media diets is not independent of overall exposure levels. Hence, to accurately estimate the relation of types of media diets and different dimensions of the public agenda, we have also controlled for the effects of the frequency of news consumption both in offline and online media[17]. Furthermore, past studies show that people's reliance on different media types, e.g. online news sites or offline newspapers, might have different effects on people's perceptions of salient issues (Althaus & Tewksbury, 2002). Therefore, we have included several variables to measure frequency of news consumption in different media channels and the use of several types of devices.

---

[16]We use *Estudio General de Medios* (AIMC, 2016) and *comScore*, the official online audience meter for Spain (Interactive Adversiting Bureau, 2011) as sources to calculate the market share of the online and offline news providers.

[17]In the models using observed data, we have included a more nuanced measure of news consumption (Online News Sources Visits), based on the actual total number of visits to online news media that we have previously identified, instead of the reported number sources visited (Online News Sources Reported)

Overstated News Diet

This variable measures the difference between types of news diets reported and observed ($M=$-3.10, $SD=$0.350). We include it in our models of observed news consumption to control for the differences we find between the level of concentration of observed and reported online diets. In the observed model, this control helps to subtract the effect of those individuals misreporting their level of media concentration, from the effect that observed levels of media concentration have on the public agenda. Although controlling for this would make much sense in the reported model, we cannot compute this variable for the 735 individuals for whom we have reported data given that we only have navigation data for half of them. It is worthy to note though, that When we include this control for the subsample of individuals for which we have both survey and navigation data in the reported model, results not only hold but also improve substantially (see model in the Appendix F).

Table 4.1 and Table 4.2 contain descriptive statistics for the main variables included in the study, both for the total sample and the tracked pool of subjects. All non-dichotomous control variables have been recoded from 0 to 1 for the sake of comparability of their effects.

Tab. 4.1.: Descriptive Statistics of the Main Variables of Interest for the Survey Sample

| Statistic	N	Mean	St. Dev.	Min.	Max.
Common Agenda	725	1.900	0.790	0	3
Agenda Capacity Nominal	725	5.900	2.900	0	13
Agenda Diversity	725	0.530	0.200	0.000	1.000
Political Interest	725	0.400	0.250	0.000	1.000
Freq. Online News Consumption	725	0.140	0.270	0.000	1.000
Freq. Newspaper use for News	725	0.550	0.340	0.000	1.000
Freq. Television use for News	725	0.220	0.300	0.000	1.000
Freq. Radio use for News	725	0.500	0.400	0.000	1.000
HHI Offline Reported News Media Diet	725	0.330	0.240	0.000	1.000
HHI Online Reported News Media Diet	725	0.300	0.240	0.000	1.000
Online News Sources Reported	725	4.500	3.700	0	28

4.4 Results

We start our analysis by examining the level of concentration of the news media diets using the reported data. Figure 4.2 shows the distribution of HHI statistic for the online and offline news consumption activity of the full sample ($N=$725). A visual inspection of the left skewed distribution shown on both graphics yields evidence that people's news diets, both in the online and offline domain, vastly include outlets

Descriptive Statistics of the Main Variables of Interest for the Tracked Sub-Sample

Statistic	N	Mean	St. Dev.	Min.	Max.
Common Agenda	372	1.800	0.800	0	3
Agenda Capacity Nominal	372	6.000	2.900	0	13
Agenda Diversity	372	0.530	0.200	0.000	1.000
Political Interest	372	0.400	0.250	0.000	1.000
Freq. Online News Consumption	372	0.140	0.270	0.000	1.000
Freq. Newspaper use for News	372	0.570	0.350	0.000	1.000
Freq. Television use for News	372	0.230	0.310	0.000	1.000
Freq. Radio use for News	372	0.520	0.410	0.000	1.000
HHI Offline Reported News Media Diet	372	0.300	0.240	0.000	1.000
HHI Online Observed News Media Diet	372	0.610	0.280	0.000	1.000
HHI Overstated News Media Diet	372	-0.310	0.350	-0.940	0.850
Online News Sources Visits	372	102	129	0	694

with small audience shares in the news markets. Table 4.1 also shows the means for both measures, which equal 0.330 for the HHI offline and 0.300 for the HHI online reported. These scores, as well as the distributions in Figure 4.2, suggest, first, that people prefer getting news from websites that have relatively small number of audiences in relation to the overall online news market and, second, that they choose, although in a lesser extent, offline news providers that equally represent small shares in their corresponding markets, i.e. television audience, radio listeners or newspapers readers.

Fig. 4.2.: Distribution of Type of Reported Online and Offline News Diets for the Full Sample
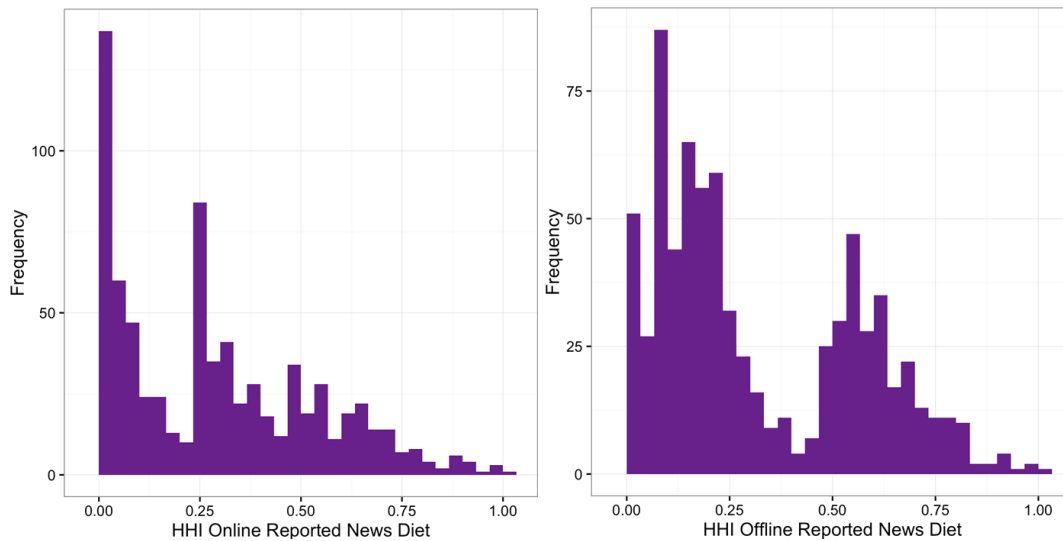
Note: HHI measure level of concentration of the media diet, the higher it is, the more concentrated is the media diet on those outlets higher shares in the news market.

We use these variables, HHI reported online and offline, and three OLS models to estimate effects of news media diets on the common agenda, the agenda diversity, and the individuals' carrying capacity. Table 4.3 shows the results. In these models,

the type of offline news media diet does not significantly affect the common agenda (see column 2). As for the online news diet, there is a positive and significant effect over the common agenda (0.290*), which suggests that concentrated digital news diets, or HHI scores close to 1, are positively associated with the extent to which people share the same public concerns. Additionally, people with digital news diets that include media sites with higher market shares, i.e. legacy outlets, are more likely to have a more diverse public agenda (0.091**) and finally, although positive, the effect over the agenda capacity is not significant (1.000).

As said, we do not find a significant relationship between the type of offline diets and the common agenda, but this variable is positively and significantly associated with the agenda diversity (0.058*) and the agenda capacity (0.950**)[18].

These models also include a variable measuring the number of news sources that people declare having visited while navigating the web. This variable does not affect either the existence of the common agenda or the agenda diversity. Yet, it has a small positive significant effect over the agenda capacity (0.080*), suggesting that the number of online sources people report to have visited increases the number of raw issues that people see as politically relevant. Overall, these results support our argument that it is necessary to assess the type of news media diet to fully understand the relationship between the online news domain and the public agenda. The models on Table 4.3 bring evidence to confirm the relations predicted in the **H4.1** and partially support **H4.2**. They suggest that people do share common ground to the extent that their online news diets are concentrated around news providers that have a major control of the news market, which happen to be legacy news media, i.e. news providers that predate the Internet era. Furthermore, they show that online news diets concentrated in mainstream media are associated with an increase of the diversity of issues that people regard as important. Yet, we do not find a significant relation with the agenda capacity.

We now aim to determine whether this pattern is consistently held when we use observed data instead of recalled media habits, as we did in the previous analyses. First, we examine again the types of news diets that are predominant in the online domain. To this end, we use browsing behavior data and measure the level of concentration of the observed news diets. Our underlying motivation is to compare reported and observed news diets and check whether it brings similar results. Figure 4.3 plots the distribution of the observed online media diets and Table 4.2 includes the mean of the HHI statistics using observed data, which equals 0.610. The comparison between the means of the HHI scores using reported ($M=0.300$) and observed data and their distribution yield clear evidence in support of the argument

---

[18]We cannot compare the magnitude across models because the dependent variables are measured in different scales.

Tab. 4.3.: OLS Estimation of the Common Agenda and the Agenda Diversity and Capacity for the Full Sample

	Agenda Diversity (1)	Common Agenda (2)	Agenda Capacity (3)
HHI Offline Reported News Diet	0.058*	-0.003	0.950**
	(0.033)	(0.130)	(0.470)
HHI Online Reported News Diet	0.091**	0.290*	1.000
	(0.043)	(0.170)	(0.620)
Online News Sources Reported	0.005	0.010	0.080*
	(0.003)	(0.012)	(0.042)
Female	0.045***	0.030	0.560***
	(0.015)	(0.059)	(0.210)
Age (25-35)	0.002	0.092	0.092
	(0.023)	(0.091)	(0.330)
Age (36-50)	0.062***	0.320***	0.460
	(0.022)	(0.086)	(0.310)
Age (= or >50)	0.092***	0.350***	1.300***
	(0.023)	(0.093)	(0.330)
Education (College)	-0.0002	-0.042	0.380*
	(0.015)	(0.061)	(0.220)
Device (Mobile phone)	0.007	0.043	-0.170
	(0.018)	(0.073)	(0.260)
Device (Tablet)	-0.019	-0.076	-0.130
	(0.027)	(0.110)	(0.390)
Political Interest	-0.023	0.310**	-1.700***
	(0.031)	(0.120)	(0.450)
Freq. Online News Consumption	0.004	-0.049	0.130
	(0.029)	(0.120)	(0.420)
Freq. Newspaper use for News	0.031	0.039	0.260
	(0.023)	(0.092)	(0.330)
Freq. Television use for News	-0.013	-0.140	0.530
	(0.026)	(0.100)	(0.370)
Freq. Radio use for News	0.010	-0.066	0.310
	(0.021)	(0.085)	(0.310)
Constant	0.380***	1.400***	4.200***
	(0.036)	(0.140)	(0.520)
Observations	725	725	725
R^2	0.084	0.062	0.099
Adjusted R^2	0.065	0.042	0.080
Residual Std. Error (df = 709)	0.190	0.770	2.800
F Statistic (df = 15; 709)	4.300***	3.100***	5.200***

Note: *$p<0.1$; **$p<0.05$; ***$p<0.01$

that news diets are more concentrated than our survey data shows. A plausible explanation for these differences is that subjects in our sample reported visiting a good amount of small and peripheral news sites, or niche sites, when they were asked about their news consumption habits that, according to our observed navigation data, they rarely visit. Why they do so is beyond the scope of our research. Over-reporting might reflect higher levels of political knowledge or interest or be owed to social desirability. Yet the nature of our data and measures does not allow us to drill down on the nuances of these relationships, and the findings should be subject for future research.

Fig. 4.3.: Distribution of Type of Reported and Observed Online News Diets
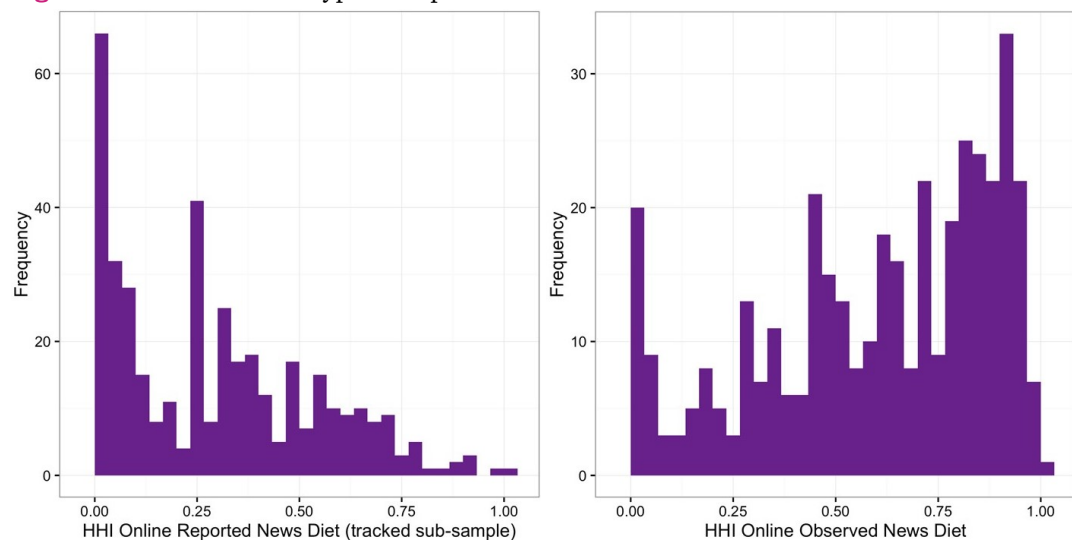
Turning to the relationship between the type of news diets observed and the extent to which people share a public agenda, Table 4.4, Column (2) reveals that the level of observed concentration in online news diets also predicts the extent to which people share common ground. The magnitude of the effect is stronger when using observed (0.640***) instead of reported data (0.290*). However, it is worth noting that when we drop from the model the variable controlling for over-reporting of media visits the effect of observed media concentration ceases to be significant. This suggests that this is not a robust effect. Moreover, the fact that this variable has a strong and significant positive relation with all of our dependent variables (agenda diversity 0.140***, common agenda 0.410**, agenda capacity 2.000***) probably calls for a more nuanced analysis of effects of the media diet on the agenda for different groups (individuals that report and are observed to have concentrated media diets, individuals that report but are not observed to have concentrated media diets, individuals that report and are observed to have fragmented media diets and individuals that report but are not observed to have fragmented media diets).

When analyzing the relation between the total amount of visits to different news sources online and the public agenda, our model predicts that news exposure online

Tab. 4.4.: OLS Estimation of the Common Agenda and the Agenda Diversity and Capacity for the Tracked Sub-Sample

	Agenda Diversity (1)	Common Agenda (2)	Agenda Capacity (3)
HHI Online Observed News Diet	0.130** (0.060)	0.640*** (0.240)	2.200*** (0.840)
HHI Offline Reported News Diet	0.087* (0.048)	-0.230 (0.190)	1.800*** (0.670)
Online News Visits	-0.0002* (0.0001)	-0.001** (0.0004)	-0.002* (0.001)
HHI Overstated News Diet	0.140*** (0.044)	0.410** (0.170)	2.000*** (0.620)
Female	0.045** (0.021)	-0.039 (0.083)	0.520* (0.290)
Age (25-34)	0.024 (0.030)	0.100 (0.120)	0.480 (0.420)
Age (36-50)	0.053* (0.030)	0.310*** (0.120)	0.560 (0.420)
Age (= or >50)	0.096*** (0.031)	0.380*** (0.120)	1.600*** (0.430)
Education (College)	0.019 (0.021)	0.012 (0.084)	0.760** (0.300)
Device (Mobile phone)	-0.028 (0.026)	-0.022 (0.100)	-0.410 (0.370)
Device (Tablet)	-0.006 (0.050)	-0.120 (0.200)	0.150 (0.690)
Political Interest	0.012 (0.044)	0.480*** (0.170)	-1.300** (0.610)
Freq. Online News Consumption	-0.023 (0.041)	-0.280* (0.160)	0.038 (0.580)
Freq. Newspaper use for News	0.010 (0.032)	-0.017 (0.130)	-0.095 (0.450)
Freq. Television use for News	-0.011 (0.036)	-0.160 (0.140)	0.800 (0.500)
Freq. Radio use for News	0.009 (0.031)	-0.150 (0.120)	0.580 (0.430)
Constant	0.410*** (0.056)	1.600*** (0.220)	3.800*** (0.790)
Observations	372	372	372
R^2	0.110	0.110	0.130
Adjusted R^2	0.065	0.070	0.092
Residual Std. Error (df = 355)	0.200	0.770	2.700
F Statistic (df = 16; 355)	2.600***	2.700***	3.300***

Note: *$p<0.1$; **$p<0.05$; ***$p<0.01$

undermines the common ground (-0.001**), the agenda diversity (-0.002*), and the carrying capacity (-0.002*). In line with the previous results, this again confirms that it is necessary to identify the type of sources that subjects visits to fully understand the extent to which online news consumption weakens the public agenda. Only by means of a more nuanced analysis that can disentangle types of news providers can we tackle the impact of the digital news domain over the public agenda.

With regards to political interest, our models using both observed and reported data consistently confirm that those more interested in politics are more likely to share the most salient issues in the public agenda (0.480***, 0.310***). However our findings also suggest that they are less likely to develop an expansive public agenda (-1.300***, -1.700***). One plausible explanation for theses results is that highly politically interested individuals are more informed yet since they are also more able to assess the news they tend to more "accurately" identifying the few problems that are more prominent, which tend to be the public agenda. Nonetheless, these findings should be further studied in future research.

To sum up, we predicted in **H4.1** that the digital media does not erode public consensus over the issues that people believe are important. Additionally our **H4.2** predicted that online news media increases the level of diversity of the public agenda. Our results seem to confirm our hypothesis although our reported models yield more robust evidence than our observed ones. The results for our observed models indicate that more analyses are needed to disentangle the effects of media diets on the agenda for different groups. While our results are more robust for our reported models (see Table 4.3 and especially the model in the Appendix F), agenda capacity seems not to be affected by the reported digital news media diet. More importantly though, the major contribution of this research is that it is necessary to identify types of news media diets instead of only considering the raw number of news sources visited when aiming to have a nuanced understanding of the impact of the online news domain on the public sphere.

Furthermore, and somehow counterintuitively, our analysis also reveals that concentrated digital news diets increase the likelihood to have a more diverse public agenda. To understand the underlying logic of this relationship, it is important to highlight that having concentrated news media diets does not imply accessing a fewer number of news sources, but that those sources visited are mostly legacy outlets. Therefore, and despite the literature reviewed above pointed to a high redundancy in the media agenda of legacy news outlets, as we expected, there might be a non-negligible diversity in the type of issues reported by them, which positively impacts the diversity of the public agenda.

4.5 Discussion

In this Chapter 4 we have examined the relationship between online news media diets and the public agenda. While many extant studies reviewed above contend that the online domain undermines the public sphere, our evidence suggests that it does not. We find that online media has not significantly changed media diets, which continue to be concentrated on outlets that have higher shares in the market of news —or on legacy media. The claim that the explosion of available sources and news outlets would erode the public sphere by fragmenting audiences has not actually been found verifiable. As this and other studies show, people continue to focus their attention on relevant and important outlets when they use the Internet to access news and navigate content online. This guarantees that the public continue to share a common ground, despite the fragmentation of outlets affecting the supply of information. If anything, by lowering the costs of accessing diverse sources and news outlets, online media seem to encourage exposure to more than one legacy media, which might account for the effect that we find between concentrated news diets and greater diversity and capacity. Hence, and in spite of theoretical predictions to the contrary, subjects agree on a set of common experiences when their online news diets are based on those outlets with higher shares in the news market. Our findings also suggest that reported news diets tend to over-state visits to peripheral or small news providers, which seems to be (clearly in our reported data) negatively associated with public agenda and agenda diversity. In this regard, we bring strong evidence that suggests that reported news diets are biased towards small peripheral news outlets that people rarely visit.

However, our study does not provide evidence for a causal link between the type of news diets and the extent to which people share a public agenda. Our measures are based on non-temporal data, and questions remain open about whether the relationships this study has unveiled are causal. Future studies using longitudinal observed data or based on an experimental design could contribute to identifying the mechanism through which news media diets affect the public agenda and the pathways among the variables. We do not control for informal ways of acquiring political information, through family of friendship ties, which are also potential explanatory variables for some of the relations that our study unveils.

With respect to other limitations, one might argue that our findings may be partially explained by the administration of the questionnaire leading subjects to state a higher number of most important problems. Yet, as it has already been stated before, subjects were always informed of the possibility to leave an item blank, which leads us to contend that our constructs for the public agenda are accurate measures of the dimensions studied here. There is another technical issue that deserves discussion:

as stated, this is a domain-level study, and we do not include subdomains in the browsing tracking analysis. Hence, we do not identify the type of content that people access when navigating the web. Future research should also analyze subdomains to filter out subjects that do not navigate political information at all. Finally, although some limitations exist within the representativeness of our sample, our findings make an important contribution to understand the behavior of active online news users. Ultimately, they allow us to identify predominant types of news diets and their impact on the public agenda in the digital age.

Conclusion

5

The main goal of this study has been, first, to identify whether there is fragmentation in the online news domain and, secondly, in case it exists, to measure whether the fragmentation is eroding the public agenda and, by extension, democratic life. Our findings provide qualified support for debunking the theses that equates the digital news domain with public fragmentation. We have contributed to this debate with novel measures and data (from Spain) and by tackling fragmentation at three different levels. In each measurement level, we have found consistent evidence that digital technologies have not weakened -–i.e. fragmented-– the public domain to the extent that some expected (Chaffee & Metzger, 2001; Papacharissi, 2002; Pariser, 2011; Sunstein, 2009; Tewksbury, 2005). As a result, we have also found that there is still common ground to anchor informed citizens; and, thus, public space for political discussion seems to be still guaranteed.

In our study, fragmentation has been related, at the macro-level, to the structural characteristics of the network of news providers and the audience behavior and at the micro-level, to the existence of shared public concerns —i.e. public agenda. While previous studies have made inferences about the transformation of the public space and audience behavior mainly by studying the organization and content of digital news providers, our work has improved on such research by: first, connecting the two dimensions of the news domain —i.e. the supply and the demand side-– and measuring their structural similarities simultaneously; and, secondly, by relating observed types of news diets to actual perceptions about the public agenda.

At the macro-level (Chapter 3), we have shed light on the limited ability of news providers to drive audience behavior using hyperlink strategies. We have also brought novel evidence to better understand how people navigated content online. Overall, we have found high levels of audience overlapping among news providers. In other words, according to our results, there is no evidence of fragmentation in the demand side of the news domain. This structural feature suggests that people widely navigate online news content. The opposite is true for the news network, that is, our results suggest that there are high levels of fragmentation in the provision of news content.

We have operationalized fragmentation in the online domain by means of network community analysis. This method has allowed us to understand the structural organization of news consumption and has revealed that when navigating content online audience responds to a greater extent to individual motivations than to the structure of news providers. A supplementary analysis, i.e. ERGM analysis, has unveiled stronger connections among new media and legacy media, suggesting that audience overlapping is higher within the same type of media sites. Future research should delve into this finding. Although, our data does not allow us to disentangle the causal mechanisms explaining these (audience) patterns, a plausible explanation might be that distinctive generational motivations for news consumption play a role in shaping news diets. Along the same lines, the community analysis has also suggested that the ideological dimension might be an important variable for understanding the organization of news consumption. Yet again, we encourage future research to address this relationship with novel data that include audience ideological leaning.

With regard to the supply dimension, and in line with previous theoretical accounts, the community analysis' results have suggested that news providers are organized around segregated groups defined by economic and geographical criteria. The former criteria, which is the main driver of this network, is based on the business groups that each news media belongs to. More precisely, our results have shown that news media outlets are by far more strongly connected with sites from the same parent corporation. The geographical criteria might be also paying a role, although to a lesser extent than the previous one, in organizing the network of news providers. Languages as well as regional proximity seem to strengthen the relationships between media outlets, which causes the emergence of few additional isolated clusters in the supply side of the news domain.

We hypothesized though, that the structure of the audience and the news networks would be significantly correlated if the audience responded to the connections created by media outlets. By applying correlation techniques for network structures, our analyses have revealed that news consumption patterns are only slightly correlated to the networks of news providers. We interpret this result as evidence of a lack of fragmentation in the news domain because although the provision of news is highly fragmented (Q=0.68), with clear fault lines associated to media ownership structures, the audience does not mirror this structure (Q close to zero).

From a meso-level, in Chapter 4, we have delved into the relation between news consumption patterns and news media connections to understand how they shape power relations in the online news domain. The main goal of this analysis has been to determine the extent to which the legacy media have lost their monopoly as elites within the political communication process. To this end, we have proposed

an empirical framework that discerns the structural mechanisms that confer power to media outlets by assessing the level of endorsement that news content receives (authorities) and their control over the audience flow (brokers). We draw one main conclusion from the examination, at this level, of the structural position of different types of news providers both in the audience network and the hyperlink network: legacy news media are still in control of the most powerful positions in the digital news ecology. They are the most authoritative sources of information, as revealed by the patterns of news citations or hyperlinks that news media send to each other. Further, they also occupy those positions that grant a higher level of control of the audience flow. In line with the results at the macro-level analysis, we have brought novel evidence to prove that young generations, which have distinctive patterns of news navigation, also confer power to new digital outlets. These findings have clear implications for reassessing theoretical and empirical accounts about the emergence of new elites within the online news domain (Castells, 2009; Groshek & Tandoc, 2017) and the extent to which legacy media still play a fundamental role as agenda setters and gatekeepers.

It is important to highlight that the empirical framework that we have applied at the macro- and the meso- levels offers an advantage with respect to previous works that have analyzed the structural characteristics of online audience behavior (Barnett & Park, 2005; Taneja, 2016; Taneja & Webster, 2016; Webster & Ksiazek, 2012; Wu & Ackland, 2014). While previous network analysis assessing audience behavior and the structure of the web have identified significant connections by disregarding edge strength -–which in our case represents the amount of audience overlapping or the number of citation amongst outlets– we have followed previous research in network science (Ronen et al., 2014) and have retained this information in all of our analyses. We argue that this methodological approach provides a more robust representation of the online news consumption patterns and, ultimately, enhances the reliability of our analyses. Moreover, the network techniques that we have applied in Chapters 3 and 4 go beyond a descriptive effort. As previous methodological research has shown, they are advanced analyses in that they operate on the basis of null models and help to determine departure from randomness (Borge-Holthoefer & González-Bailón, 2015).

Finally, in Chapter 4, we have aimed to analyze the fragmentation of the online news domain at the individual or micro-level. To this end we have used both reported and observed -from browsing tracking systems- measures of online news exposure to identify the level of concentration of digital news media diets. We have found that online media does not erode the common ground binding (democratic) polities together, mainly because news consumption continues to be concentrated in a few mainstream news outlets. Perhaps more surprisingly, we have found that concentrated online news media diets are associated with more diverse agendas and

a greater capacity to account for public problems. Hence, our results at the micro-level seem to confirm our hypothesis, that digital media do not cause audiences to be more fragmented and do not, as a result, erode the public space. Our results for the main finding that media concentration helps bind communities together and increases agenda diversity and capacity, however, are more robust for reported than for observed measures of news consumption. This fact clearly calls for a more nuanced analysis of the effects of the media diet on the public agenda by taking into account effects for different groups. Our analyses suggests that levels of consistency between reported and observed exposure should be taken into account and inform future analysis of the effects of online news diets in the public agenda.

Apart from this caution, other caveats are in order too. First, although we have not found evidence for audience fragmentation, there is still some room for audiences to be organized according to their ideological biases. We have neither analyzed news content nor counted with information about audience' political stances. Therefore, we have not been able to discern the extent to which political leaning might be a driver of online audience. This is something that future research should tackle.

Secondly, the emergence of new regulatory attempts to impose taxes on news-linking activity in Europe (European Commission, 2016) makes even more important to reassess, in the short-term, levels of fragmentation in the online news domain using longitudinal data and considering the differences in the news structure and audience behavior before and after these interventions. In light of our results, if the European Parliament follows the same policy course than Spain, one might expect that it can fragment the provision of news information as well as reduce the number of new media that aggregate to some extend news information from several sources.

Finally, our network approach has had the advantage of making explicit an online structure that remained hidden in previous works. Yet it does not replace the examination of individual characteristics to explain social phenomena, rather it supplements it. It is an additional resource that has allowed us here —in Chapter 2 and 3– to better understand the relationship between news consumption and the public domain and provide evidence to reassess claims of public fragmentation at different levels.

Beyond all of these limitations, this thesis makes an important contribution to inform a future cross-country and more ambitious study of the digital news domain. The empirical framework designed and applied to the Spanish case can be used regardless of the national context. Yet, Spain is just an observation within a larger interrelated media landscape and our results might not hold when assessing new cases.

There are several reasons that justify a future study to further tackle levels of fragmentation in the online news domain by analyzing a multinational scenario. On one hand, the role of digital-born news media in Europe varies across country. Spain has provided a good opportunity to study a news market where new media outlets are prominent. Here, digital-born organization flourished after the struggles that legacy media suffered during the economical crisis, between 2008 and 2015. This scenario might explain the predominant positions, as brokers of the overall audience flow, that these types of news providers occupy amongst the youngest public. However, in Germany and the UK legacy media remain strong and they have a longer tradition of publicly-funded journalism (Nicholls, Shabbir, & Nielsen, 2016). These conditions might make it more difficult for new media to challenge the power of legacy organizations even amongst the youngest public. This hypothetical scenario would significantly change the results we have found and it deserve a further analysis. Beyond the reallocation of media power though, our results suggest that different types of news media diets are associated with levels of common ground for public discussion. In this regards, France provides an interesting case to test whether our results hold in a country where digital-born media are very prominent. If news media diet were less concentrated in this country, fears of public fragmentation could become a reality.

On the other hand, we have only analyzed two dimensions of the online news domain. Future studies should consider a more comprehensive operationalization of the online news domain. In other words, we have simultaneously analyzed the provision and consumption of information and we have not found patterns of fragmentation in the public domain. Yet, a more ambitious study should include the analysis of the several other layers where news media operate in the digital sphere, i.e. Twitter and Facebook. These social platforms have an increasing role in the distribution of news information. Previous research suggests that nodes i.e. media outlets that are prominent in one information domain might not be prominent in another. More precisely, evidence from network science (De Domenico, Solé-Ribalta, Omodei, Gómez & Arenas, 2013) has shown that analyzing the several layers where nodes are located can alter their centrality. A multidimensional approach could offer a more comprehensive analysis of the overall role that different types of news media have. Hence a natural step forward to advance our knowledge of the digital news domain implies a multilayer structure study. This approach can offer a more nuanced assessment of the overall performance of new and legacy media in the online news domain.

Moreover, the power of news media has been analyzed here by looking into the structure of interdependence that users and producers of information create. We have mapped the position of news media along network paths that allow media organizations to be more efficient in the diffusion of their content and generate

higher visibility for their brands. Yet, the nature of our data, has not allowed us to assess the extent to which these positions impact in their ability to engage wider audiences and increase the reach of their journalism. By using dynamic data from social platforms of news distribution i.e. Twitter and Facebook future studies can drill down on the consequences of these new actors in increasing or limiting the influences of legacy and new media.

In concluding, it is worthy to mention that all of these lines of future research already belong to a new study program. It is based on the evidences presented here and tackles the levels of fragmentation across countries in Europe and the role of news providers from a multilayered perspective including social media platforms.

Contributions

The work included in this thesis has resulted in the following papers:

- Majó-Vázquez, S., Cardenal, A. S., & González-Bailón, S. (2015). *Online media networks and audience flow: Mapping the fragmentation in news production and consumption on the web.* Manuscript submitted for publication.

- Majo-Vázquez, S., Cardenal, A. S., Sagarra, O., & Colomer, P. (2016). *Challenging power in the flow of digital news: Emergent authorities and audience brokers.* Manuscript submitted for publication.

- Majó-Vázquez, S., Cardenal, A. S., Galais, C. (2017). *Lack of common ground? Capacity and diversity of the public agenda in the digital age.* Manuscript submitted for publication.

- Majo-Vázquez, S., Cardenal, A. S., Sagarra, O., & Colomer, P. (2016). *Audience brokers and news discoverers: The role of new media in the digital domain.* Building a European Digital Space Proceedings of the 12th International Conference on Internet, Law & Politics. Universitat Oberta de Catalunya, Barcelona, July, 2016. pp. 607-632.

- Majó-Vázquez, S. (2015). *A network analysis of online audience behaviour: Towards a better comprehension of the agenda setting process*. IDP. Revista de Internet, Derecho y Política. No. 20, pp. 61-74. UOC <http://journals.uoc.edu/index.php/idp/article/view/n20-majo/n20-majo-pdf-en> http://dx.doi.org/10.7238/idp.v0i20.2599

Previous drafts of the chapters of this thesis were presented at the following international and national conferences:

- Majo-Vázquez, S., Cardenal, A. S., Sagarra, O., & Colomer, P. (2016). *Audience brokers and news discoverers: The role of new media in the digital news domain.* Presented at the 9th CIS-Harvard Summer Seminar on Sociological and Political Research, Harvard University, Cambridge, MA.

- Majó-Vázquez, S., Cardenal, A. S., & González-Bailón, S. (2016). *Online media networks and audience flow: Mapping the fragmentation in news production and consumption on the web.* Presented at the XXIV World Congress of Political Science, Poznań, Poland.

- Majo-Vázquez, S., Cardenal, A. S., Sagarra, O., & Colomer, P. (2016). *Challenging power in the flow of digital news: Emergent authorities, news media and audience brokers.* Presented at the Social Media & Society Conference, London, UK.

- Majo-Vázquez, S., Cardenal, A. S., Sagarra, O., & Colomer, P. (2016). *Challenging power in the flow of digital news: Emergent authorities, news media and audience brokers.* Presented at Internet, Law and Politics Congress UOC, Barcelona, Spain.

- Majó-Vázquez, S., Cardenal, A. S., & González-Bailón, S. (2016). *Online media networks and audience flow: Mapping the fragmentation in news production and consumption on the web.* Presented at ICA, Fukuoka, Japan.

- Majó-Vázquez, S., Cardenal, A. S., & González-Bailón, S. (2015). *Online media networks and audience flow: Mapping the fragmentation in news production and consumption on the web.* Presented at MZES Big Data Conference, Mannheim, Germany.

- Majó-Vázquez, S., Cardenal, A. S., & González-Bailón, S. (2015). *Online media networks and audience flow: Mapping the fragmentation in news production and consumption on the web.* Presented at the XII Congreso AECPA, San Sebastián, Spain.

- Majó-Vázquez, S., Cardenal, A. S. (2015). *A network analysis of the online audience behavior: Towards a better comprehension of the public agenda.* Presented at the 65th Annual ICA Conference, San Juan, Puerto Rico.

Bibliography 7

Abbar, S., Zanouda, T., & Borge-Holthoefer, J. (2016). Robustness and Resilience of cities around the world. *arXiv Preprint arXiv:1608.01709*.

Ackland, R. (2005). Voson. Retrieved from http://uberlink.com

Ackland, R. (2010). WWW Hyperlink Networks. In B. S. and M. S. D. Hansen (Ed.), Analyzing Social Media Networks with NodeXL: Insights from a connected world. Morgan-Kaufmann.

Ackland, R., & Gibson, R. (2004). Mapping Political Party Networks on the WWW.

Adamic, L. A., & Glance, N. (2005). The Political Blogosphere and the 2004 U.S. Election: Divided They Blog, 36–43.

AIMC. (2014). *Estudio General de Medios*, resumen abril 2013 a marzo 2014. Retrieved from http://www.aimc.es/

AIMC. (2016). *Estudio General de Medios*. Retrieved May 1, 2015, from http://www.aimc.es/-Datos-EGM-Resumen-General-.html

Albert, R., Jeong, H., & Barabási, A.-L. (2000). Error and attack tolerance of complex networks. Nature, 406(6794), 378–382.

Aldred, J. (2008). The world's 50 most powerful blogs. Retrieved February 2, 2016, from https://www.theguardian.com/technology/2008/mar/09/blogs

Alexa Internet, I. (2014). About the Alexa Traffic Rankings. Retrieved July 14, 2014, from www.alexa.com/help/traffic-learn-more

Alexa Internet, I. (2015). Alexa. Top Sites in Spain. Retrieved January 20, 2015, from http://www.alexa.com/topsites/countries/ES

Allen, R. L., & Izcaray, F. (1988). Nominal agenda diversity in a media-rich, less-developed society. *Communication Research*, 15(1), 29–50.

Althaus, S. L., & Tewksbury, D. (2000). Patterns of Internet and traditional news media use in a networked community. *Political Communication*, 17(1), 21–45.

Althaus, S. L., & Tewksbury, D. (2002). Agenda Setting and the "New" News: Patterns of Issue Importance Among Readers of the Paper and Online Versions of the New York Times. *Communication Research*, 29(2), 180–207. http://doi.org/10.1177/00936502 02029002004

Anderson, C., Bell, E., & Shirky, C. (2012). Post-industrial Journalism: Adapting to the Present: a Report. *Columbia Journalism School.*

APM. (2013). *Informe Anual de la Profesión Periodística 2013*. Madrid.

APM. (2015). *Informe Anual de la Profesión Periodística 2015*. Madrid.

Athey, S., Mobius, M., & Pal, J. (2012). The impact of news aggregators on internet news consumption: The case of localization.

Barabasi, A.-L. (2002). Linked: How everything is connected to everything else and what it means. Plume Editors.

Barberá, P. (2014). How Social Media Reduces Mass Political Polarization. Evidence from Germany, Spain, and the US. Job Market Paper, New York University.

Barnett, G. A., & Park, H. W. (2005). The structure of international Internet hyperlinks and bilateral bandwidth. *In Annales des telecommunications* (Vol. 60, pp. 1110–1127).

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

BBC. (2016). Have your say. Retrieved from http://www.bbc.com/news/have_your _say

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.

Bennett, W. L. (2003). New media power. *Contesting Media Power*, 17–37.

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

Bode, L. (2016). Political news in the news feed: Learning politics from social media. *Mass Communication and Society*, 19(1), 24–48.

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 http://ssrn.com/abstract=2686703

Borodin, A., Roberts, G. O., Rosenthal, J. S., & Tsaparas, P. (2005). Link analysis ranking: algorithms, theory, and experiments. ACM Transactions on Internet Technology (TOIT), 5(1), 231–297.

Boynton, G. R., & Glenn W Richardson, J. (2016). Agenda setting in the twenty-first century. *New Media & Society*, 18(9), 1916–1934. http://doi.org/10.1177/1461444815616226

Brogi, E., Ginsborg, L., Ostling, A., Parcu, P. L., & Simunjak, M. (2015). Monitoring Media Pluralism in Europe – Testing and Implementation of the Media Pluralism Monitor 2015. San Domenico di Fiesole (FI).

Burt, R. S. (2000). The network structure of social capital. *Research in Organizational Behavior*, 22, 345–423.

Cagé, J. (2016). Saving the Media: Capitalism, Crowdfunding, and Democracy. Harvard University Press.

Cage, J., Viaud, M.-L., & Herve, N. (2015). The Production of Information in an Online World. Available at SSRN.

Cańo, A. (2016, March 3). *Carta abierta del director de EL PAÍS a la Redacción del periódico*. El País. Retrieved from http://elpais.com/elpais/2016/03/03/actualidad/1457031570_991358.html

Cantijoch, M., Cutts, D., & Gibson, R. (2012). Internet use and political engagement: The role of e-campaigning as a pathway to online political participation. *International Political Science Association XXII World Congress* (p. 31). Madrid.

Carpini, M. X. D., & Keeter, S. (1997). What Americans know about politics and why it matters. New Haven and London: Yale University Press.

Casanueva, M. (2007). *El diario "Público" nace sin editoriales*. Retrieved August 15, 2015, from http://www.elmundo.es/elmundo/2007/09/20/comunicacion/1190308 120.html

Castells, M. (2007). Communication, power and counter-power in the network society. *International Journal of Communication*, 1(1), 29.

Castells, M. (2009). Communication Power (first). New York: Oxford University Press, Inc.

Ceron, A., Curini, L., & Iacus, S. M. (2016). First-and second-level agenda setting in the Twittersphere: An application to the Italian political debate. *Journal of Information Technology & Politics*, 13(2), 159–174.

Chadwick, A. (2013). The hybrid media system: Politics and power. Oxford University Press.

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

Chaffee, S. H., & Wilson, D. G. (1977). Media rich, media poor: Two studies of diversity in agenda-holding. *Journalism and Mass Communication Quarterly*, 54(3), 466.

Chung, C. J., Nam, Y., & Stefanone, M. A. (2012). Exploring online news credibility: The relative influence of traditional and technological factors. *Journal of Computer-Mediated Communication*, 17(2), 171–186.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, S95–S120.

Coleman, R., & McCombs, M. (2007). The young and agenda-less? Exploring age-related differences in agenda setting on the youngest generation, baby boomers, and the civic generation. *Journalism & Mass Communication Quarterly*, 84(3), 495–508.

Coleman, R., McCombs, M., Shaw, D., & Weaver, D. (2009). Agenda setting. *The Handbook of Journalism Studies*, 147–160.

Colino, A. (2013). Evolución de la financiación de los medios de comunicación en España. ¿Hacia un modelo más sostenible? Papeles de Europa, 26(1).

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

Cook, C., Heath, F., & Thompson, R. L. (2000). A Meta-Analysis of Response Rates in Web- or Internet-Based Surveys. *Educational and Psychological Measurement*, 60(6), 821–836. http://doi.org/10.1177/00131640021970934

Cook, T. E. (1998). Governing with the news: The news media as a political institution. University of Chicago Press.

Couldry, N., & Curran, J. (2003). Contesting media power: Alternative media in a networked world. Rowman & Littlefield Publishers.

Croteau, D. (2006). The growth of self-produced media content and the challenge to media studies. *Critical Studies in Media Communication*, 23(4), 340–344.

Ctxt.es. (2016). #CTXT. El contexto lo es todo. Retrieved March 15, 2016, from https://en.
goteo.org/project/ctxt

Culbertson, H. M. (1992). Measuring agenda diversity in an elastic medium: Candidate position papers. *Journalism Quarterly*, 69(4), 938–946.

Curran, J., Iyengar, S., Lund, A. B., & Salovaara-Moring, I. (2009). Media system, public knowledge and democracy A Comparative Study. *European Journal of Communication*, 24(1), 5–26.

Davis, A. (2009). Politics, Journalism and New Media: Virtual Iron Cages in the New Culture. New Media, Old News: Journalism and Democracy in the Digital Age, 121.

De Domenico, M., Solé-Ribalta, A., Omodei, E., Gómez, S., & Arenas, A. (2015). Ranking in interconnected multilayer networks reveals versatile nodes. *Nature communications*, 6.

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

Dellarocas, C., Katona, Z., & Rand, W. (2013). Media, aggregators, and the link economy: Strategic hyperlink formation in content networks. *Management Science*, 59(10), 2360–2379.

Dilliplane, S., Goldman, S. K., & Mutz, D. C. (2013). Televised exposure to politics: New measures for a fragmented media environment. *American Journal of Political Science*, 57(1), 236–248.

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.

Downie, L., & Schudson, M. (2009). The reconstruction of American journalism. *Columbia Journalism Review*, 19, 2009.

Dutton, W. H., Gillett, S. E., McKnight, L. W., & Peltu, M. (2004). Bridging broadband Internet divides: reconfiguring access to enhance communicative power. *Journal of Information Technology*, 19(1), 28–38.

Edelman. (2016). 2016 Edelman Trust Barometer.

Elconfidencial. (2015, February 16). Nacho Cardero: *"El Confidencial aspira a ser el mejor diario de este país."* Elconfidencial.com. Retrieved from http://www.elconfidencial.com/comunicacion/2015-02-16/encuentro-digital-con-nacho-cardero-director-de-el-confidencial_712964/

Eldiario.es. (2015, October 14). *Nace El Español, el nuevo medio de Pedro J. Ramírez*. Eldiario.es. Madrid. Retrieved from http://www.eldiario.es/politica/Nace-nuevo-medio-Pedro-Ramirez_0_441306037.html

Elespanol.com. (2015, February 28). *EL ESPAÑOL finaliza su campaña de "crowdfunding."* Elespanol.com. Retrieved from https://www.youtube.com/watch?v=fUCtsZSJgBU&feature=youtu.be

Ennew, C., Lockett, A., Blackman, I., & Holland, C. P. (2005). Competition in Internet retail markets: The impact of links on web site traffic. *Long Range Planning*, 38(4), 359–372.

European Commission. (2016). *Directive of the European Parliament and of the Council on Copyright in the Digital Single Market*. Brussels. Retrieved from https://ec.europa.eu/digital-single-market/en/news/proposal-directive-european-parliament-and-council-copyright-digital-single-market

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.

Fenton, N. (2010). New media, old news: Journalism and democracy in the digital age. Sage Publications.

Fernández-Santos, E. (2012). *Roures sobre el cierre del diario Público*. Retrieved August 16, 2015, from http://sociedad.elpais.com/sociedad/2012/02/25/actualidad/13 30149573_222114.html

Fernandez, R. M., & Gould, R. V. (1994). A dilemma of state power: Brokerage and influence in the national health policy domain. *American Journal of Sociology*, 1455–1491.

Flaxman, S., Goel, S., & Rao, J. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, 80(Special Issue), 298–320.

Flaxman, S., Goel, S., & Rao, J. M. (2013). Ideological segregation and the effects of social media on news consumption. Available at SSRN.

Fortunati, L., Sarrica, M., O'Sullivan, J., Balcytiene, A., Harro-Loit, H., Macgregor, P., Roussou, N., Slaverría, R., De Luca, F. (2009). The influence of the Internet on European journalism. *Journal of Computer-Mediated Communication*, 14(4), 928–963.

Freeman, L. C. (1979). Centrality in social networks conceptual clarification. Social Networks, 1(3), 215–239. Froman, M. B. G. (2016). The 2016 National Trade Estimate Report on Foreign Trade Barriers. Washington D.C. Retrieved from https://ustr.gov/sites/default/files /2016-NTE-Report-FINAL.pdf

Garnham, N. (1995). The media and the public sphere. *The Information Society Reader*, 357–365.

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

Garrett, R. K., & Resnick, P. (2011). Resisting political fragmentation on the Internet. Daedalus, 140(4), 108–120. Gentzkow, M., & Shapiro, J. M. (2011). Ideological Segregation Online and Offline. *The Quarterly Journal of Economics*, 126(4), 1799–1839. http://doi.org/10.1093/qje/qjr044

Girvan, M., & Newman, M. E. J. (2002). Communitiy structure in social and biological networks. *Proceedings of the National Academy of Science*, 99(12), 7821–7826. http://doi.org/
10.1073/pnas.122653799

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

González-Bailón, S. (2009). Opening the black box of link formation: Social factors underlying the structure of the web. *Social Networks*, 31(4), 271–280.

González-Bailón, S., & Wang, N. (2013). The Bridges and Brokers of Global Campaigns in the context of Social Media. *SSRN Work. Pap.*

Google. (2013). Google creates €60m Digital Publishing Innovation Fund to support transformative French digital publishing initiatives. Retrieved February 1, 2014, from http://googleblog.blogspot.com.es/2013/02/google-creates-60m-digital-publishing.html

Google. (2016). European copyright: there's a better way. Retrieved January 1, 2016, from https://blog.google/topics/public-policy/european-copyright-theres-better-way/

Google Digital News Initiative. (2015). Digital News Initiative. Retrieved January 20, 2017, from https://www.digitalnewsinitiative.com/faq/

Guess, A. M. (2015). Measure for Measure: An Experimental Test of Online Political Media Exposure. *Political Analysis*, 23(1).

Guess, A. M. (2016). Media Choice and Moderation: Evidence from Online Tracking Data.

Gurevitch, M., Coleman, S., & Blumler, J. G. (2009). Political communication—Old and new media relationships. *The ANNALS of the American Academy of Political and Social Science*, 625(1), 164–181.

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

Habermas, J. (1994). Three Normative Models of Democracy. Costellations, 1(1), 1–10. http://doi.org/doi: 10.1111/j.1467-8675.1994.tb00001.x

Hallin, D. C., & Mancini, P. (2004). Comparing media systems: Three models of media and politics. Cambridge University Press.

Helft, M., & Lohr, S. (2006, November 20). 176 Newspapers to Form a Partnership With Yahoo. The New York Times. Retrieved from http://www.nytimes.com/2006/11/20/technology/20yahoo.html

Hermida, A., & Thurman, N. (2008). A clash of cultures: The integration of user-generated content within professional journalistic frameworks at British newspaper websites. *Journalism Practice*, 2(3), 343–356.

Hindman, M., Tsioutsiouliklis, K., & Johnson, J. A. (2003). "Googlearchy": How a few heavily-linked sites dominate politics on the web. Princeton University Press.

Holt, K., Shehata, A., Strömbäck, J., & Ljungberg, E. (2013). Age and the effects of news media attention and social media use on political interest and participation: Do social media function as leveller? *European Journal of Communication*, 28(1), 19–34.

ICIJ. (2016). Panama Papers. Retrieved March 4, 2016, from https://panamapapers.icij.org/

Independent. (2016, February 12). The Independent becomes the first national newspaper to embrace a global, digital-only future. Independent. London, UK. Retrieved from http://www.independent.co.uk/news/media/press/the-independent-becomes-the-first-national-newspaper-to-embrace-a-global-digital-only-future-a6869736.html

Interactive Adversiting Bureau. (2011). Comscore, seleccionado para la medición de la audiencia online en España. Retrieved August 18, 2015, from http://www.iabspain.net/noticias/sector-digital/comscore-seleccionado-para-la-medicion-de-la-audiencia-online-en-espana/

Jarvis, J. (2016). STOP THE PRESSES! The Independent is Dead. Long Live the Independent. Retrieved January 1, 2016, from http://buzzmachine.com/

Kahneman, D. (2011). Thinking, fast and slow (2013th ed.). New York: Farrar, Straus and Giroux.

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 http://www.jstor.org/stable/1048167

Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM* (JACM), 46(5), 604–632.

Kleinberg, J. M., Kumar, R., Raghavan, P., Rajagopalan, S., & Tomkins, A. S. (1999). The web as a graph: measurements, models, and methods. In *Computing and Combinatorics* (pp. 1–17). Springer.

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.

Koehly, L. M., Goodreau, S. M., & Morris, M. (2004). Exponential Family Models for sampled and census Network data. *Sociological Methodology*, 34(1), 241–270. http://doi.org/10.1111/j.0081-1750.2004.00153.x

Krackhardt, D. (1987). QAP partialling as a test of spuriousness. *Social Networks*, (9), 171–186.

Krackhardt, D. (1988). Predicting with networks: nonparametric multiple regression analysis of dyadic data. *Social Networks*, (10), 359–381.

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, 50(1), 537–567.

Ksiazek, T. B. (2011). A Network Analytic Approach to Understanding Cross-Platform Audience Behavior. *Journal of Media Economics*, 24(4), 237–251. http://doi.org/10.1080/08997764.2011.626985

Larsson, A. O. (2013). Staying in or going out? Assessing the linking practices of Swedish online newspapers. *Journalism Practice*, 7(6), 738–754.

Lazarsfeld, P. F., Berelson, B., & Gaudet, H. (1944). The people's choice: How the voter makes up his mind in a presidential election. New York: Duell, Sloan and Pearce. New York and London: Columbia University Press.

Lee, J. K. (2007). The effect of the Internet on homogeneity of the media agenda: A test of the fragmentation thesis. *Journalism & Mass Communication Quarterly*, 84(4), 745–760.

Lee, J. K. (2009). Incidental Exposure to News: Limiting Fragmentation in the New Media Environment. The University of Texas at Austion.

Lee, N. Y. (2016). Two Different Motivations on Agenda Setting: Need for Orientation and Motivated Reasoning. *International Journal of Public Opinion Research*, 28(4), 484–510.

Leibovich, M. (2016, October 15). Inside her race to the finish. New York Times, p. 1. New York.

Lewis, S. C. (2012). From journalism to information: The transformation of the Knight Foundation and news innovation. *Mass Communication and Society*, 15(3), 309–334.

Lippmann, W. (1922). Public opinion. New York: Macmillan.

Livingstone, S. (1999). New media, new audiences? *New Media and Society*, 1(1), 59–66.

Local Media Consortium. (2016). Local Media Consortium. Retrieved February 2, 2017, from http://www.localmediaconsortium.com/

Lusher, D., Koskinen, J., & Robins, G. (2012). Exponential random graph models for social networks: Theory, methods, and applications. Cambridge University Press.

Majó-Vázquez, S., Cardenal, A. S., & González-Bailón, S. (2015). Online Media Networks and Audience Flow: Mapping the Fragmentation in News Production and Consumption on the Web.

Majo-Vázquez, S., Cardenal, A. S., Sagarra, O., & Colomer, P. (2016). Audience Brokers and News Discoverers: The Role of New Media in the Digital Domain. *Proceedings of the 12th International Conference on Internet*, 607–632.

Mansell, R. (2004). Political economy, power and new media. *New Media and Society*, 6(1), 74–83.

Maspublico.org. (2012). MasPublico.org. Retrieved April 10, 2016, from http://www.verkami.com/projects/1736-maspublicoorg

Matthes, J. (2006). The need for orientation towards news media: Revising and validating a classic concept. *International Journal of Public Opinion Research*, 18(4), 422–444.

McChesney, R. W., & Pickard, V. (2011). Will the last reporter please turn out the lights: The collapse of journalism and what can be done to fix it. New Press, The.

McCombs, M. (2005). A look at agenda-setting: Past, present and future. *Journalism Studies*, 6(4), 543–557.

McCombs, M. E., & Shaw, D. L. (1972). The Agenda-Setting Function of Mass Media. *Public Opinion Quarterly*, 36(2), 176–187. Retrieved from http://0-www.jstor.org.cata leg.uoc.edu/stable/2747787

McCombs, M., & Shaw, D. (1972). The agenda-setting function of mass media. *Public Opinion Quarterly*, (36), 176–187.

McCombs, M., & Zhu, J.-H. (1995). Capacity, diversity, and volatility of the public agenda trends from 1954 to 1994. *Public Opinion Quarterly*, 59(4), 495–525.

McDonald, D. G., & Dimmick, J. (2003). The conceptualization and measurement of diversity. *Communication Research*, 30(1), 60–79.

Meraz, S. (2009a). Is there an elite hold? Traditional media to social media agenda setting influence in blog networks. *Journal of Computer-Mediated Communication*, 14(3), 682–707.

Meraz, S. (2009b). Is There an Elite Hold? Traditional Media to Social Media Agenda Setting Influence in Blog Networks. *Journal of Computer-Mediated Communication*, 14(3), 682–707. http://doi.org/10.1111/j.1083-6101.2009.01458.x

Meraz, S., & Papacharissi, Z. (2013). Networked gatekeeping and networked framing on Egypt. *The International Journal of Press/Politics*, 18(2), 138–166.

Messner, M., & DiStaso, M. W. (2008). The source cycle: How traditional media and weblogs use each other as sources. *Journalism Studies*, 9(3), 447–463.

Minder, R. (2015, November 5). Spain's News Media Are Squeezed by Government and Debt. The New York Times. New York. Retrieved from http://www.nytimes.com/2015/11/06/world/europe/as-spains-media-industry-changes-rapidly-some-worry-about-objectivity.html?_r=0

Mitchell, A., Gottfried, J., Shearer, E., & Lu, K. (2017). How Americans Encounter, Recall and Act Upon Digital News. Retrieved from http://assets.pewresearch.org/wp-content/uploads/sites/13/2017/02/08183209/PJ_2017.02.09_Experiential_FINAL.pdf

Mori, I. (2015). Trust in Professions. Retrieved from https://www.ipsos-mori.com/Assets/Docs/Polls/ipsos-mori-veracity-index-2015-topline.pdf

Mullin, B. (2017, February 9). Flush from its Trump-bump, ProPublica is staffing up to cover the president. Pointer.org. Retrieved from www.poynter.org/2017/flush-from-its-trump-bump-propublica-is-staffing-up-to-cover-the-president/448405/

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: a more flexible representation of substantive theory. *Psychological Methods*, 17(3), 313.

Mutz, D. C., & Young, L. (2011). Communication and Public Opinion Plus Ca Change? *Public Opinion Quarterly*, 75(5), 1018–1044.

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

Napoli, P. M. (2008). Toward a model of audience evolution: New Tecnologies and the Transformation of Media Audiences (Working Paper Series No. 15). Retrieved from http://fordham.bepress.com/mcgannon_working_papers/15

Neuman, R. W., Guggenheim, L., Mo Jang, S., & Bae, S. Y. (2014). The dynamics of public attention: Agenda-setting theory meets big data. *Journal of Communication*, 64(2), 193–214.

Neuman, W. R. (1990). The threshold of public attention. *Public Opinion Quarterly*, 54(2), 159–176.

Neuman, W. R. (2001). The impact of the new media. Mediated Politics: Communication in the Future of Democracy, 299–320.

Newman, M. (2010). Networks: an introduction. Oxford University Press.

Newman, M. E. J. (2012). Communities, modules and large-scale structure in networks. *Nature Physics*, 8(1), 25–31.

Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, 69(2), 26113. http://doi.org/10.1007/978-3-540-88192-6_64

Newman, N., Fletcher, F., Levy, D. A., & Nielsen, R. K. (2016). Digital News Report 2016. London, UK.

Newman, N., Levy, D. A., & Nielsen, R. K. (2015). Reuters Institute Digital News. Report 2015. Tracking the future of news. Retrieved from https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Reuters Institute Digital News Report 2015_Full Report.pdf

Newton, K. (2001). Media and Communication Questions. European Social Survey Core Questionnaire Development. London.

Nielsen, R. K., & Schrøder, K. C. (2014). The relative importance of social media for accessing, finding, and engaging with news: an eight-country cross-media comparison. *Digital Journalism*, 2(4), 472–489.

Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, 32(3), 245–251.

Osborne, P. (2015). Why I have resigned from the Telegraph. Retrieved January 1, 2015, from https://www.opendemocracy.net/ourkingdom/peter-oborne/why-i-have-resigned-from-telegraph

Papacharissi, Z. (2002). The virtual sphere: The internet as a public sphere. *New Media & Society*, 4(1), 9–27. http://doi.org/10.1177/14614440222226244

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

Pariser, E. (2011b). The filter bubble: What the Internet is hiding from you. Penguin UK.

Park, S., Ko, M., Lee, J., Song, J., & others. (2013). Agenda Diversity in Social Media Discourse: A Study of the 2012 Korean General Election. In ICWSM.

Pavlik, J. V. (2001). Journalism and new media. Columbia University Press.

Pew Research Center. (2015). State of the News Media 2015.

Picard, E. (2008). 'Hyperlinking and Advertising Strategy. In J. Turow & L. Tsui (Eds.), The Hyperlinked Society. Questioning connection in digital age. (pp. 159–164). Ann Arbor, MI:University of Michigan Press.

Pilny, A., & Shumate, M. (2012). Hyperlinks as extensions of offline instrumental collective action. *Information, Communication & Society*, 15(2), 260–286.

Posada de la Concha, P., Gutiérrez, A., & Hernández, H. (2015). Impacto del Nuevo Artículo 32.2 de la Ley de Propiedad Intelectual.

Price, A., & Grann, V. R. (2012). Portrayal of complementary and alternative medicine for cancer by top online news sites. *The Journal of Alternative and Complementary Medicine*, 18(5), 487–493.

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

Prior, M. (2008). Are hyperlinks "weak ties"? In J. Turow & L. Tsui (Eds.), The Hyperlinked Society: Questioning Connections in the Digital Age (pp. 227–249). University of Michigan Press Ann Arbor, MI.

Prior, M. (2009). The Immensely Inflated News Audience: Assessing Bias in Self-Reported News Exposure. *Public Opinion Quarterly*, 73(1), 130–143.

Rawls, J. (2009). A theory of justice. Harvard University Press.

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

Rhoades, S. A. (1993). The herfindahl-hirschman index. Fed. Res. Bull., 79, 188.

Riordan, K. (2014). Accuracy, Independence, and Imparciality: How legacy media and digital natives approach standards in the digital age. Reuters Institute for the Study of Journalism. Retrieved from https://reutersinstitute.politics.ox.ac.uk/publication/accuracy-independence-and-impartiality

Robinson, S. (2007). "SOMEONE'S GOTTA BE IN CONTROL HERE" The institutionalization of online news and the creation of a shared journalistic authority. *Journalism Practice*, 1(3), 305–321.

Robles, J. M., Molina, Ó., & De Marco, S. (2012). Participación política digital y brecha digital política en España. Un estudio de las desigualdades digitales. Arbor, 188(756), 795–810.

Ronen, S., Gonçalves, B., Hu, K. Z., Vespignani, A., Pinker, S., & Hidalgo, C. A. (2014). Links that speak: The global language network and its association with global fame. *Proceedings of the National Academy of Sciences*, 111(52), E5616–E5622.

Rovall, M., & Bergstrom, C. . (2008). Maps of Random Walks on Complex Network Reveal Community Structure. In *Proceedings of the National Academy of Sciences*, (p. 105(4), 1118-1123.).

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

Sagarra, O., Font-Clos, F., Pérez-Vicente, C. J., & Díaz-Guilera, A. (2014). The configuration multi-edge model: Assessing the effect of fixing node strengths on weighted network magnitudes. EPL (*Europhysics Letters*), 107(3), 38002.

Schoepp, S. (2016, March 23). Reporteros en el Salón. Ctxt.es. Madrid. Retrieved from http://ctxt.es/es/20160323/Politica/4979/reporteros-salon-periodismo-espa%25C3%25B1a-Comunicaci%25C3%25B3n-Viva-la-prensa-escrita-Documentos-CTXT.htm

Shaw, D. L., & Hamm, B. J. (1997). Agendas for a public union or for private communities? How individuals are using media to reshape American society. *Communication and Democracy*, 209–230.

Shoemaker, P. J., & Vos, T. (2009). Gatekeeping theory. Routledge.

Snijders, T. A. B. (2011). Statistical models for social networks. *Annual Review of Sociology*, 37, 129–151. http://doi.org/10.1073/pnas.0706851105

Snijders, T. A. B., de Bunt, G. G., & Steglich, C. E. G. (2010). Introduction to stochastic actor-based models for network dynamics. *Social Networks*, 32(1), 44–60.

Strömbäck, J., & Shehata, A. (2010). Media malaise or a virtuous circle? Exploring the causal relationships between news media exposure, political news attention and political interest. *European Journal of Political Research*, 49(5), 575–597. http://doi.org/10.1111/j.1475-6765.2009.01913.x

Stroud, N. J. (2010). Polarization and partisan selective exposure. *Journal of Communication*, 60(3), 556–576.

Sunstein, C. R. (2001). Republic 2.0. Princeton N.J.: Princeton University Press.

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

SWR. (2016). Mitarbeiter fordern Revolution von unten. Retrieved March 31, 2016, from http://www.swr.de/swr2/kultur-info/krise-beim-spiegel-mitarbeiter-fordern-revolution-von-unten/-/id=9597116/did=17196744/nid=9597116/yexa7/index.html

Tan, Y., & Weaver, D. H. (2013). Agenda diversity and agenda setting from 1956 to 2004: What are the trends over time? *Journalism Studies*, 14(6), 773–789.

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

Taneja, H., & Wu, A. X. (2014). Does the Great Firewall really isolate the Chinese? Integrating access blockage with cultural factors to explain Web user behavior. *The Information Society*, 30(5), 297–309.

Teorell, J., Torcal, M., & Montero, J. R. (2007). Political participation: Mapping the terrain. *Citizenship and Involvement in European Democracies: A Comparative Perspective*, 17, 334–357.

Tepper, F. (2015). Snapchat redesign moves "Discover" to top of stories tab. Retrieved July 15, 2015, from http://techcrunch.com/2015/07/13/snapchat-redesign-moves-discover-to-top-of-stories-tab/

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

Tewksbury, D. (2005). Seeds of Audience Fragmentation: Specialization in the Use of Online News Sites, *The. J. Broad. & Elec. Media*, 49, 332.

The Pew Research Center. (2015). Millennials and Political News. Retrieved from http://www. journalism.org/files/2015/06/Millennials-and-News-FINAL-7-27-15.pdf

Times, N. Y. (2014). Innovation. New York. Retrieved from http://www.scribd.com/ doc/224608514/The-Full-New-York-Times-Innovation-Report

Trilling, D., & Schoenbach, K. (2012). Skipping current affairs: The non-users of online and offline news. *European Journal of Communication*, 267323112453671.

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

Tsui, L. (2008). The hyperlinks in Newspapers and Blogs. In J. Turow & L. Tsui (Eds.), The Hyperlinked Society. Questioning Connection in Digital Age. (pp. 70–84). Ann Arbor, MI: The University of Michigan Press.

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

Turow, J. (2007). Breaking up America: Advertisers and the new media world. Chicago, IL: University of Chicago Press.

Turow, J., & Tsui, L. (2009). The Hyperlinked society: Questioning Connections in the Digital Age. University of Michigan Press.

Valentino, N. A., Hutchings, V. L., Banks, A. J., & Davis, A. K. (2008). Is a worried citizen a good citizen? Emotions, political information seeking, and learning via the internet. *Political Psychology*, 29(2), 247–273.

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 http://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.

Vergeer, M., & Franses, P. H. (2016). Live audience responses to live televised election debates: time series analysis of issue salience and party salience on audience behavior. *Information, Communication & Society*, 19(10), 1390–1410. http://doi.org/10.1080 /1369118X.2015.1093526

Verkami.com. (2015). Crític Periodisme d'Investigació. Retrieved April 10, 2016, from http://www.verkami.com/projects/9138-critic-periodisme-d-investigacio

Vogt, N., & Mitchell, A. (2016). Crowdfunded Journalism: A Small but Growing Addition to Publicly Driven Journalism. Washington. Retrieved from http://www.journalism.org/2016/01/20/crowdfunded-journalism/

Wanta, W., & Ghanem, S. (2007). Effects of agenda setting. *Mass Media Effects Research: Advances through Meta-Analysis*, 37–51.

Weber, M. S., & Monge, P. (2011). The flow of digital news in a network of sources, authorities, and hubs. *Journal of Communication*, 61(6), 1062–1081.

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. http://doi.org
/10.1111/j.1460-2466.2011.01616.x

White, D. M. (1950). The Gate Keeper: A Case Study in the Selection of News. *Journalism and Mass Communication Quarterly*, 27(4), 383.

Williams, A. P., Trammell, K. D., Postelnicu, M., Landreville, K. D., & Martin, J. D. (2005). Blogging and Hyperlinking: use of the Web to enhance viability during the 2004 US campaign. *Journalism Studies*, 6(2), 177–186.

Witherow, J., & Ivens, M. (2016, April 1). A letter from the editors. The Times. London, UK. Retrieved from http://www.thetimes.co.uk/edition/news/a-letter-from-the-editors-fdrq0p2p0

Wojcieszak, M. E., & Mutz, D. C. (2009). Online groups and political discourse: Do online discussion spaces facilitate exposure to political disagreement? *Journal of Communication*, 59(1), 40–56.

Wolfsfeld, G., Yarchi, M., & Samuel-Azran, T. (2015). Political information repertoires and political participation. New Media & Society, 1461444815580413.

Wu, L., & Ackland, R. (2014). How Web 1.0 fails: the mismatch between hyperlinks and clickstreams. *Social Network Analysis and Mining*, 4(1), 1–7.

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.

Xalabarder, R. (2016). Press Publisher Rights in the New Copyright in the Digital Single Market Draft Directive. London, UK. Retrieved from https://zenodo.org/record/183788/files/CREATe-Working-Paper-2016-15.pdf

Zhu, J.-H. (1992). Issue competition and attention distraction: A zero-sum theory of agenda-setting. *Journalism & Mass Communication Quarterly*, 69(4), 825–836.

In-strenght Distribution of the News and Audience Network

<div style="text-align: right">**A**</div>

Fig. A.1.: In-strength Distribution of the News Network
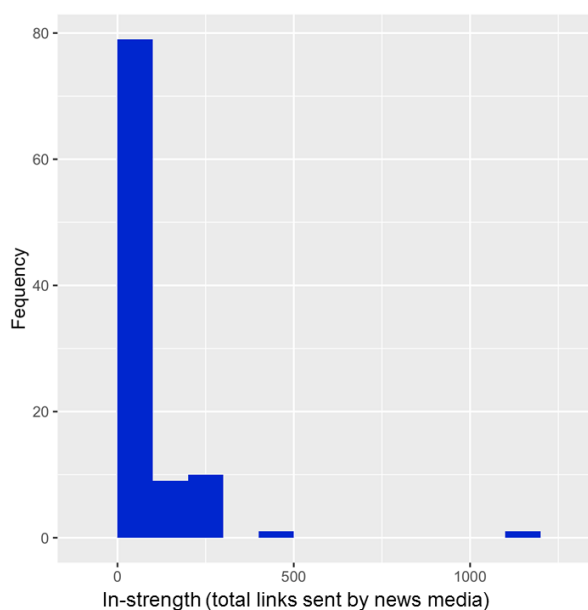
Fig. A.2.: In-strength Distribution of the Audience Networks Representing News Consumption of the General and Young Population
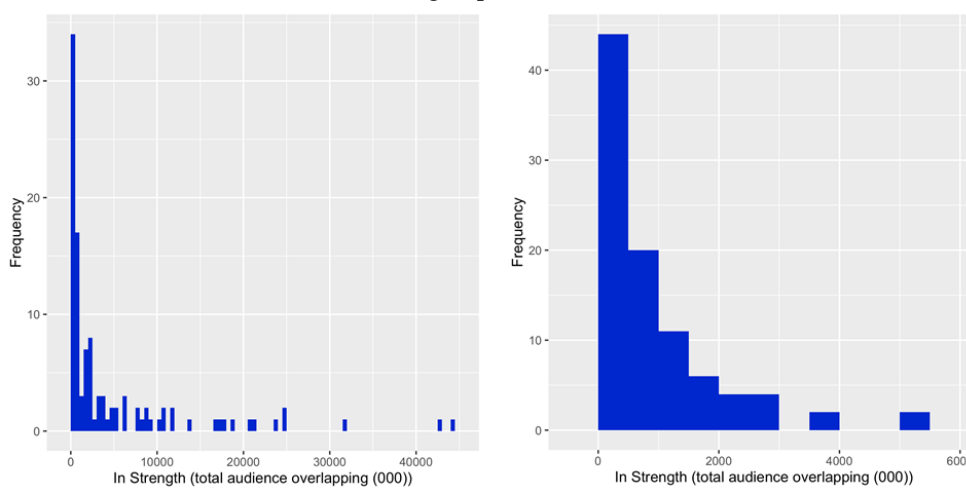
Desktop and Mobile Survey Samples

Fig. B.1.: Survey Question About Getting Access to Television News

Note: Participants where asked about their preferences on television channels to get access to news. The image above shows the design of this question for the web-based survey. Logos of 25 television channels were provided and randomize for each participant. Top channels in terms of audience where included in the list.

Fig. B.2.: Survey Question about the Second Most Important Problem
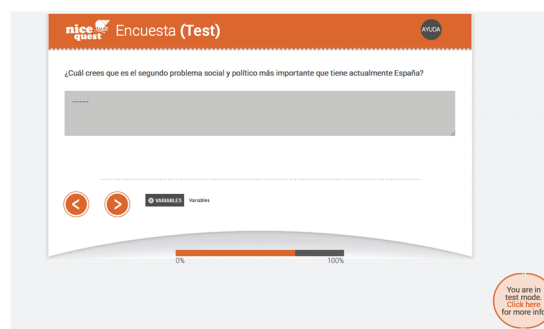
Note: Participants were asked about the three most important problems facing Spain. These questions were administered using the graphic interface shown above.

Fig. B.3.: Survey Question about other Most Important Problems
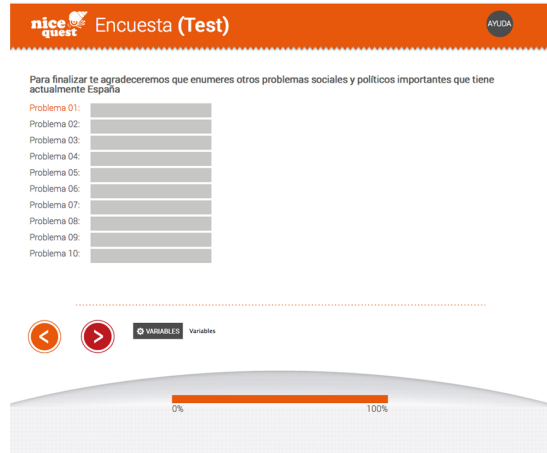
Note: Participants were asked to list other most important problems facing Spain. This question was administered using the graphic interface above and, as we did on the previous questions, participants were told they could leave an item blank by hitting the right arrow.

Fig. B.4.: Mobile Questionnaire Sample

Note: Participants were allowed to answer the questionnaire using their mobile devices, cell phones, and tablets. The graphical interface was designed and adapted to be used on these devices too. Several tests were run to guarantee they work well on any platform.

Survey Instrument and Question Wording

<div style="text-align: right">C</div>

Presentation

We would like to thank you for participating in this study about the Internet and political information.

The survey you are about to answer belongs to a non-profit research study led by a university.

Please, before you answer, read carefully the following information:

Duration: It will take you 10 minutes to complete this survey

What is it about?

This survey includes 19 questions. There are no correct or incorrect answers.

Please, answer with sincerity.

Is my privacy protected?

Yes, it is. All your data will be anonymized before being transferred to us. Hence, your privacy is guaranteed.

Which is the goal of this survey?

This information will be used solely for an academic study led by a Spanish university.

Thank you very much for your participation!

Screening Questions

1. Do you use the Internet as a source of political news?

- No

- Yes

1.2 How often do you use Internet as a source of political news?

- Everyday

- 4 to 5 days per week

- 2 to 3 days per week

- During the weekends

- Almost never

2. Do you read blogs of political information? Are you a news blog subscriber?

- No

- Yes

2.1 Please could you tell us which blogs of political information do you read?

3. Is there a politician among your Facebook friends? (Local representatives, MPs, presidents) Or do you follow a political party on Facebook?

- No

- Yes

4. Do you follow a political party or politician on Twitter?

- No

- Yes

4.1 Which political party or politician do you follow on Twitter?

Many thanks for your time and answering this survey!

Pre-questionnaire

5. Could you please tell us where you live? (Province)

6. Could you please tell us in which town or city you live?

7. Gender

  • Man

  • Woman

8. Age

9. Level of studies

  • No studies

  • Primary School

  • Secondary School

  • College, Master Studies, PhD

9.1 Type of studies

We would like to know how you keep up with current political events.

10. How do you prefer to access news?

  • Newspapers

  • Online newspapers

  • Television

  • Radio

  • Internet (e.g., social networks, blogs, forums, online aggregators)

11. How often do you read political news in a newspaper (offline edition)?

- Everyday

- 4 to 5 days per week

- 2 to 3 days per week

- During the weekends

- Almost never

- Never

11.1 Which newspaper do you read?

- El País

- 20 minutos

- El Mundo

- La Vanguardia

- La Voz de Galícia

- El Periódico

- Abc

- El Correo

- La Nueva Espańa

- Faro de Vigo

- La Razón

- Levante

- Heraldo de Aragón

- El Diario Vasco

- Información Alicante

- La Verdad

- Diario de Navarra

- Ideal de Andalucía

- Última Hora

- El Norte de Castilla

- El Día

- Expansión

- El Diario Montańés

- El Comercio

- La Provincia

- El Punt Avui

- Diario de León

- Ara

- Hoy Diario de Extremadura

- Sur

- Las Provincias

- Canarias 7

- Diario de Noticias

- El Progreso

- Córdoba

- None of the mentioned above. I read another newspaper. Tell us the name:

12. Do you pay a newspaper subscription?

- No

- Yes

12.1 To which newspaper do you subscribe?

13. How do you prefer to access news on the Internet?

- PC

- Mobile phone

- Tablet

14. How often do you watch news on television?

- Everyday

- 4 to 5 days per week

- 2 to 3 days per week

- During the weekends

- Almost never

- Never

14.1 On which television channels do you prefer to watch news?

- Tele 5

- Antena 3

- La1

- La Sexta

- Cuatro

- Canal Sur

- La2

- TV3

- TVG

- Telemadrid

- TvCan

- 8TV

- Etb2

- CMT

- Aragón TV

- Tpa

- Canal Extremadura

- 3/24

- Etb1

- Ib3

- None of the mentioned above. I watch news on another tv channel. Tell us the name:

15. How often do you listen the radio to get news?

- Everyday

- 4 to 5 days per week

- 2 to 3 days per week

- During the weekends

- Rarely

- Never

15.1. From which radio station do you prefer to access news?

- Ser

- Onda Cero

- Cope

- RNE

- Rac 1

- Catalunya Ràdio

- Canal Sur Radio

- EsRadio

- Radio Euskadi

- Radio Galega

- Euskadi Irratia

- None of the mentioned above. I listen another radio station. Tell us the name:

We will now ask you some questions about politics.

16. Are you interested in politics?

- Very much interested

- Somehow interested

- Not very much interested

- Not at all

17. When talking about politics, we use the expressions left and right.

Please, could you tell us where you are located on a line from 0 to 10?
0 equals extreme left and 10 extreme right

Extreme Left	0	1	2	3	4	5	6	7	8	9	10	Extreme Right

18. Do you always vote for the same political party?

- No

- Yes

19. As you well know, there are several political parties and coalitions in Spain.

Which political party or coalition do you vote for?

(Select one option from the list below)

- PP

- PSOE

- IU (ICV en Catalunya)

- UPyD

- Podemos

- Ciudadanos

- CIU

- Amaiur

- PNV

- ERC

- BNG

- CC-NC

- Compromís-Equo

- Geroa Bai

- Another party. Name :

- Blank ballot

- Null vote

- I do not vote

20. Please tell us your current situation:

- Self-employed

- Employee

- Student

- Domestic worker

- Unemployed

- Do not work (retired, disabled worker)

21. What is your current job?

22. Finally, could you please tell us how many surveys, including this one, you have answered within the last three months?

0	1	2	3	4	5	6	7	8	9	10+

We would like to thank you for participating in this survey. Your information will be used for a non-profit study led by a Spanish university.

Thank you !

Post-questionnaire

We would like to thank you for participating in this survey about the Internet and political information.

We will ask you 7 questions about news on the Internet. Please, be honest when answering them.

This survey is anonymous and your privacy is protected. This information will be used for a non-profit study led by a Spanish university.

Thank you very much for your time!

1. Can you remember any news website(s) that you have visited during the last two weeks to keep up with current events?

Use the following lines to write their names:

2. You will find below a list of news websites. During the last two weeks, have you visited any of them?

- Elmundo.es

- Elpais.com

- Abc.es

- elconfidencial

- 20minutos.es

- Lavanguardia.com

- telecinco.es

- meneame.net

- Publico.es

- Rtve.es

- Europapress.es

- Eldiario.es

- Libertaddigital.com

- Antena3.com

- Huffingtonpost.es

- Elperiodico.com

- Lavozdegalicia.es

- Lainformacion.com

- cuatro.com

- cadenaser.com

- lasexta.com

- vozpopuli.com

- elcorreo.com

- lasprovincias.es

- Ara.cat

- Diariosur.es

- Larazon.es

- Levante.emv.com

- Heraldo.es

- Que.es

- Elplural.com

- Laverdad.es

- Diarioinformacion.com

- infolibre.es

- Elmundotoday.com

- Tv3.cat

- Farodevigo.es

- Naciodigital.com

- Canarias7.es

- Diariovasco.com

- Elconfidencialdigital.com

- Cope.es

- None of the mentioned above. Tell us the name:

3. News media can be classified according to their editorial lines. We normally use the expressions left, centre, and right when talking about media slant.

Can you classify the websites that you visited within these categories?

4. During the last two weeks, have you got political news that you had not intentionally searched for?

- No

- Yes

5. We would like to know your opinion about the general situation in our country.

6. What do you think is the most important political or social problem in Spain? You can skip this question if you want.

7. What do you think is the second most important political or social problem in Spain? You can skip this question if you want.

8. What do you think is the third most important political or social problem in Spain? You can skip this question if you want.

9. Finally could you tell us whether there are any other important political or social problems in Spain? You can skip this question if you want.

Thank you very much!

D

Sociodemographic Characteristics of the Spanish Population and the Survey Sample

Tab. D.1.: Sociodemographic Characteristics of the Spanish Population and the Sample Studied

Variables	Mean	Standard Deviation	Min.	Max.
CIS				
Age	46.280	17.180	18	92
Education	2.730	0.880	1	4
Political interest	2.250	0.940	1	4
Ideology	4.800	1.930	1	10
Sample				
Age	36.130	13.730	18	74
Education	3.480	0.580	2	4
Political interest	2.760	0.760	1	4
Ideology	3.970	2.070	0	10

Tab. D.2.: Intercode Reliability Scores for each Variable Measuring the Public Agenda

	Percent Agreement	N Agreements	Disagreements	Cases	Decisions
MIP1	89.77	676	77	753	1506
MIP2	86.71	653	100	753	1506
MIP3	78.88	594	159	753	1506
MIP4	77.02	580	173	753	1506
MIP5	77.95	587	166	753	1506
MIP6	77.68	585	168	753	1506
MIP7	79.01	595	158	753	1506
MIP8	82.07	618	135	753	1506
MIP9	85.25	642	111	753	1506
MIP10	89.90	677	76	753	1506
MIP11	92.56	697	56	753	1506
MIP12	83.89	707	46	753	1506
MIP13	94.82	714	39	753	1506

Most Important Problems in Catalonia

In justifying our decision to not include Catalonia in our sample, it is worthy to mention that in a very short period of time, those that would have chosen an independent Catalan state -–in a multi-choice question[1] regarding their preferred constitutional arrangement-– almost tripled between 2006 (Muñoz & Tormos, 2015) and 2015. Consequently, the saliency of this problem in the public agenda sharply increased before we launched our study. Two major events close to the dates of launching our study boosted the salience of this issue: an unofficial referendum for the independence that took place in November 2014 and the call for regional elections that took place in September 2015, and were framed as an independence referendum by the pro-independence parties. Consequently, the Catalan-Spanish relations became the fourth most important problem for the Catalan population at the time our study took place (see Figure E.1).

Fig. E.1.: Most Important Problems for Catalan Population Over Time
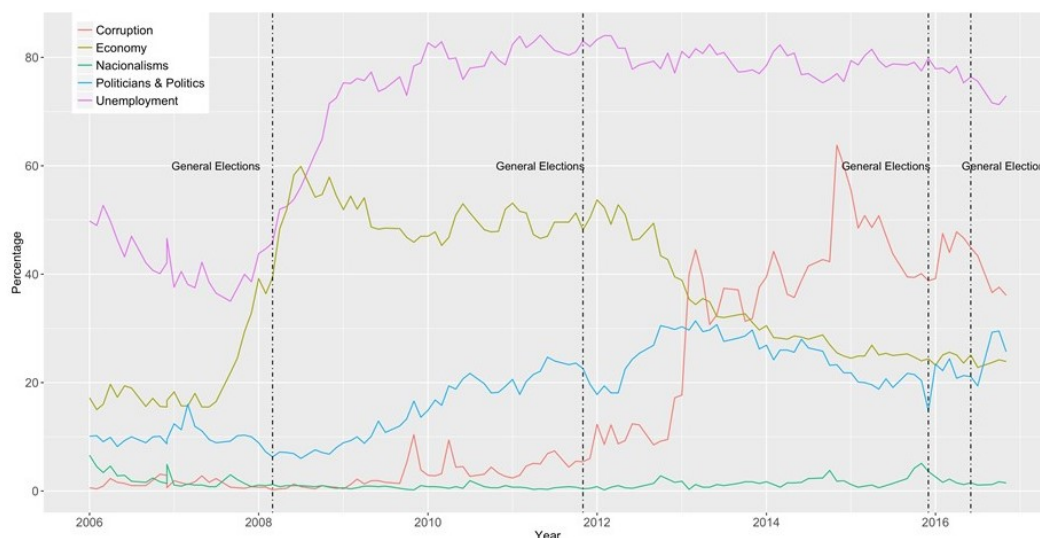
Note: Data was collected from the *Catalan Public Barometer* by the *Centre d'Estudis d'Opinió* a polling center funded by the Catalan Government.

More precisely, 12.2% of Catalans regarded the issue as among the most salient ones in October 2014 and it remained in the same position in March 2015 (7.7%), according to the *Catalan Public Barometer* by the *Centre d'Estudis d'Opinió*, a polling

---

[1]The other options in the multi-opinion question were federalism, autonomy (i.e., the *status quo*) and regionalism.

center funded by the Catalan Government. In stark contrast, the Spanish population did not perceived this issue as important. Figure 4.1 shows the evolution of the top 4 most important problems in Spain from 2006 to 2016; none of them are related to the Catalan political situation. Yet, we have included in this figure the evolution of nationalisms regarded as problem by the Spanish population to illustrate its negligible importance in the nation-wide context. When we launched our study, this item was considered the tenth most important problem, listed only by 3.8% of the Spanish population. Then, in March 2015, just 0.9% of the population considered nationalisms among the most important problems.

F OLS Estimation of Common Agenda and Agenda Diversity and Capacity

Tab. F.1.: OLS Estimation of Agenda Diversity, Common Agenda and Carrying Capacity (Reported Data from Individuals of the Tracked Sub-Sample)

	Agenda Diversity	Common Agenda	Agenda Capacity
HHI Online Reported News diet	0.127**	0.635***	2.248***
	(0.060)	(0.237)	(0.837)
HHI Offline Reported News Diet	0.087*	-0.227	1.770***
	(0.048)	(0.190)	(0.671)
Online News Sources Reported	-0.0002*	-0.001**	-0.002*
	(0.0001)	(0.0004)	(0.001)
HHI Overstated New Diet	0.019	-0.221	-0.262
	(0.043)	(0.171)	(0.604)
Female	0.045**	-0.039	0.524*
	(0.021)	(0.083)	(0.293)
Age (25-34)	0.024	0.101	0.484
	(0.030)	(0.119)	(0.419)
Age (36-50)	0.053*	0.312***	0.563
	(0.030)	(0.119)	(0.421)
Age (= or >50)	0.096***	0.380***	1.556***
	(0.031)	(0.122)	(0.430)
Education (College)	0.019	0.012	0.759**
	(0.021)	(0.084)	(0.297)
Device (Mobile phone)	-0.028	-0.022	-0.413
	(0.026)	(0.105)	(0.369)
Device (Tablet)	-0.006	-0.120	0.146
	(0.050)	(0.196)	(0.691)
Political Interest	0.012	0.481***	-1.307**
	(0.044)	(0.172)	(0.609)
Freq. Online News Consumption	-0.023	-0.281*	0.038
	(0.041)	(0.163)	(0.576)
Freq. Newspaper use for News	0.010	-0.017	-0.095
	(0.032)	(0.127)	(0.448)
Freq. Television use for News	-0.011	-0.158	0.803
	(0.036)	(0.142)	(0.500)
Freq. Radio use for News	0.009	-0.151	0.578
	(0.031)	(0.121)	(0.429)
Constant	0.413***	1.550***	3.773***
	(0.056)	(0.222)	(0.786)
Observations	372	372	372
R^2	0.106	0.110	0.131
Adjusted R^2	0.065	0.070	0.092
Residual Std. Error (df = 355)	0.195	0.772	2.727
F Statistic (df = 16; 355)	2.617***	2.733***	3.347***

Note: *$p<0.1$; **$p<0.05$; ***$p<0.01$

List of Figures

List of Tables

Colophon

This thesis was typeset with $\text{\LaTeX}\,2_\varepsilon$. It uses the *Clean Thesis* style developed by Ricardo Langner.

Download the *Clean Thesis* style at `http://cleanthesis.der-ric.de/`.