Fabio Giglietto, Nicola Righetti, Giada Marino, Luca Rossi

Multi-Party Media Partisanship Attention Score. Estimating Partisan Attention of News Media Sources Using Twitter Data in the Lead-up to 2018 Italian Election

Ente di afferenza:
Universit Verona (univr)

Fabio Giglietto, Nicola Righetti, Giada Marino and Luca Rossi

Multi-Party Media Partisanship Attention Score

Estimating Partisan Attention of News Media Sources Using Twitter Data in the Lead-up to 2018 Italian Election

Saggi

ABSTRACT

The ongoing radical transformations in communication ecosystems have brought up concerns about the risks of partisan selective exposure and ideological polarization. Traditionally, partisan selective exposure is measured by cross-tabulating survey responses to questions on vote intentions and media consumption. This process is expensive, limits the number of news outlets taken into account and is prone to the typical biases of self-reported data. Building upon previous works and with a specific focus on the online media environment, we introduce a new method to measure partisan media attention in a multi-party political system using Twitter data from 2018 Italian general election. Our first research question addresses the effectiveness of this method by measuring the extent to which our estimates correlate with partisan newspaper consumption measured by the latest Italian National Election Studies (ITANES) survey. Once established the reliability of our method, we employ these scores and measures to analyze the Italian digital media ecosystem in the lead-up to March 2018 election. The traditionally high level of political parallelism that characterizes both the Italian press and TV sectors is only partially reflected in a digital media ecosystem where partisan selective and cross-cutting exposure seem to coexist. Results also point out that certain online partisan communities tend to rely more on exclusive news media sources.

Keywords: selective exposure, attention, polarization, insularity, Italian 2018 election.

1. Introduction

Allegations of biased political news reports and partisan coverage of public issues are common throughout the entire history of media. However, during the recent years, a combination of factors ranging from the declining public trust in established media (Zuckerman, 2018) to the rise of social media and popular alternative

digital media sources (Benkler, Faris and Roberts, 2018) have brought up concerns toward the social and political implications of a surge in partisan selective exposure (Lazarsfeld, Berelson and Gaudet, 1968) driven by an increasingly fragmented media environment. Along the same line, the concept of «echo-chambers» theorized by Sunstein (2017) describes the effects of combining the tendency to self-entrap into like-minded groups of individuals with the feature that allows social media users to craft their networks as a result of platform affordances' constraints, algorithmic filters, prioritization and personal choices (Bakshy, Messing and Adamic, 2015a).

On the supply side, by offering effective opportunities for the development of micro (and macro) niches of audiences, social media have fostered the success of hyper-partisan media. Such hyper-partisan journalistic and quasi-journalistic sources (McNair, 2017) that nowadays compete for users' online attention with mainstream news organizations, often disregard established ethical values in professional news reporting or even deliberately publish false or misleading news stories (Bhat, 2018).

To better understand the challenges posed by this transformation to modern liberal democracies, a wide range of studies have been carried on with the aim of estimating the partisan consumption of both traditional news media and social media, and to assess its effect on political and ideological polarization. Concerns over the effects of increased polarization are especially high in two-party political systems, such as the United States, where a range of systemic peculiarities might have a relevant impact on patterns of self-segregation of political users and sources' polarization on social media. Conversely, studies considering other context and multi-party systems are still limited.

Guided by the intent of investigating partisan political news consumption on social media within a highly different context than the US, this study introduces and assesses the effectiveness of an original method aimed at measuring the attention devoted by different partisan online communities to digital news media sources using Twitter data in a multi-party political system.

To showcase the method, we employ its measures to analyze the Italian online media system (634 online news sources) in the lead-up to the 2018 election. Italy, a multi-party parliamentary democracy, has been often described as «polarized-pluralist» (Hallin and Mancini, 2004) due to the limited development of journalistic professionalization, poorly widespread elite press (Lizzi and Pritoni, 2014) and high level of political parallelism, both in the press and in the television sectors. The resulting picture provides an assessment of the system-wide level of polarization, allows to discuss the role played by cross-partisan and insular news media sources for different partisan online communities and highlights the implications of the method for future studies.

2.   Literature review and research questions

Classifying the political leaning of news media outlets is a relevant issue in political communication research. A traditional method to accomplish this goal consists in asking people to directly classify news outlets on a left-to-right ideological scale (Pew Research Center, 2018), or in linking parties to news media sources by asking respondents their vote intention and their information diet (Mancini and Roncarolo, 2018). This process is expensive, inherently limits the number of news outlets taken into account and is prone to the typical biases of self-reported data (Haenschen, 2019; Prior, 2009).

To overcome these limits, a new strand of studies begins using digital data to measure media partisanship both in the offline and online media system, producing three main approaches in the field of political and computer science.

The first aims at estimating the political bias of news media outlets by relying on analysis of articles' content, for example counting names of think tanks and policy groups cited both by newspapers and parliamentarians (Groseclose and Milyo, 2005), comparing their language similarity (Gentzkow and Shapiro, 2010) or analyzing perspective about Supreme Court cases (Ho and Quinn, 2008).

The second strand of research aims at classifying the political leaning of a variety of items like documents, news articles or blogs, by exploiting textual features as well as other digital information. It includes, for example, studies aimed at analyzing textual data with automated or semi-automated methods to classify documents according to their policy positions (Laver, Benoit and Garry, 2003), ideological perspectives (Lin, Xing and Hauptmann, 2008) or party affiliation (Yu *et al.*, 2008), while other studies take advantage of features like news stories votes cast by users of social news aggregator services to classify news articles (and users) as liberal or conservative (Zhou *et al.*, 2011), hypertextual cocitations to estimate the political orientation of hypertext documents (Efron, 2004), or hyperlinks to classify political blogs (Lin and Cohen, 2008).

With regard to these strands of research, we can observe that classifying online news media sources by analyzing news stories textual content is theoretically possible but highly resource consuming. Many studies dealing with classification problems employ semi-automated methods of analysis which require a part of hand coding or data preparation. Furthermore, exploiting features like hyperlinks is not always possible (e.g. in the Italian journalistic practices is not common to link external contents in news stories).

The third strand of research includes studies aimed at estimating the political leaning of Twitter users, for example, by analyzing the accounts they follow (Barberá, 2015), their tweets, retweets and retweeters (Ming Fai Wong *et al.*, 2016),

or through automated content analysis. Similarly to what is done through a question-naire, the political orientation of social media users can be used to classify, in turn, the political leaning of the online content they share (Bakshy, Messing and Adamic, 2015a). In this way, the observation of people's digital behaviour – instead of their responses to the researcher's questions – can be used to link political parties to news articles and consequently to news media outlets.

The last approach was adopted by a recent study on the American online media environment during the 2016 U.S. Presidential Election (Faris *et al.*, 2017). This study proposed a method – named «Media Partisanship Attention Scores» (MPAS) – aimed at automatically determining the political leaning of news outlets by classifying, first, the political orientation of social media users, and then, the news sources they share online.

This method was conceived for the US two-party political system and required some adaptations to work for the Italian multi-party political system. In this paper, we introduce a variation of the MPAS named «Multi-Party Media Partisanship Attention Score» (MP-MPAS) designed to measure the attention devoted by online partisan communities to different news media in the Italian context.

Given the modification introduced and the fact that the validation of the original MPAS was based on data only available for US news outlets (Bakshy, Messing and Adamic, 2015b), we estimated the effectiveness of our method by comparing its results with data about the partisan consumption of eleven major Italian newspapers obtained through a survey. We thus formulated the following research question:

> *RQ1.* To what extent do survey-based estimates of news media partisan alignment overlap with estimates of partisan attention based on Twitter data?

Multiple studies using a wide range of methodologies attempted to empirically verify the existence of phenomena like echo chambers and filter bubbles (Pariser, 2011; Flaxman, Goel and Rao, 2013). Estimating the partisan attention of the news sources circulating online is useful to better understand the political configuration of the online news media ecosystem, and more specifically its degree of ideological polarization and partisan communities' self-segregation.

Despite the debate about political polarization also features studies pointing out its positive outcomes (e.g. Layman, Carsey and Horowitz, 2006), most of the scholars agree on its adverse effects on the health of the public debate within a liberal democracy (Dahl, 1998). Highly polarised communities tend to lack the common ground that is necessary for an effective public debate and are vulnerable to the spread of problematic information (Jack, 2017).

Concerns on political polarization and its relationship with selective exposure (Stroud, 2010), i.e. the preference for like-minded media contents (Holbert *et al.*, 2010), brought a renewed interest in understanding its main drivers. Most of the studies point out the cognitive individual's disposition for like-minded contents and the implications of digital technologies on political polarization (see Gentzkow, 2006; Newman *et al.*, 2017).

Both the reduction of cognitive dissonance (Festinger, 1962) and the homophilic leaning seems to encourage individuals to avoid messages clashing with their opinions (Garrett and Stroud, 2014), favouring the exposure to like-minded media outlets (Dvir-Gvirsman, 2017).

The hybrid media system (Chadwick, 2017) is characterised by the newly prominent role of the Internet, and specifically of social media in news consumption practices. According to Prior (2007), these changes in media technology environment are crucial for political polarization development: the audience ability to craft their information diet have contributed to add further concerns about the phenomenon of partisan selective exposure. Additionally, media outlets, often pushed by the decreasing advertising revenues, have boosted the production and circulation of politically biased news stories in order to bait the audience attention (Iyengar and Hahn, 2009). As a consequence, Hollander (2008) argued that this news sources behaviour further contributed to the rise of homophilic niches of publics.

According to Sunstein (2017), in this highly fragmented media environment the homophilic effect of recommendation, sorting and filtering algorithms should also be taken into account (Bakshy, Messing and Adamic, 2015a). Due to the effect of these algorithms, social media news consumers risk finding themselves entrapped in ideological echo-chambers, where existing opinions and biases are fostered and contact with different opinions, instead, limited.

Beside the algorithms, the role played by users' social media contacts on news exposure has to be taken into account. Messing and Westwood (2014), for example, argued that social factors augment the probability that users are stumbling upon political news that contradicts their political beliefs. «Social media news communities» (Saez-Trumper, Castillo and Lalmas, 2013) – i.e. groups of active social media users interested in news consumption and news redistribution through sharing practices – play a crucial role in the process of news delivery and news exposure, multiplying the opportunities of circulation of a news story. In this context, it is thus central to shed some light on how partisan users behave within the social media environment in terms of interactions with political news stories.

Weeks and colleagues (2017) argued that partisan users of social media are particularly engaged in interacting with, and in selective sharing of (Shin and

Thorson, 2017), political news stories favouring their political opinions or their favourite candidate. In this way, each partisan community on social media can strategically exploit information, supporting the idea that social media are a fertile ground for political polarization growing and hyperpartisan news sources spreading.

Guided by the intent to shed light on the level of self-segregation and partisanship of the Italian digital media system in the lead-up to 2018 election and in light of the high level of political parallelism typical of «polarized-pluralist» media systems (Hallin and Mancini, 2004), we expected to observe an uneven distribution of the attention devoted by different partisan communities to digital media sources. At this extent, leveraging on the idea of insularity as a measure of the inequality of the distribution of social media shares performed by different partisan online communities on the stories published by a certain news outlet, we formulate the following research question:

> *RQ2.* To what extent are Italian sources of political news online characterized by high ideological insularity on social media?

3. Measure and methods

Building upon the abovementioned Media Partisanship Attention Scores (Faris *et al.*, 2017), we relayed on Twitter data to characterize the political leaning of news sources. MPAS is based on the frequency of sharing media sources among users who retweeted messages from either of the two main presidential candidates (@realdonadtrump and @hillaryclinton). The general idea is to categorize users first (based on the proportion of their retweets) and then, in turn, the news sources they shared. Faris and colleagues validated their Twitter metric against partisan aligned measures estimated by researchers with special access to Facebook data (Bakshy, Messing and Adamic, 2015a) reporting a high correlation (rho = .94) between the two methods (Faris *et al.*, 2017: 134).

In order to apply the MPAS method to the Italian multi-party political system, we initially identified a set of official Twitter accounts[1] of the main Italian

---

[1] We specifically collected tweets that matched a «retweets_of:» rule for the following Twitter accounts: angealfa, alternativa_pop (Popular Alternative, or «Alternativa Popolare»), bealorenzin, civica_popolare (Popular Civic List, or «Civica Popolare»), giulianopisapia, campoprog (Progressive Camp, or «Campo Progressista»), giorgiameloni, fratelliditalia (Brothers of Italy, or «Fratelli d'Italia», FdI), forza_italia, berlusconi (Forza Italia, FI), verditalia, insieme2018, partsocialista (Together, or «Insieme»), pbersani, articolounomdp, si_sinistra, nfratoianni, possibileit, civati, pietrograsso, robersperanza, lauraboldrini, libe-

political parties and their leaders, considering both the main candidates and prominent political personalities. Using DiscoverText[2], we collected the retweets of these accounts during January 2018 ($N = 216,765$) to estimate the partisan leaning of each contributor based on the proportion of their retweets for each party. Subsequently, we employed DMI-TCAT[3] to collect all the tweets published by the top 5,000 contributors in this retweet dataset between February 1 and March 4, 2018 ($N = 4,385,877$). In order to identify the news domains they shared, we extracted about 1.3 million URLs contained in their tweets. Because Twitter shortens the links shared on the platform[4], we converted them back to their original form (Rudis, 2016). Considering only the tweets containing a link, the Twitter dataset we analyzed included 3,945 users, 3,500,575 tweets and 19,274 unique domains. We further cleaned up the list of domains by removing news web portals (msn.com, news.google.com, etc.), content sharing websites (YouTube, Vimeo, Google Docs, etc.) and those mentioned by less than 31 tweets, that is the average of the tweet distribution per domain. Following this procedure, we obtained a list of 1,372 unique domains.

We matched these domains with those available in a dataset of 2018 election-related political news-stories (Giglietto, 2018), and using the average partisan affiliations of contributors who referenced a certain domain, we calculated a set of ten scores (one for each party we considered, excluding Progressive Camp and Popular Alternative that did not participate in the elections) ranging from 0 to 1 and add up to 1 for each unique news domain. These scores, which we named «MP-MPAS» or «Multi-party Media Partisanship Attention Score», indicate the distribution of partisan attention received by each news domain. After having removed domains shared by too few contributors to calculate the MP-MPAS and those with less than 2 unique URLs in the dataset we ended up with a list of 634 news sources marked by the MP-MPAS score.

Building upon the sets of such MP-MPAS metrics, we designed an insularity score to measure the degree by which a news media source is prominently shared by online actors affiliated to a single party. In other terms, insularity is a measure of the audience polarization of an online media outlet. Considering each party j and media source i whose we calculated MP-MPAS scores, and the Gini coefficient G, we

---

ri_uguali (Free and Equal, or «Liberi e Uguali», LeU), matteosalvinimi, leganordpadania, noiconsalvini, legasalvini (Northern League, or «Lega Nord», LN), luigidimaio, beppe_grillo, mov5stelle (Five Star Movement, or «Movimento Cinque Stelle», M5S), matteorenzi, pdnetwork, paologentiloni (Democratic Party, or «Partito Democratico», PD), emmabonino, radicali, piu_europa (+Europa), maurizioacerbo, direzioneprc, potere_alpopolo (Power to the People, or «Potere al Popolo»).

[2] https://discovertext.com/. More specifically, we relied on the «retweets_of:» rule made available by Twitter Enterprise search API.

[3] https://github.com/digitalmethodsinitiative/dmi-tcat/wiki.

[4] https://help.twitter.com/en/using-twitter/url-shortener.

Figure 1. Insularity formula

$$I_{(i)} = \frac{(MP\,MP\,AS_{max(i,j)} + G_{(i,j)})}{MAX\,(MP\,MP\,AS_{max(i,j)} + G_{(i,j)})}$$

defined the insularity of a news media source $I_{(i)}$ the index [0,1] calculated according to the formula represented in figure 1.

The measure takes into account two properties of the MP-MPAS set of scores obtained by each news outlet. The maximum score [0,1] indicates the amount of attention devoted to the news source by the dominating partisan community. The Gini coefficient [0,1], a widely used measure of the statistical dispersion of a distribution, is used to fine-tune the measure by assessing cases where the attention toward a news outlet is concentrated (higher Gini index) or spread across different communities (lower Gini index).

Finally, using the properties of the insularity distribution (M = .65, SD = .16) and the distance from the mean, we identified four classes of news media sources with different degrees of insularity: «High», «Moderate», «Low», «None» (see tab. 4, presented in «Findings» section).

Using these scores, we also proceeded to adjudicate the media sources with low to high insularity to the partisan community with the highest scores in the set. Media sources attributed to the «None insularity» class were instead added to a «Cross Partisan» category. In other terms, when a media source is adjudicated to a party, it means that the source received a significantly partisan attention (in terms of sharing activity), concentrated among users who retweeted that party. On the other hand, a media source was adjudicated to the «Cross Partisan» category when a heterogeneous audience shared it.

Following the estimates of the partisan attention devoted to different media outlets according to the MP-MPAS method, we assessed the effectiveness of this method by comparing it against the partisan newspapers consumption gathered during the latest iteration of the Italian National Election Studies (ITANES) survey as reported in the tab. 3.2. of Mancini and Roncarolo (2018: 59)[5].

Since MP-MPAS and ITANES data are not directly comparable because of the differences in the respective methodologies and data sources, we pre-processed

[5] Throughout this paper we thus refer to «ITANES data» as the data reported in Mancini and Roncarolo (2018).

Table 1. Preprocessing output of ITANES data

	LeU	PD	M5S	FI	LN	FdI	Tot
Corriere della Sera	48	181	205	97	97	36	664
la Repubblica	200	417	254	54	36	18	979
La Stampa	24	85	61	28	38	19	255
Il Sole 24 Ore	17	56	122	50	61	28	334
Il Giornale	0	5	33	52	60	22	172
Libero	2	13	43	22	47	19	146
il Fatto Quotidiano	34	25	482	17	25	25	608
Il Messaggero	8	30	72	30	25	14	179
Avvenire	4	6	12	10	5	1	38
QN - Il Resto del Carlino	2	29	36	22	26	14	129
Leggo	2	12	37	17	17	8	93
Tot	341	859	1357	399	437	204	3597

the available data through a two-step process. First, we kept only the 6 political parties for which an explicit vote intention was reported (thus excluding the «Other», «Not voting», «Not yet decided/No answer» and «Other parties in coalition with PD» categories), and the 11 major Italian newspapers for which an explicit reading preference was expressed too (excluding the «Other», «I have not read any newspaper/No answer»). Second, since the ITANES data were structured as a matrix m x n where the rows m were the newspapers and the columns n were the political parties, and whose elements $a_{(i,j)}$ were the percentages given by the number of respondents who claimed to read a specific newspaper and to vote for a specific political party out of the total number of readers of that newspaper, we converted these percentages to absolute values by multiplying them by their respective total and rounding up the results. By following this procedure, we obtained a sample of 3,597 respondents who declared to read 11 newspaper and expressed vote intentions for 6 political parties (see tab. 1).

We used Principal Component Analysis (PCA) and chi-square test of «goodness of fit» results as a benchmark to estimate the effectiveness of MP-MPAS scores with specific reference to the 11 major Italian newspapers available in ITANES data (RQ1).

We first applied PCA on both the dataset to identify their main factors and to calculate their correlation via Pearson's correlation coefficient. A significant positive correlation between the two main dimensions identified by the PCA for both MP-MPAS and ITANES dataset would support the idea that the newly introduced method is effective.

We then analyzed ITANES data by using the chi-square test to check statistically significant differences in the newspapers consumption between groups of

respondents who declared different vote intentions. In order to do so, we used the proportions of claimed vote intention shares for each party to calculate the chi-square expected partisan consumption for each newspaper. In other terms, we started from a theoretical situation of equal newspaper consumption across all party voters and we expected to find a proportion of readers for each newspaper equal to the shares of vote intentions for each party.

Following the chi-square test, we analyzed the adjusted standardized residuals applying Bonferroni correction for multiple comparisons[6]. Finally, we adjudicated each newspaper to the group of voters characterized by the most statistically significative chi-square contribution. In the case where no group of voters had shown a statistically significant preference towards a newspaper, that newspaper was labelled as cross-partisan.

We then used the adjudication and insularity distribution derived from the MP-MPAS to analyze the levels of insularity among Italian online political news system as a whole (RQ2).

All the analyses were performed using R software for statistical computing (R Core Team, 2018). Replication data and R code are available at https://dataverse.harvard.edu/dataverse/mine2018.

4.   Findings

The first research question assesses the effectiveness of MP-MPAS by weighting its alignment with measures of partisan media consumption computed through survey data (RQ1). By means of a Principal Component Analysis (PCA) run on both MP-MPAS and ITANES data, we mapped partisan communities and media sources (see figures 2 and 3). Along with many similarities between the output of the two methods, the comparison also points out some differences. The most substantial is the different proportion of variance explained by the two principal components (62.6% in the case of MP-MPAS and 87.5% in the case of ITANES). The differences in the explained variance may be caused by the different numerosity of the two samples ($N = $ 634 news sources in the case of MP-MPAS and $N = 11$ in the case of ITANES survey), which may influence the structural complexity of the data.

Besides the differences, the political communities actually gather together according to their ideological proximity in both the models, resembling the political

---

[6] A commonly used approach used to further investigate a statistically significant chi-square test result (Sharpe, 2015).

coalitions that have run for the national election. Indeed, Pearson's correlation coefficient calculated between the first principal components of MP-MPAS and ITANES data is 0.95 ($p < .01$) and the one calculated between the second principal components is 0.97 ($p < .01$).

The PCA identified two principal dimensions (fig. 2): the first reflects the traditional left-to-right ideological scale, while the second opposes M5S to all other parties.

Mapping the eleven newspapers (fig. 3) across the same dimensions, we can observe that in both visualizations *la Repubblica*, *Avvenire*, *La Stampa* and *Corriere della Sera* appear in the top left quadrant and *il Fatto Quotidiano* in the bottom left one. *Il Giornale*, *Libero*, *QN - Il Resto del Carlino*, *Il Messaggero* and *Leggo* all appear, although with differences on the vertical axis, in the right side of the chart. Finally, *Il Sole 24 Ore* appears almost at the centre of the MP-MPAS visualization, it is instead positioned in the bottom right quadrant of ITANES map, with relatively low scores on both dimensions.

To get a more in-depth understanding of each media outlets, we further compared the political leaning attributed to each newspaper by MP-MPAS and ITANES method. Both similarities and differences appeared (see tab. 2). A substantial agreement emerged concerning *il Fatto Quotidiano*, adjudicated to the M5S's partisan community by both methods, as well as *Il Messaggero*, classified as cross-partisan, and *Il Giornale*, *Libero* and *Il Resto del Carlino*, adjudicated to the

Figure 2. PCA of MP-MPAS (N = 634 observations) and ITANES data (N = 11 observations)
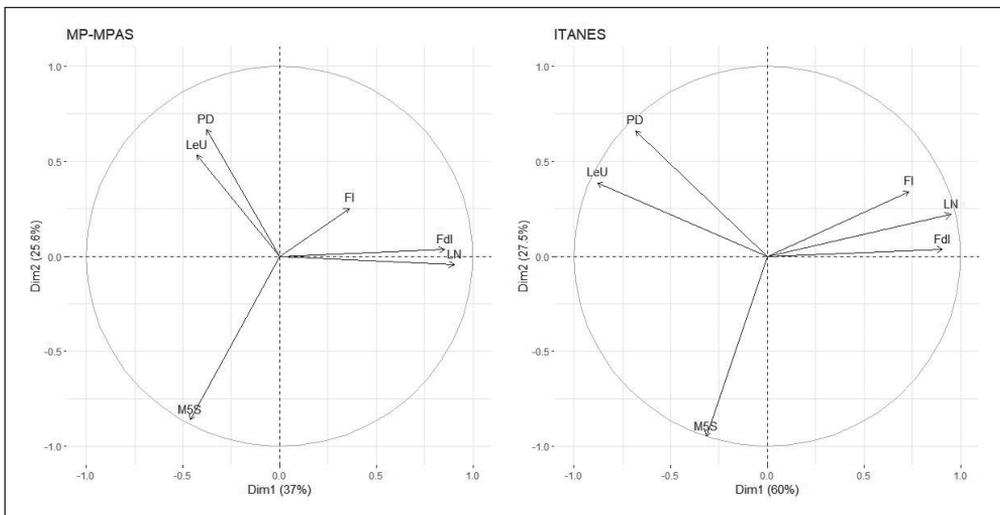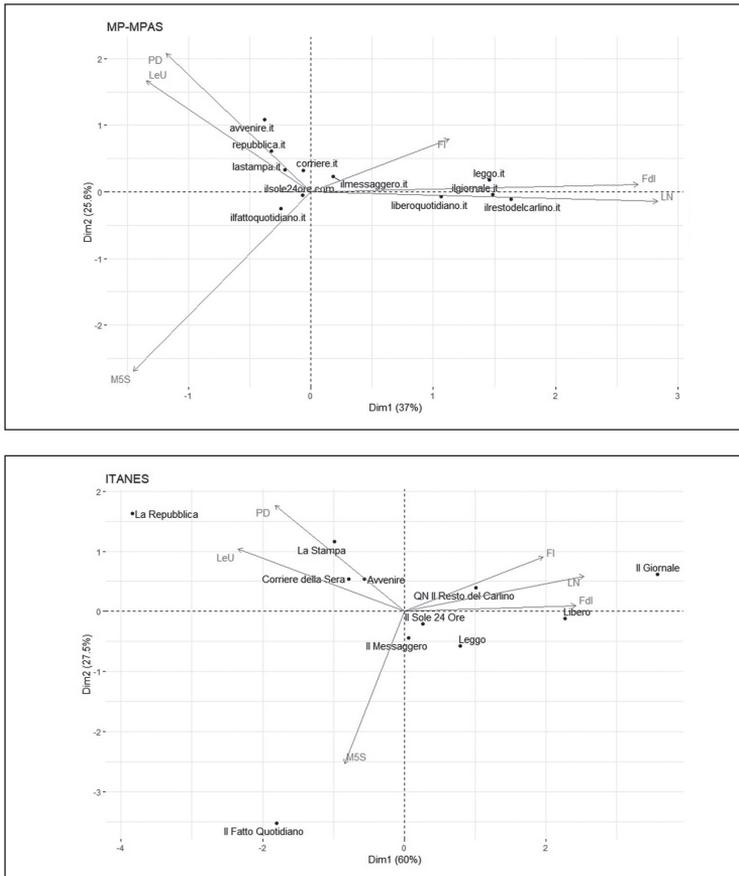
Figure 3. The map of political parties and newspapers resulted from PCA.

League (see tables 2 and 3). Conversely, some differences arose between the other newspapers adjudication.

According to MP-MPAS data, *Avvenire* is a cross-partisan news source, while according to ITANES data it is read by those who express a vote intention for Forza Italia (tab. 3). However, the data available for this newspaper are based on a too small sample to perform a chi-square test and draw reliable conclusions (see tab. 1).

According to MP-MPAS data, *Corriere della Sera* is a cross-partisan news source, while according to ITANES data it is prominently read by those who express a vote intention for Forza Italia more than expected by chance. Nevertheless, the difference between the expected and actual proportions of Forza Italia's readers for this newspaper is not so large, since it is significant only at the .05 level (tab. 3).

Table 2. Adjudication based on MP-MPAS score and ITANES data

Newspaper	MP-MPAS	ITANES
Avvenire	Cross-Partisan	FI*
Corriere della Sera	Cross-Partisan	FI*
il Fatto Quotidiano	M5S	M5S***
Il Giornale	LN	LN***
Il Messaggero	Cross-Partisan	Cross-Partisan
Il Sole 24ore	Cross-Partisan	LN**
Leggo	LN	Cross-Partisan
Libero	LN	LN***
QN - Il Resto del Carlino	LN	LN*
la Repubblica	Cross-Partisan	PD***
La Stampa	Cross-Partisan	PD**

Statistical significance (Bonferroni correction): * *p* < .05; ** *p* < .01; *** *p* < .001

According to MP-MPAS data, *Il Sole 24 Ore* is a cross-partisan news source, while according to ITANES data it is read by those who express a vote intention for the League more than expected by chance (tab. 3). In this regard, it is noteworthy that the website of this newspaper – as well as those of many other major news media outlets – is associated with multiple subdomains (i.e. infodata. ilsole24ore.com, alleyoop.ilsole24ore.com, etc.). In all these cases, we benchmarked MP-MPAS results against ITANES data by using the primary domain (i.e. www.il-sole24ore.com). However, two out of eight of the *Il Sole 24 Ore* subdomains were identified as close to the League's community.

According to MP-MPAS data, *Leggo* is close to the League's online community, while according to ITANES data it is a cross-partisan source, although the result is not statistically significant (tab. 3). The dissimilarity in the offline and online distribution of the newspaper could be accounted for some of the observed differences in the estimated political leaning. The audience of the print version of *Leggo* may be considered cross-partisan because it is a newspaper freely distributed in public spaces (such as metro and train stations), while its digital version could circulate into more circumscribed political networks.

Finally, both *la Repubblica* and *La Stampa* are cross-partisan news sources according to MP-MPAS data, while according to ITANES are prominently read by those who expressed the intention to vote for the Democratic Party (tab. 3). Besides the difference in the final estimate for *la Repubblica*, Democratic Party's community ranked highest in the MP-MPAS distribution too, while with regard to *La Stampa*,

Table 3. Chi-square adjusted standardized residuals. Highest score in bold

|	Avvenire	Corriere della Sera	il Fatto Quotidiano	Il Giornale	Il Messaggero	Il Sole 24 Ore	Leggo	Libero	QN – Resto del Carlino	la Repubblica	La Stampa
LeU	0.22	–1.98	–3.27**	–4.24***	–2.29	–2.74*	–2.41	–3.35**	–3.07*	11.69***	–0.04
PD	–1.17	2.04	–11.43***	–6.45***	–2.23	–3.05*	–2.48	–4.24***	–0.37	**13.73***	3.54**
M5S	–0.78	–3.64**	**21.14***	–5.02***	0.69	–0.45	0.41	–2.06	–2.30	–7.61***	–4.55***
FI	**2.99***	**2.88***	–6.51***	7.99***	**2.41**	2.26	**2.21**	1.53	2.16	–5.56***	–0.06
LN	0.19	1.94	–6.07***	**9.13***	0.74	**3.42***	1.81	**7.41***	**2.78***	–8.11***	1.35
FdI	–0.81	–0.28	–1.66	4.04***	1.24	2.14	1.22	3.84***	2.54	–5.18***	1.23

Statistical significance (Bonferroni correction): * $p < .05$; ** $p < .01$; *** $p < .001$

even though the Democratic Party's MP-MPAS loading is one of the highest, it follows a short distance behind the League and M5S.

Based on all the reported evidence, we concluded that there is a moderate overlap between survey-based estimates of news media partisan alignment and estimates of partisan attention based on Twitter data (RQ1).

Moving to the second research question on the ideological insularity of the Italian online news media system (RQ2) we found a nuanced scenario. The insularity distribution (fig. 4) is positively skewed (.35) thus slightly oriented toward the low insularity side. Consequently, the number of news sources characterized by an insularity score that falls below the average exceeded those above it (about 53% vs 47%). However, we also observed that the minimum (.33), the mean (.65) and the median (.64) are relatively high, indicating that the distribution is shifted toward the high insularity side. Furthermore, the process of adjudication resulted in an uneven distribution of news sources per partisan community (see tab. 5). Over half of the news sources has been indeed adjudicated to one of the two populist parties. The prominent online activism of the League online community on Twitter also affected the low number of news sources adjudicated to the other members of the centre-right

Figure 4. Histogram of insularity scores
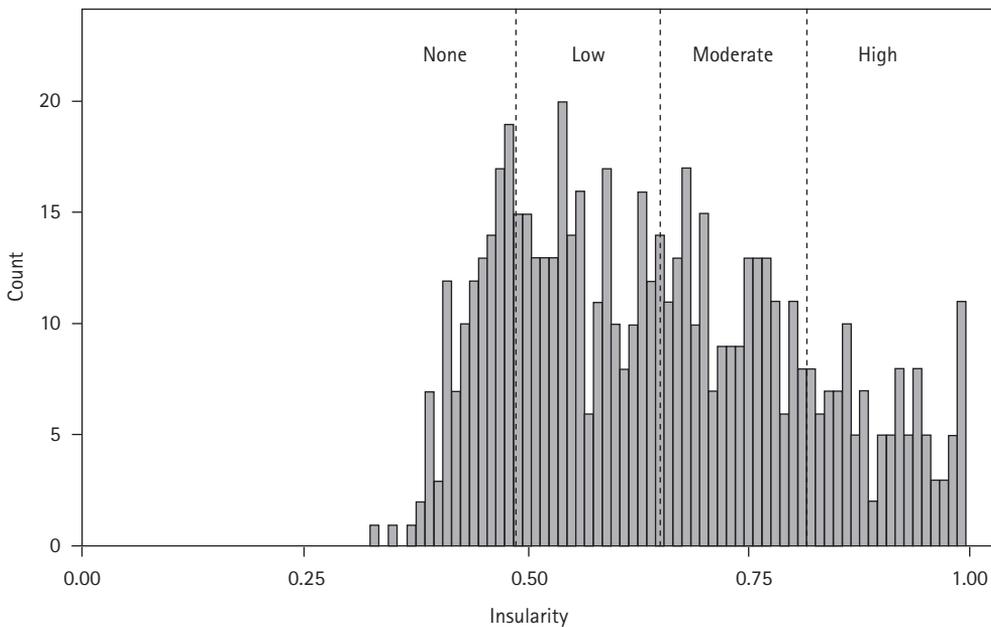
Table 4. News sources by insularity classes

Insularity class	N	%
High	116	18.3
Moderate	181	28.5
Low	216	34.1
None	121	19.1
All	*634*	*100*

Table 5. Results of the adjudication process

	Media sources adjudicated	%
League (LN)	215	33.9
Five Star Movement (M5S)	177	27.9
Cross-partisan	121	19.1
Democratic Party (PD)	60	9.5
Free and Equal (LeU)	49	7.7
Forza Italia (FI)	4	0.6
Power to the People (PP)	4	0.6
+Europe	3	0.5
Insieme	1	0.2
Civica Popolare	0	0
Brotherhood of Italy (FdI)	0	0
All	*634*	*100*

coalition (Forza Italia and Brothers of Italy). A similar role has been played in the centre-left coalition by the Democratic Party.

These observations let us conclude that the Italian online news media system is only moderately characterized by a widespread insularity. While a certain degree of audiences' ideological self-segregation and homogeneity exists, a relatively high number of cross-partisan media sources tend to receive attention by heterogeneous political communities.

That said, the four insularity classes are not evenly distributed (χ^2 (6, N = 501) = 51.877, $p < .001$) among the four political communities with the largest number of adjudicated news media sources (tab. 5). Populist parties' online communities (Five Star Movement and the League) relied on news sources characterized by higher insularity than the other ones (Democratic Party and Free and Equal. See fig. 5).

Figure 5. Insularity class by MP-MPAS adjudication
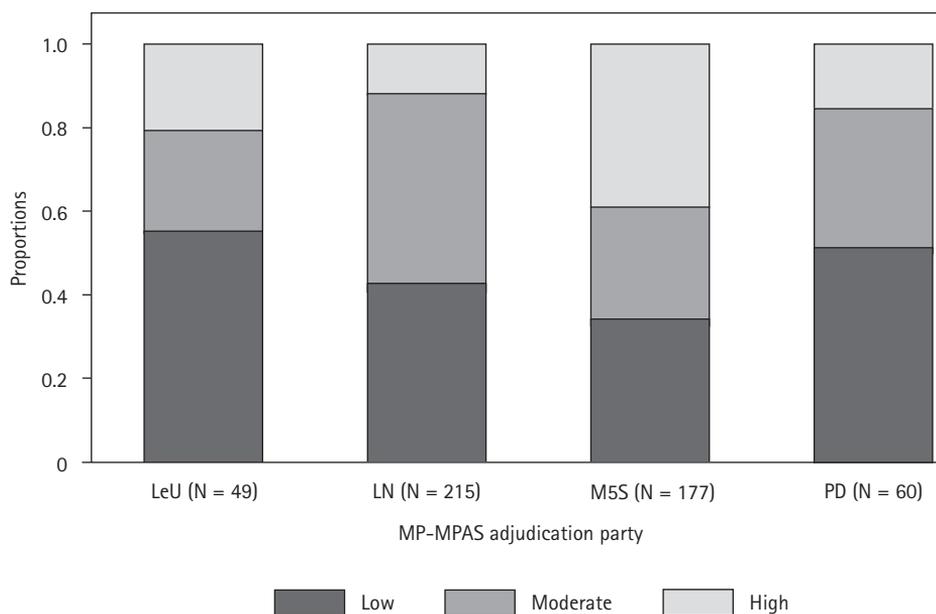
Table 6. Pearson's chi-square standardized residuals

	Low	Moderate	High
LeU	1.94	−1.67	−0.38
LN	0.27	4.16***	−5.07***
M5S	−2.56	−3.04*	6.50***
PD	1.60	−0.34	−1.49

Statistical significance (Bonferroni correction): * $p < .05$; *** $p < .001$

Further analysis of standardized Pearson residual (tab. 6) also points out differences between the two populist parties, insofar news media sources attributed to the Five Star Movement were more concentrated in the «High» insularity class ($p < .001$) and less concentrated in the «Moderate» insularity class ($p < .05$), while those attributed to the League (LN) were more concentrated in the «Moderate» insularity class ($p < .001$) and less concentrated in the «High» insularity class ($p < .001$) than would be expected by chance (Sharpe, 2015).

ComPol

101

5.   Discussion and conclusions

The growing concerns over a potential surge in political and ideological polarization in the context of a fragmented media environment are increasingly leading scholars to concentrate their efforts to analyse how partisan communities are exposed to media.

This paper contributes to this debate by introducing a method to estimate the partisan attention received by a large set of online news media sources in a multi-party political context using Twitter data. Besides presenting the method, we report data suggesting its effectiveness and employ the procedure to assess the partisan attention devoted to 634 online news media sources in the lead-up to 2018 Italian general election.

Assessing the effectiveness of a new method to measure partisan attention toward media sources in a digital media system proved to be challenging. Proper validation of the method would, in fact, require estimates provided by an already established alternative procedure. However, such estimates do not exist for the Italian context. We thus resorted to test the effectiveness of our outcomes against estimates reported by Mancini and Roncarolo (2018) in their analysis of newspaper partisan consumption during the 2018 Italian general election.

We are well aware of the limitations brought by a comparison drawn between procedures that are not only methodologically different but also aimed at a different goal in a very different context. Claiming to read a newspaper is, in the first place, something different than sharing on Twitter an article from the online website of the corresponding newspaper. However, both these actions underline a certain level of exposure to the contents produced by that source. Furthermore, Twitter is by no means representative of the general population and part of the activity we observed have been performed by social media bots (Bessi Ferrara, 2016). The context studied is also very different in terms of the wider level of online media fragmentation.

That said and considering the aforementioned limitations, the similarity between the outcomes of the two methods on eleven major Italian newspapers are remarkable. Results of the principal component analysis surface two main dimensions that, given the high level of correlation between the two sets of scores, point out a shared set of features that link partisan communities and news outlets both in digital and traditional contexts. Furthermore, while the first dimension represents a traditional left-to-right ideological scale, the second describes a complementary dimension insofar it counters M5S, a self-proclaimed post-ideological movement that openly rejects to run as part of an electoral coalition with other actors, to all other parties.

In light of these and other findings concerning effectiveness reported in the previous section, we confidently proceeded to use MP-MPAS scores and related insularity measure to observe the distribution of the attention devoted by partisan and cross-partisan communities to a set of 634 online news media in the lead-up of the Italian 2018 general election.

Unlike what we expected, insularity is not a pervasive feature within the system. Highly insular news sources account for less than 20% of the total, while cross-partisan media sources account for about 20%. The remaining news media are evenly distributed between the «Low» and the «Moderate» insularity classes. Besides the strict realm of the research question at stake, we should probably be careful in drawing conclusions based on an entirely new measure that was never applied before to study other online media systems or the same system in another period of time. Without a proper comparative approach (either in a cross-country or longitudinal perspective), our attempt to characterize the Italian online news system is inevitably inconclusive. For this reason, and with the aim of fostering such comparative studies, we decided to publicly release the dataset of scores and measures discussed in this paper.

While it is impossible to draw conclusive results on the system itself, interesting differences arise when we observe the distribution of insularity among news media sources adjudicated to certain parties. As pointed out by the results reported in the previous section, insularity is significantly higher in news media sources prominently used by the Five Star Movement and, at a less extent, in those prominently used by the League. In other terms, the online communities of the two parties often described as populists, tend to rely on exclusive informative sources (sources that only their community tend to rely on) more than other parties. On the other hand, both communities also seem to rely on (and thus being exposed to) cross-partisan news sources.

Fabio Giglietto
Università di Urbino Carlo Bo
Via Aurelio Saffi 15, 61029 Urbino
E-mail: fabio.giglietto@uniurb.it

Nicola Righetti
Università di Urbino Carlo Bo
Via Aurelio Saffi 15, 61029 Urbino
E-mail: nicola.righetti@uniurb.it

Giada Marino
Università di Urbino Carlo Bo
Via Aurelio Saffi 15, 61029 Urbino
E-mail: giada.marino@uniurb.it

Luca Rossi
IT University of Copenhagen
Rued Langgaards Vej 7, 2300 København
E-mail: lucr@itu.dk

References

Bakshy, E., Messing, S. and Adamic, L.A. (2015a). Exposure to Ideologically Diverse News and Opinion on Facebook. *Science*, 348 (6239), 1130-1132.

Bakshy, E., Messing, S. and Adamic, L. (2015b). *Replication Data for: Exposure to Ideologically Diverse News and Opinion on Facebook* [Data set]. Harvard Dataverse, https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/LDJ7MS.

Barberá, P. (2015). Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data. *Political Analysis*, 23 (1), 76-91.

Benkler, Y., Faris, R. and Roberts, H. (2018). *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. New York: Oxford University Press.

Bessi, A. and Ferrara, E. (2016). Social Bots Distort the 2016 U.S. Presidential Election Online Discussion. *First Monday*, 21(11), https://doi.org/10.5210/fm.v21i11.7090.

Bhat, P. (2018). Advertisements in the Age of Hyper-Partisan Media, in Robert E. Gutsche Jr. (ed.) *The Trump Presidency, Journalism, and Democracy*. New York: Routledge.

Chadwick, A. (2017). *The Hybrid Media System: Politics and Power*. New York: Oxford University Press.

Dahl, R.A. (1998). *On Democracy*. New Haven, CT-London: Yale University Press.

Dvir-Gvirsman, S. (2017). Media Audience Homophily: Partisan Websites, Audience Identity and Polarization Processes. *New Media & Society*, 19 (7), 1072-1091.

Efron, M. (2004). The Liberal Media and Right-wing conspiracies: Using Co-citation Information to Estimate Political Orientation in Web Documents, in *CIKM '04. Proceedings of the Thirteenth ACM International Conference on Information and Knowledge Management* (pp. 390-398). New York: ACM.

Faris, R., Roberts, H., Etling, B., Bourassa, N., Zuckerman, E. and Benkler, Y. (2017). *Partisanship, Propaganda, and Disinformation: Online Media and the 2016 US Presidential Election*. Berkman Klein Center Research Publication 2017-6, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3019414.

Festinger, L. (1962). *A Theory of Cognitive Dissonance* (Vol. 2). Stanford, CA: Stanford University Press.

Flaxman, S., Goel, S. and Rao, J.M. (2013). *Ideological Segregation and the Effects of Social Media on News Consumption*, https://bfi.uchicago.edu/sites/default/files/research/flaxman_goel_rao_onlinenews.pdf.

Garrett, R.K. and Stroud, N.J. (2014). Partisan Paths to Exposure Diversity: Differences in Pro- and Counterattitudinal News Consumption. *Journal of Communication*, 64 (4), 680-701.

Gentzkow, M. (2006). Television and Voter Turnout. *The Quarterly Journal of Economics*, 121 (3), 931-972.

Gentzkow, M. and Shapiro, J.M. (2010). What Drives Media Slant? Evidence From U.S. Daily Newspapers. *Econometrica*, 78 (1), 35-71.

Giglietto, F. (2018). *Observations of Facebook Engagement Around Italian Political News* [Data set]. Harvard Dataverse, https://doi.org/10.7910/DVN/RTDS4M.

Groseclose, T. and Milyo, J. (2005). A Measure of Media Bias. *The Quarterly Journal of Economics*, 120 (4), 1191-1237.

Haenschen, K. (2019). Self-Reported Versus Digitally Recorded: Measuring Political Activity on Facebook. *Social Science Computer Review*, January, https://journals.sagepub.com/doi/10.1177/0894439318813586.

Hallin, D.C. and Mancini, P. (2004). *Comparing Media Systems: Three Models of Media and Politics*. Cambridge: Cambridge University Press.

Ho, D.E. and Quinn, K.M. (2008). Measuring Explicit Political Positions of Media. *Quarterly Journal of Political Science*, 3, 353-377.

Holbert, R., Garrett, R.K. and Gleason, L. (2010). A New Era of Minimal Effects? A Response to Bennett and Iyengar. *Journal of Communication*, 60, 15-34.

Hollander, B.A. (2008). Tuning Out or Tuning Elsewhere? Partisanship, Polarization, and Media Migration from 1998 to 2006. *Journalism & Mass Communication Quarterly*, 85 (1), 23-40.

Iyengar, S. and Hahn, K.S. (2009). Red Media, Blue Media: Evidence of Ideological Selectivity in Media Use. *Journal of Communication*, 59 (1), 19-39.

Jack, C. (2017). Lexicon of Lies: Terms for Problematic Information. *Data & Society*, 3.

Laver, M., Benoit, K. and Garry, J. (2003). Extracting Policy Positions from Political Texts Using Words as Data. *American Political Science Review*, 97 (2), 311-331.

Layman, G.C., Carsey, T.M. and Horowitz, J.M. (2006). Party Polarization in American Politics: Characteristics, Causes, and Consequences. *Annual Review of Political Science*, 9, 83-110.

Lazarsfeld, P.F., Berelson, B. and Gaudet, H. (1968). *The People's Choice: How the Voter Makes up His Mind in a Presidential Campaign*. New York: Columbia University Press.

Lin, F. and Cohen, W.W. (2008). The MultiRank Bootstrap Algorithm: Self-Supervised Political Blog Classification and Ranking Using Semi-Supervised Link Classification,

in E. Adar, M. Hurst, T. Finin, N. Glance, N. Nicolov and B. Tseng (eds.) *Proceedings of the Second International Conference on Weblogs and Social Media.* Menlo Park, CA: AAAI Press.

Lin, W.H., Xing, E. and Hauptmann, A. (2008). A Joint Topic and Perspective Model for Ideological Discourse, in W. Daelemans, B. Goethals, and K. Morik (eds.) *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. Lecture Notes in Computer Science*, vol. 5112 (pp. 17-32). Berlin: Springer.

Lizzi, R. and Pritoni, A. (2014). Sistema degli interessi e sistema mediatico in Italia attraverso un'indagine sulla stampa nazionale (1992-2013). *Comunicazione Politica*, 2, 287-312.

Mancini, P. and Roncarolo, F. (2018). Tanto tuonò che piovve: la campagna elettorale nei giornali e in televisione, in ITANES (ed.) *Vox Populi. Il voto ad alta voce del 2018* (pp. 47-62). Bologna: Il Mulino.

McNair, B. (2017). After Objectivity? Schudson's Sociology of Journalism in the Era of Post-factuality. *Journalism Studies*, 18 (10), 1318-1333.

Messing, S. and Westwood, S.J. (2014). Selective Exposure in the Age of Social Media: Endorsements Trump Partisan Source Affiliation When Selecting News Online. *Communication Research*, 41 (8), 1042-1063.

Ming Fai Wong, F., Wei Tan, C., Sen, S. and Chiang, M. (2016). Quantifying Political Leaning from Tweets, Retweets, and Retweeters. *IEEE Transactions on Knowledge and Data Engineering*, 28 (8), 2158-2172.

Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D. and Nielsen, R.K. (2017). *Reuters Institute Digital News Report 2017*, https://ssrn.com/abstract=3026082.

Pariser, E. (2011). *The Filter Bubble: What the Internet Is Hiding from You*. London: Penguin.

Pew Research Center (2018). *News Media and Political Attitudes in Italy*, http://www.pewglobal.org/fact-sheet/news-media-and-political-attitudes-in-italy/.

Prior, M. (2007). *Post-broadcast Democracy: How Media Choice Increases Inequality in Political Involvement and Polarizes Elections*. Cambridge: Cambridge University Press.

Prior, M. (2009). Improving Media Effects Research through Better Measurement of News Exposure. *The Journal of Politics*, 71 (3), 893-908.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Wien: The R Foundation, https://www.R-project.org/.

Rudis B. (2016). *longurl: Expand Short URLs. R package version 0.3.0*, https://CRAN.R-project.org/package=longurl.

Saez-Trumper, D., Castillo, C. and Lalmas, M. (2013). Social Media News Communities: Gatekeeping, Coverage, and Statement Bias., in *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management* (pp. 1679-1684). New York: ACM.

Sharpe, D. (2015). Your Chi-Square Test Is Statistically Significant: Now What?. *Practical Assessment, Research & Evaluation*, 20 (8), 1-10.

Shin, J. and Thorson, K. (2017). Partisan Selective Sharing: The Biased Diffusion of Fact-Checking Messages on Social Media. *The Journal of Communication*, 67 (2), 233-255.

Stroud, N.J. (2010). Polarization and Partisan Selective Exposure. *Journal of Communication*, 60 (3), 556-576.

Sunstein, C.R. (2017). *#Republic: Divided Democracy in the Age of Social Media.* Princeton, NJ: Princeton University Press.

Weeks, B.E., Ardèvol-Abreu, A. and Gil de Zúñiga, H. (2017). Online Influence? Social Media Use, Opinion Leadership, and Political Persuasion. *International Journal of Public Opinion Research*, 29 (2), 214-239.

Yu, B., Kaufmann, S. and Diermeier, D. (2008). Classifying Party Affiliation from Political Speech. *Journal of Information Technology & Politics*, 5 (1), 33-48.

Zhou, D.X., Resnick, P. and Mei, Q. (2011). Classifying the Political Leaning of News Articles and Users from User Votes, in *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media* (pp. 417-424). Menlo Park, CA: AAAI Press.

Zuckerman, E. (2018). *DRAFT Chapter 1: The Necessity for Trust.* Washington, DC: Knight Commission on Trust, Media and Democracy, https://medium.com/trust-media-and-democracy/renewing-americans-trust-b16a63eb5e63.