News by Popular Demand: Ideology, Reputation, and Issue Attention in Social Media News Sharing

Natalia Aruguete[*] Ernesto Calvo [†] Tiago Ventura [‡]

October 14, 2019

Abstract

Theories of news sharing behavior and editorial gatekeeping are central to current studies in Political Communication. Thereby, deriving theoretically sound parameters and developing testable implications using social media data is central to the literature. This is also an area were Computational Social Science (CSS) can greatly contribute to theory building and testing. In this article, we describe a CSS method that contributes to our theoretical understanding of news sharing and editorial gatekeeping. We model news sharing as a function of three behavioral parameters (cognitive congruence, media reputation, and user's attention to issues) with observational social network data as input. More important, we introduce readers to an extension of our model that allows us to assess the level of congruence between users and media offerings, contributing to the literature on gatekeeping. We show intuitive and interesting results from #Bolsonaro in Brazil, #Maldonado in Argentina, and the #TravelBan in the US.

Number of words: 7,750

Keywords: News Sharing, Gatekeeping, Social Media

[*]UNQ, Conicet

[†]GVPT, University of Maryland, Corresponding author: ecalvo@umd.edu

[‡]GVPT, University of Maryland

News by Popular Demand:

Ideology, Reputation, and Issue Attention in Social Media News Sharing

Why do social media users share links to news articles? How important are ideological considerations, the reputation of a news organization, and the attention of users to particular issues? Theories of news sharing behavior and editorial gatekeeping are central to current studies in Political Communication. Thereby, deriving theoretically sound parameters and developing testable implications using social media data is central to the literature. This is also an area were Computational Social Science (CSS) can greatly contribute to theory building and testing. In this article, we describe a CSS method that contributes to our theoretical understanding of news sharing and editorial gatekeeping. We model news sharing as a function of three behavioral parameters (cognitive congruence, media reputation, and user's attention to issues) with observational social network data as input. More important, we introduce readers to an extension of our model that assesses the level of congruence between users and media offerings, contributing to the literature on gatekeeping. We show intuitive and interesting results using data from #Bolsonaro in Brazil, #Maldonado in Argentina, and the #TravelBan in the US.

Our study on news sharing and gatekeeping takes as input a matrix of social media embeds and deliver estimates of the importance of cognitive congruence, media reputation, and issue attention in news sharing. Our estimates allow researchers to (i) understand news consumption in different regions of a social network; (ii) decompose the demand for content revealed by users; and (iii) evaluate gatekeeping incentives through estimates of the optimal editorial line if news organizations were solely interested in maximizing readership.[1] All three contributions are theoretically informed and of substantive interest to students of the relationship between news consumption and social media.

Our research contributes to the growing field of news sharing (Thurman et al., 2019; García-Perdomo

[1] Understanding the "optimal" editorial line of news organizations is critical if we want to compare the level of congruence of a media outlet with the public. While assuming that editors may simply maximize readers is unrealistic, this is a required baseline to understand differences between editorial choices and news demands by readers.

et al., 2018; Arendt et al., 2016; Kümpel et al., 2015; Bright, 2016; Boczkowski et al., 2018), providing a valuable tool to derive sharing behavior from aggregate level data. The literature on news sharing is broad, considering subjective, social, rational, and emotional factors that explain sharing behavior. As described by Kümpel (2019) and Boehmer and Tandoc Jr (2015), individual's sharing behavior can be explained by factors such as the trustworthiness of the source (Suh et al., 2010; Wang et al., 2012), the content of the message (Macskassy and Michelson, 2011), the attention to issues (Boyd et al., 2010; Rudat et al., 2014), and the linkages between users and their peer groups. Empirical testing, however, requires a statistical model to extract informative parameters from observational data. Results of our study also inform recent debates on gatekeeping behavior, assessing congruence between media organizations and users in different regions of a social media network.

1. A model of News embedding in Social Media

In a recent article, Kümpel et al. (2015) conduct a meta-analysis of 461 articles on *news sharing* published between 2004 and 2014. The authors show that articles on news sharing published every year in peer-reviewed journals increased from 10 in 2004-2005 to over two hundred in 2013-2014. Indeed, with the rise of social media, news sharing and news sharing behavior have become a central topic in the communication's literature, key to our understanding of news consumption and editorial gatekeeping. In complex and user centered digital environments, concepts such as *novel* or *newsworthy* are increasingly unable to model the decisions by editors or the consumption patterns of users (Shoemaker and Reese, 2013). As Kumpel, Karnowski, and Keyling note, "online news sites increasingly rely on [*news sharing*] referrals from social media to improve their website traffic, article views, and ultimately their economic success (Kümpel et.al. 2015:1).[2]

A recent survey of news editors by Hong Tien Vu called attention to this issue, finding that "editors are willing to adjust their editorial decisions based on web metrics." (Vu, 2014) So far, however, we lack a statistical strategy to model news sharing behavior in observational data. We also lack a strategy to

[2]Journalists are increasingly encroached by users that also deliver news. The delivery of online content, therefore, is openly challenging journalistic monopoly over the news gates (Waisbord, 2017).

test whether editors respond to demands by users.

Our answer is to model news sharing behavior from observational data. First, we estimate three key behavioral parameters that explain news embedding: ideological congruence, reputation, and issue attention. Then, we explain how these parameters can be used to evaluate the demand for news of a particular ideological color. Finally, we estimate the optimal editorial line of a news organization *if* the editor is solely interested in maximizing readership. This unrealistic assumption allows us to compare the editorial line of an organization against the observed consumption patters of users. Our results not only inform on the existence of a social media news gap (Boczkowski and Mitchelstein, 2013; Bright, 2016), but also describe the news sharing behavior that explains this gap in different regions of a social media network.

Our point of departure is the assumption that readership matters. This is far from being universally accepted in political communication, as editorial choices are also explained by professional considerations, media owners' preferences, editorial expertise, and vendors' demands. The micro and macro factors that explain news content are well described by the *hierarchy of influences* model (Shoemaker and Reese, 1996). Different from White's approach (1950), Shoemaker et al. (2001) describe the effect of work routines on gatekeeping behavior (Shoemaker et.al, 2001: p. 235). The reader-seeking orientation of editors, however, has not been formalized in the existing literature.

The assumption that what readers think should be of importance is often downplayed in journalism and in professional training, with journalistic integrity and consumer demand often perceived as being in conflict with each other. In the past, control over the flow of information was the undisputed prerrogative of traditional news media organizations (Bagdikian, 1983; Galtung, 1971; Hardt, 1979). Prior to the rise of online journalism, editorial decisions were centralized by journalists and expected to be insensitive to audience effects (Gans, 2004).

However, the use of news dashboards and media monitoring tools, ubiquitous in most news rooms, shows that media organizations are attentive to the consumption patterns of readers, reinforcing the findings in Vu (2014). There is value in monitoring users' demands for news and the performance of journalists, journal articles, and media products. While editors and journalists hope to defend their role as gatekeepers and to defend their autonomy to define news-worthiness (Domingo, 2008; Singer et al., 2011), readership increasingly conditions what is newsworthy. The click-culture and the agenda-audience (Anderson, 2011) have transformed the journalist's routines.

In the last twenty years, scholars have studied how editors adapt their gatekeeping strategies to social media (Barzilai-Nahon, 2008) and how users activate content that is congruent with their beliefs (Aruguete and Calvo, 2018; Entman, 2004).[3] As new technologies feature more prominently in the dissemination of media frames, more attention is now devoted to the decision of readers to view and share content they agree with.[4] These significant advances, however, are hampered by the lack of a statistical model to understand news sharing in social networks. In this article, we seek to fill this gap.

2. Modeling news sharing behavior

Let us begin with an example that describes the importance of our three key parameters of interest: ideology, reputation, and issue attention. Consider a vector of social media users (rows) that embed links to media organizations (columns). Table 1 provides an example with each user $u_i \in I$ sharing news published by media organizations, $m_j \in J$. For presentation purposes, let us assume that media organizations are listed from left to right by conservatism, so that m_1 is less conservative than m_2, and $m_1 < m_2 < ... < m_j$.

In Table 1 we see a higher number of embeds by user u_2, 25, and a lower number of embeds for media

[3]While it is widely acknowledge that journalists monitor media traffic (Domingo, 2008; Lowrey and Woo, 2010; McKenzie et al., 2011), little research explains how media metrics affects the editors decisions and the coverage of news events (Vu, 2014). However, there is little doubt that users are becoming increasingly relevant for explaining news creation, publication, and propagation (Singer et al., 2011).

[4]Vu notices that the hierarchies of influence's model (Shoemaker y Reese, 1996) was created when traditional news organizations where both centralized and hierarchical. At the time, the audience was just one among many extra-media factors that editors could consider. However, in the current digital environment, user behavior and media fragmentation have gain considerable relevance for explaining gatekeeping behavior. Shoemaker and Vos (2009) agree in the need to acknowledge the behavior of digital media readers and their effect on news salience.

Table 1: Observational data with
Users u_i (row) and Media m_j (columns)

$$\begin{array}{c c}
\begin{array}{ccccc} \mathrm{m}_1 & \mathrm{m}_2 & \mathrm{m}_3 & . & \mathrm{m}_J \end{array} & \\
\begin{pmatrix} 0 & 2 & 0 & 4 & 1 \\ 5 & 8 & 2 & 6 & 4 \\ 10 & 5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 5 & 9 \\ . & . & . & . & . \\ 0 & 5 & 0 & 1 & 7 \end{pmatrix} & \begin{array}{c} \mathrm{u}_1 \\ \mathrm{u}_2 \\ \mathrm{u}_3 \\ \mathrm{u}_4 \\ . \\ \mathrm{u}_I \end{array}
\end{array}$$

m_3, 2. Readers may also notice that u_3 is embedding more progressive news while u_4 is embedding more conservative news. Everything else equal, we can think of row means as summary information of the overall attention that a user gives to news hyperlinks and column means as an indicator of the overall prevalence of a media organization in the data. In a statistical model, random intercepts by row and column will describe the mean counts of embeds by each user u_i and for each media m_j. For each user, however, we also expect that they will share a higher proportion of media news from organizations that are ideologically closer. In our statistical model, therefore, a random slope describes the user's taste for ideological congruence.

While we estimate all quantities of interest simultaneously, Figures 1, 2, and 3 provide an intuition of the relationship parsed out by each parameter set. Figure 1 describes a media outlet that is broadly shared by a larger subset of users on the left, center, and right of the political spectrum, in this case *Media B*. Indeed, we expect that higher reputation outlets will be more broadly share by users even if they are relatively distant ideologically. Figure 2 describes a user, "User Left", who is embedding links to media outlets across the ideological spectrum, indicating higher interest or attention on the subject being reported. Figure 3, finally, describes a higher propensity to share links to media that is located in the same location as users "Left", "Center", and "Right". A statistical model of media sharing needs to parse out from Table 1 the marginal effects of the different behavioral choices. In what follows, we use i to index information about the users and j to index information about the media outlets, dropping the "u" and "m" descriptors.

The Model

Let us now describe in greater detail our modeling strategy. We consider a utility function where each social media user i minimizes cognitive dissonance on issue k, given her preferred ideological position, x_i^k, and an editorialized set of news, L_j^k, that is created or published by actor j. We define editorialized news as content that is posted by an organization, with an *ideological charge* that is observed by both the user and the editor. Media actor j may be a news organization, a candidate, a political group, a social media peer, a friend, or any entity that publishes information in social media that is accessed by user i. At this time we only assume that media entities have a separate online page that can be accessed (and shared) through hyperlinks that may be inserted in social media publications. The utility that user i derives from the editorialized news, as we already noted, minimizes cognitive dissonance between the ideological leaning of the news' content and her existing ideological preference.

As important, the utility of user i also increases with the perceived reputation R_j^k of media actor j. That is, users want to minimize cognitive dissonance when reading news, but cognitive congruence is valuable insofar as information is trustworthy. For example, if the National Enquirer and Fox News were to be located in the exact same ideological position, we expect that users will still value the National Enquire lower than Fox News. Therefore, we expect that users will share the latter considerably more than the former. By assumption, the reputation of each media actor j is fixed in the short term, which means that the reputation of an outlet only changes over time and it is not affected by the current news the user shares. Given that news organizations have reputations that are fixed in the short-term, they maximize readership by altering the ideological leaning of the content they publish or the issues they cover.

By assumption, cognitive dissonance is *negative* while reputations are *positive*. That is, users are more likely to share posts that agree with their ideological beliefs and see a declining utility from news that are further removed or openly challenge their beliefs. While it is possible that users share "ironically" news that are further removed from them ideologically, there is no empirical evidence that shows systematic sharing of dissonant content in social media.

Users also receive a *positive* utility for information they agree with if it is published by a reputable news organization. The reputation scores and the ideological leaning of news organizations, as observed by users, may or may not be correlated. A user may perceive that the reputation of a news organization is high because it agrees with its beliefs. Readers of Fox news, for example, may consider that the publications of this organization are of high reputation precisely because it minimizes cognitive dissonance. Readers of the New York Times, on the other hand, may perceive that Fox News is both biased and of low quality because news published by this organization fail to align with the individual's preferences. Other readers, however, may perceive that ideology and reputation are separate dimensions, orthogonal to each other. For example, a conservative reader may perceive that the NYT and Fox News are of high reputation and that the New York Post, while conservative and congruent with her beliefs, is of low reputation. The extent to which ideology and reputation are interrelated is something that we can test for empirically.

Both ideology and reputation are issue-dependent. That is, users may perceive the New York Times as leftist when reading *world news*, but see this same organization as centrist when reading *Real State news*. Readers may also perceive that reputation varies by issue, considering the book reviews of the New Yorker as being of higher reputation than those of the New York Times, even if they do not differ in ideological terms. Therefore, ideological proximity and reputation may vary by issue as well as by organization.

Issues may also be more or less important to readers, who may spend more or less time reading about issue k, A_i^k. [5] Therefore, a Reader i will perceive a utility from "voting" (reading, liking, or

[5] This is what other authors have described as the gate-watching effect of social media on news organizations (Bruns 2005).

sharing news) on issue k by organization j as described in Equation (1):

$$U^k_{(ij)} = -\alpha^k_i \left(x^k_i - L^k_j\right)^2 + A^k_i + R^k_j + \gamma^k_{ij} \tag{1}$$

In Equation (1), the quadratic term $\alpha^k_i \left(x^k_i - L^k_j\right)^2$ describes the disutility of a publication by media j on issue k, with ideological leaning L that is further removed from the reader's preferred ideological position, x^k_i. For every unit of increase in cognitive dissonance, the utility of reader i declines by $-\alpha$. The parameter $-\alpha$ also has a natural interpretation as the weight that a reader attaches to the ideological leaning of a media organization on issue k. When browsing for news about Donald J. Trump, for example, ideology may weigh more heavily on the user's decision to activate content than when browsing news about Justin Bieber, $\alpha^{Bieber}_i < \alpha^{Trump}_i$. As in Gelman et.al. (2004), α, x, and L can be consider latent parameters. We will provide a computationally simpler alternative in the next section, where we input information about x_i and L_j.

Equation (1) also shows that news published by a more reputable actor, R^k_j, increase the utility of reader i. The importance of reputation varies by issue k. For example, reputation may matter more when reading about Donald Trump than when reading about Justin Bieber, $R^{Bieber}_{ij} < R^{Trump}_{ij}$. As we will show, reputation will also vary by the location of users in different regions of a network. Finally, users may also give different attention to an issue, A^k_i, sharing a higher than average number of post with social media peers. Equation (1) also includes an stochastic term that captures overdispersion, γ^k_{ij}, by user and media outlet.

Readers may recognize equation (1) as a multilevel specification with a random slope, α, and two random intercepts, A and R. The random slope captures the weight that readers attach to ideological congruence, while the random intercepts describe the importance of reputation and user attention in regards to issue k.

Sharing news can take many forms, such as reading (clicking), liking, or sharing content. For simplicity, we assume that editors of a news organization measure *success* by the number of times users share the content they publish. We could also consider the success of particular news, a journalist, an organizations, or any other event that we are interested in ranking. Let us for know consider as a dependent variable the number of articles of organization j that are shared by a user, with the choice function described in Equation (2), where the total number of shared news, S^k_{ij}, follows from a multinomial distribution:

$$S^k_{ij} = \tau_i \frac{e^{U^k_{ij}}}{\sum_{j=1}^{J} e^{U^k_{ij}}} \; \forall \; i,j,k \tag{2}$$

Multilevel estimation of the proposed model proceeds as in Zheng et al. (2006), with separate parameters by users (row) and media (column). However, as readers may readily observe, the computational demands of Equations (1) and (2) are very high. In what follows, we provide a few shortcuts that both simplify and enrich the model.

Binning: Reducing computational demands and model noise

A look at equation (1) shows that the proposed model estimates a rather large number of parameters. Even if we have measures of ideological location for each user x_i and for media leaning L^k_j, the total number of parameters still adds to $I * 3 + J$. For a social network with 250,000 nodes and 24 media outlets, therefore, the model estimates a total of 750,024 parameters.

Computational demands can be reduced significantly through binning, collecting model parameters by groups of users. Rather than estimating the ideology, reputation, and attention parameters by row, we bin parameters in equally sized quantiles by the ideological location of users. We then estimate a smaller set of parameters, $-\alpha^k_{q(i)}$, $R^k_{q(i),j}$, $A^k_{q(i)}$.

$$U^k_{(ij)} = -\alpha^k_{q(i)} \left(x^k_i - L^k_j \right)^2 + A^k_{q(i)} + R^k_{q(i),j} + \gamma^k_{ij} \tag{3}$$

Consider Figure 4, which describes 10 equally sized bins that collect the ideological preferences of

160 thousands high activity users from the *Bolsonaro* network [6]. First, we derive a location in the social media network for each user, collecting the first dimension of the Fruchterman-Reingold layout estimated in igraph 1.1 (Csardi and Nepusz, 2006). For the estimation of the user locations, we consider the full set of 2.9 million retweets. We then create a grouping index variable by equally sized quantiles, which is used to estimate random intercepts and slopes for each of group.

The new set of random slopes and intercepts, $-\alpha^k_{q(i)}$, $R^k_{q(i),j}$, $A^k_{q(i)}$, now describe the differences in ideology, reputation, and attention that are observed in different regions of the social media network. Therefore, we simplify estimation of the model and are still able to describe local differences in the importance that voters attach to each of the components of their sharing behavior.

Summarizing the weight of ideology, reputation, and attention on online media

After we implement the model described above, we are left with three sets of parameters that describe the importance that users attach to ideology, reputation, and attention. We may consider these parameters as exercising a "pull" on editors to accommodate the preferences of readers. Of course, some media outlets may give scant attention to the preferences of readers while other outlets may care deeply about them. Our model provides information to evaluate the extent to which the preferences of readers and those of news organizations truly align with each other.

First, let us reweight the vector of user preferences in social media by the parameters retrieved from the model. Each parameter set by quantile q can be interpreted as the weight that users give to each determinant of their sharing behavior. Some users may be very attentive to events related to issue k. Other users may care deeply about the reputation of the media outlets they embed. Meanwhile, other users may only share content that is coming from outlets they are closely aligned with them ideologically. We may re-weight the location of the median voters by the three sets of parameter estimates to determine which groups want to talk about the issue k, which group cares more about

[6] User preferences were estimated using the full set of retweets from the last week of campaigning in Brazil. Figure 1 describes the first dimension of the Fructherman-Reingold layout algorithm estimated via Igraph/R 1.0.

reputation, and which group cares more intensely about ideological congruence:

$$\text{IdeologyShift} = \frac{\sum x_i \alpha_{f(i)}}{\sum \alpha_{f(i)}} \tag{4}$$

$$\text{ReputationShift} = \bar{x}_i - \frac{\sum x_i R_{f(i)}}{\sum R_{f(i)}} \tag{5}$$

$$\text{AttentionShift} = \bar{x}_i - \frac{\sum x_i A_{f(i)}}{\sum A_{f(i)}} \tag{6}$$

Equations (4), (5), and (6) summarize the ideological shift on voters that results from the group specific weights that users attach to ideological congruence. We use the same strategy to reputation and attention, respectively. In the following section we implement our measurement strategy on three data sets from Argentina, Brazil, and the United States.

3. Three Social Media Events: #Bolsonaro, #Maldonado, and #TravelBan

We provide evidence of the usefulness of the proposed model considering three different social media events in Brazil, Argentina, and the United States. All three events took place in deeply divided political contexts and garnered significant political attention. In all three cases we have a larger showing by users with more progressive leanings, who are protesting against right-wing shift in the status quo.

First, we consider the Bolsonaro network in Brazil, using 2,943,993 tweets published by 162,107 high activity accounts on the week prior to the election of Jair Bolsonaro as President of Brazil, from September 26 through October 02, 2018. Bolsonaro is widely considered as a fringe right-wing candidate, who has stacked his administration with military officers, celebrated the use of torture by the 1964-1985 military regime, and introduced extreme legislation to reverse social policies in areas such as LGBT rights and welfare insurance. Jair Bolsonaro has been an extremely divisive political figure and, more important for this research, used a vast network of intelligence and "fake news mills" to support his presidential candidacy.

Figure 5 describes the basic layout of Bolsonaro, with pro-Bolsonaro users in Blue and anti-Bolsonaro users in red. Of the more than 2.9 million retweets analyzed in the data, 432,591 (14.7%) included hyperlinks. The number declines to 387,841 (13.2%) if we do not consider hyperlinks directed to other tweets. The most frequent news outlet embedded in the data is the pro-Bolsonaro *Oantagonista*, which represents 63,862 links, 14.7% of all hyperlinks.

In the case of the Maldonado in Argentina, we analyze 5,325,240 tweets posted by 196,066 high activity accounts in the 78 days that followed the disappearance of activist Santiago Maldonado. The disappearance of Maldonado was a deeply polarizing event. Different media outlets aligned for and against the government, which the opposition portrait as responsible. Of the more than 5 million retweets analyzed in the data, 816,694 (15.3%) included hyperlinks. The number declines to 513,659 (9.6%) if we eliminate hyperlinks to other tweets.

In the case of the TravelBan in the US, we analyze 2,031,518 retweets from 241,271 high activity accounts on January 30 and 31, following the decision of the Trump administration to restrict travel from seven majority Muslim countries. On January 30, Trump tweeted "Only 109 people out of 325,000 were detained and held for questioning. Big problems at airports were caused by Delta computer outage" and criticized democrats with a second tweet, "protesters and the tears of Senator Chuck Schumer. Secretary John F. Kelly said that all is going well with very few problems. MAKE AMERICA SAFE AGAIN!".

Figure 7 describes the basic layout of the TravelBan network, with pro-TravelBan users in Blue and anti-TravelBan users in red. Of the more than 2 million retweets analyzed in the data, 641,719 (31%) included hyperlinks. The number declines to, 485,560 (23.9%) if we do not consider hyperlinks directed to other tweets.

For each of the three networks we retrieve the matrix of users (rows) and media organizations (rows), keeping only the 24 most frequently embedded news outlets. We also retrieve the first dimension value for each user (horizontal axes in Figures 5, 6, and 7), as a proxy for x_i^k in equation 3, the quantile indices $q(i)$, and media locations, L_j^k.

As readers may have noticed, there are significant differences in the share of retweets that direct users to media outlets, from a high of 31% in the #TravelBan to a low of 9.6% in #Maldonado. There are also significant differences in embbeds within each of these networks, as we discuss next.

A visual inspection of Media Embeddings

Figures 8, 9, and 10 present 24 plots that describe the areas of the network that are activated by the top eight media outlets in #Bolsonaro, #Maldonado, and #TravelBan. The other 48 media outlets can be found in the online supplemental file. In all three of our cases, red nodes describe users that are on the left of the political spectrum while blue nodes describe users on the right of the political spectrum.

At first sight readers may notice that *Oantagonista* and *Conexaopolitica* are on the right of the political spectrum in Brazil. The former was recently founded by three conservative journalists that abandoned the weekly news magazine *Veja*. Meanwhile, *Folha*, *Veja*, and *Globo* are more readily embedded by users on the center and center-left of the political spectrum. Readers may also note a much larger share of links to Twitter on the left and more frequent links to youtube on the right of the political spectrum. Such is the result of the decision by Twitter to suspend accounts from the conservative group Movimiento Libre Brazil, who engaged from within youtube on a very active campaign of missinformation. The decision by Twitter was mirrored by Facebook, who suspended over 100,000 WhatsApp accounts on what is without a doubt the largest astroturfing campaign in any election in the region.

In Figure 9, readers can see some Argentine outlets that are clearly on the left (Página/12) or right (TN) of the political spectrum. However, other outlets such as La Nacion are embedded by most of the conservative users but also by a significant number of moderates on the center of the political spectrum. Finally, in the case of the #TravelBan, Figure 10, we see outlets that have a higher than average readership on the left of the political spectrum (New York Times) as well as those with a wider right leaning readership (Fox News).

More important for our purpose, all 24 plots provide evidence of significant variation within and across networks, with some media outlets being more widely shared by all users, some media outlets more intensely shared by users in a particular region of the network, as well as other users more actively sharing links on the #Bolsonaro, #Maldonado, and the #TravelBan networks.

Ideological congruence, Attention, and Reputation

Figures 11 and 12 provide a visual comparison of the ideology and attention parameter estimates for our three cases. As it is possible to observe, there is variation within and across cases, with users in different quantiles having a different taste for ideological congruence and attention. Figure 11 shows that, in all three cases, ideology matters more for users on the right of the political spectrum. Because ideological distance is cognitively costly, larger negative values indicate that ideology matters more. In the particular case of #Maldonado, for example, users in the 8th, 9th, and 10th quantile, which corresponds to the core of the pro-Government sub-network, display large negative estimates, reflecting a high demand for congruent news. As it is also the case with survey data, ideological congruence tends to be more modest for users in the center of the network and it increases centrifugally as we move to the extremes. If we compare the results with similar estimates of ideological congruence in election survey data, we can see that ideological congruence has a significantly larger weight in the sampled networks. In effect, most electoral data provides estimates of ideological congruence that fall in the range of [-.05,-.12], four times smaller than the empirical estimates in the social media data (Calvo and Hellwig, 2011).

Figure 12 provides estimates of attention by quantiles, with larger values indicating a higher propensity to embed links. Consistent with the visual inspection of the data in the previous section, we see that users on the left and right of the political spectrum are more likely to pay attention to #Bolsonaro,

#Maldonado, and the #TravelBan. Particularly interesting is the very high level of attention of users to the right of the political spectrum in #Maldonado, with activity that is orders of magnitude above the rest of the network. Indeed, there were in #Maldonado considerable fewer users on the right of the political spectrum but, as shown in Figure 12, these user more than compensated their low numbers with much higher rates of embed.

Modeling News Congruence with Readers: Who demands what?

In the previous section we showed that users in different locations of the social network have different taste for ideological congruence and attention. Considered jointly with the estimates of reputation, as argued earlier in this paper, we can evaluate the demands of users on media outlets. To this end, equations (5), (6), and (7) return weighted estimates of the median voter that describe the overall shift in demand when we consider the full set of parameters. We may also use Adams, Merrill, and Grofman (2005) numerical estimation algorithm to find the optimal editorial line of a news organization if it only sought to maximize readership. This unrealistic assumption allows us to understand the pressure on news organizations to take particular editorial positions and to observe whether these optimal positions diverge from what we observe in the data.

Figure 13 shows that, if we weigh the preferences of the voters by the *Attention* parameters in #Bolsonaro, the location of the median voter would move from 0 (zero) to -5, that is, to the left of the political spectrum. In other words, results from the #Bolsonaro network shows that voters on the left of the political spectrum are slightly more interested in "talking" about #Bolsonaro. On the other hand, voters on the right of the political spectrum are slightly more sensitive to ideological considerations. In other words, if a news organization publishes content that deviates from the preferred location of users, the decline in embeds is marginally larger among more conservative users. Finally, there is little difference in the demand for reputation among users, which results in a weighted location that is very similar to the original one.

Using Adams, Merrill, and Grofman (2005) we can use these parameters to estimate the optimal ideological placement of the media if they were only interested in maximizing readership. In Figure 13, the point of origin of each horizontal arrow describes the observed location of the media as it was entered into equation (3). The end point of each arrow, on the other hand, describes the optimal ideological location of each media outlet as derived solely from their readership. Shorter arrows, therefore, describe higher levels of congruence between the observed and the optimal location of a media outlet. Larger arrows, on the other hand, describe observed locations that are further away from the consumers that are more interested in embedding their news.

In the case of #Bolsonaro, for example, the optimal location of most media outlets is to the left of its observed location. In other words, most media outlets would gain larger shares of readers if they had shifted coverage in the direction of more leftist users. Very high levels of congruence between readers and news outlets, on the other hand, are observed on the right of the political spectrum, with *Oantagonista*, *Tribunadoceara*, and *Conexaopolitica* placed squarely at the location where readership is maximized. Interestingly, only *Oantagonista* is led by mainstream journalists. Others, such as Conexao Politica, can only be defined as "Fake news mills" that were created for electoral gain. In all, the outlets that are strictly located at their optimal location are generally those that cater to fringe voters on the right and left of the political spectrum.

Figure 15 presents results from the #Maldonado network. In contrast to the #Bolsonaro case, users on the right of the political spectrum are more interested in "talking" about #Maldonado while voters on left of the political spectrum see a sharper decline in embeds when media outlets are further away from them. As in the case of #Bolsonaro, voters are not all that different when we compare how much they care about reputation. Results for the Argentine case show smaller differences between the observed and optimal location of media outlets. Arrows are smallest among news outlets on the left of the political spectrum, such as *TelesurTV*, *Izquierdadiario*, *enoirsai*, and *Pagina/12*. Most news outlets on the center of the political spectrum would gain readership by shifting mildly to the right, with the notable exception of La Nacion, which would benefit from taking on a more centrist position.

Finally, results from the #TravelBan are equally interesting. First, we can see that the overall effect of attention and ideology is more modest than in the previous two cases. As in #Bolsonaro, voters on the left are marginally more interested in talking about the #TravelBan while voters on the right are slightly more ideologically demanding. As in the case of #Bolsonaro, results show that most media outlets would do better by taking an editorial position that is to the left of where they are currently observed. Rightward shifts in social media providers such, as YouTube and Twitter, provide ample evidence of a higher strategic use of the platforms by conservative users. This explains the lower level of congruence between the *observed* and *expected* placement of intermediaries as well as the lower lever of congruence between users and media organizations in the center and left of the political spectrum. Another interesting finding in Figure 15 is that some fringe media organizations such as *breibart* and the *gatewaypundit* would gain a wider readership by moving mainstream while, surprisingly, *foxnews* is located exactly where it maximizes readership.

4. Concluding Remarks

In this paper we introduce readers to a modeling strategy that takes shared embeds in social media data and estimates the importance of ideological congruence, attention, and reputation. The propose model represents an important contribution to research on news sharing and gatekeeping, describing a path to empirically test existing theories on observational data. The proposed statistical model, we show, allows us to distinguish the type of pressure that users exert on editors as well as the level of congruence between users and media organizations. Finally, we exemplify the proposed methods with data from three major social media events in Argentina, Brazil, and the United States. Finding the "optimal" editorial line may seem unrealistic at first, as editors do not blindly cater to the preferences of readers. But estimation of the "optimal" editorial line allows us to assess how congruent news organizations are with their readers.

The proposed research strategy allows researchers to extract theoretically meaningful results from observational data. While there have been extraordinary computational advances in the study of large social networks, considerable less attention has been given to the development of sound modeling strategies. Our analyses combines computational social science tools with statistical modelling to fill this gap by focusing on the behavioral determinants of news sharing and gatekeeping in observational data.

The cases of #Bolsonaro, #Maldonado, and the #TravelBan show that users on the left and right of the political spectrum have different tastes for ideology, attention, and reputation. Ideological congruence is not equally important in all regions of a political network, nor is the interest of users to "talk" about each issue. Our results converges with survey data in showing that progressive and conservative users demand higher ideological congruence than those located at the center of the three political networks.

Results from this article have important implications for the study of gatekeeping in political communication. As an important contribution to the existing literature, our article provides a statistical strategy to integrate the users' demands into existing models of gatekeeping. Our findings raise the question of why are some news organizations more ideologically attuned to the preferences of readers and to what extent is congruence related to media market shares.

Results from this article also provide a path to discuss issue specific effects on media congruence, which have been generally limited in the study of political communications. As we integrate standard models of voting to understand social media behavior, our article provides a bridge to connect the delivery of political content in social media in the context of political campaigns. The proposed methodological strategy hopes to move research in that direction, combining observational user data with voting models to understand why not all bubbles are created equal.

References

Anderson, C. W. (2011). Between creative and quantified audiences: Web metrics and changing patterns of newswork in local us newsrooms. *Journalism*, 12(5):550–566.

Arendt, F., Steindl, N., and Kümpel, A. (2016). Implicit and explicit attitudes as predictors of gatekeeping, selective exposure, and news sharing: Testing a general model of media-related selection. *Journal of Communication*, 66(5):717–740.

Aruguete, N. and Calvo, E. (2018). Time to #Protest: Selective Exposure, Cascading Activation, and Framing in Social Media. *Journal of Communication*, 68(3):480–502.

Bagdikian, B. H. (1983). The media monopoly boston.

Barzilai-Nahon, K. (2008). Toward a theory of network gatekeeping: A framework for exploring information control. *Journal of the American Society for Information Science and Technology*, 59(9):1493–1512.

Boczkowski, P. J. and Mitchelstein, E. (2013). *The news gap: When the information preferences of the media and the public diverge*. MIT press.

Boczkowski, P. J., Mitchelstein, E., and Matassi, M. (2018). "news comes across when i'm in a moment of leisure": Understanding the practices of incidental news consumption on social media. *New Media and Society*, 20(10):3523–3539.

Boehmer, J. and Tandoc Jr, E. C. (2015). Why we retweet: Factors influencing intentions to share sport news on twitter. *International Journal of Sport Communication*, 8(2):212–232.

Boyd, D., Golder, S., and Lotan, G. (2010). Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In *2010 43rd Hawaii International Conference on System Sciences*, pages 1–10. IEEE.

Bright, J. (2016). The social news gap: How news reading and news sharing diverge. *Journal of Communication*, 66(3):343–365.

Calvo, E. and Hellwig, T. (2011). Centripetal and centrifugal incentives under different electoral systems. *American Journal of Political Science*, 55(1):27–41.

Csardi, G. and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal Complex Systems*, 1695:1695.

Domingo, D. (2008). Interactivity in the daily routines of online newsrooms: Dealing with an uncomfortable myth. *Journal of computer-mediated communication*, 13(3):680–704.

Entman, R. M. (2004). *Projections of power : framing news, public opinion, and U.S. foreign policy.* University of Chicago Press, Chicago.

Galtung, J. (1971). A structural theory of imperialism. *Journal of peace research*, 8(2):81–117.

Gans, H. J. (2004). *Deciding what's news: A study of CBS evening news, NBC nightly news, Newsweek, and Time.* Northwestern University Press.

García-Perdomo, V., Salaverría, R., Kilgo, D. K., and Harlow, S. (2018). To share or not to share: The influence of news values and topics on popular social media content in the united states, brazil, and argentina. *Journalism studies*, 19(8):1180–1201.

Hardt, H. (1979). *Social theories of the press: Early German & American perspectives*, volume 6. SAGE Publications, Incorporated.

Kümpel, A. S. (2019). The issue takes it all? incidental news exposure and news engagement on facebook. *Digital Journalism*, 7(2):165–186.

Kümpel, A. S., Karnowski, V., and Keyling, T. (2015). News sharing in social media: A review of current research on news sharing users, content, and networks. *Social media+ society*, 1(2):2056305115610141.

Lowrey, W. and Woo, C. W. (2010). The news organization in uncertain times: Business or institution? *Journalism & Mass Communication Quarterly*, 87(1):41–61.

Macskassy, S. A. and Michelson, M. (2011). Why do people retweet? anti-homophily wins the day! In *Fifth International AAAI Conference on Weblogs and Social Media*.

McKenzie, C. T., Lowrey, W., Hays, H., Chung, J. Y., and Woo, C. W. (2011). Listening to news audiences: The impact of community structure and economic factors. *Mass Communication and Society*, 14(3):375–395.

Rudat, A., Buder, J., and Hesse, F. W. (2014). Audience design in twitter: Retweeting behavior between informational value and followers' interests. *Computers in human behavior*, 35:132–139.

Shoemaker, P. J., Eichholz, M., Kim, E., and Wrigley, B. (2001). Individual and routine forces in gatekeeping. *Journalism & mass communication quarterly*, 78(2):233–246.

Shoemaker, P. J. and Reese, S. D. (1996). *Mediating the message*. White Plains, NY: Longman.

Shoemaker, P. J. and Reese, S. D. (2013). *Mediating the message in the 21st century : a media Sociology perspective*. Routledge/Taylor & Francis Group, New York, third edit edition.

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

Singer, J. B., Hermida, A., Domingo, D., Heinonen, A., Paulussen, S., Quandt, T., Reich, Z., and Vujnovic, M. (2011). Participatory journalism. *Malden, MA: John Wiley & Sons*.

Suh, B., Hong, L., Pirolli, P., and Chi, E. H. (2010). Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In *2010 IEEE Second International Conference on Social Computing*, pages 177–184. IEEE.

Thurman, N., Moeller, J., Helberger, N., and Trilling, D. (2019). My friends, editors, algorithms, and i: Examining audience attitudes to news selection. *Digital Journalism*, pages 1–23.

Vu, H. T. (2014). The online audience as gatekeeper: The influence of reader metrics on news editorial selection. *Journalism*, 15(8):1094–1110.

Waisbord, S. (2017). *Crisis y posprofesionalismo en el periodismo contemporáneo*. Manantial, Buenos Aires.

Wang, X., Liu, H., Zhang, P., and Li, B. (2012). Identifying information spreaders in twitter follower networks. *School of Comput., Infor., and Decision Sys. Eng.*

Zheng, T., Salganik, M., and Gelman, A. (2006). How many people do you know in prison?: using overdispersion in count data to estimate social structure in networks. *Journal of the American Statistical Association*, 101:474.

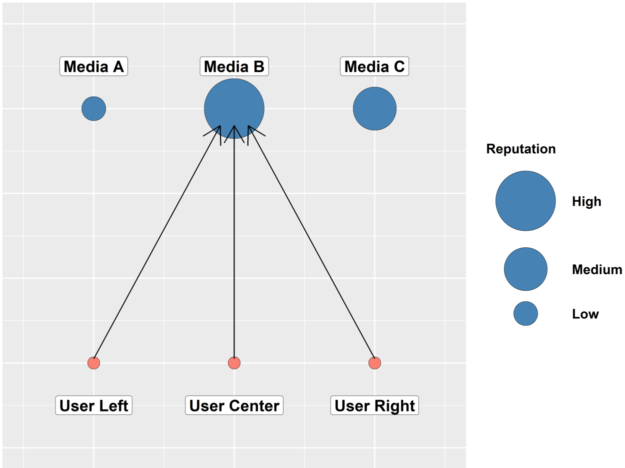


Figure 1: Effect of Reputation in a social media embeds. Users on the left, center, and right of the political spectrum embed content from Media B

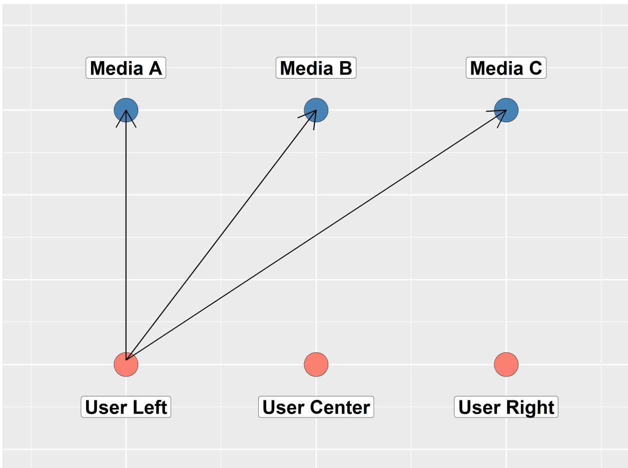


Figure 2: Effect of Attention in a social media embeds. User on the left embeds more content

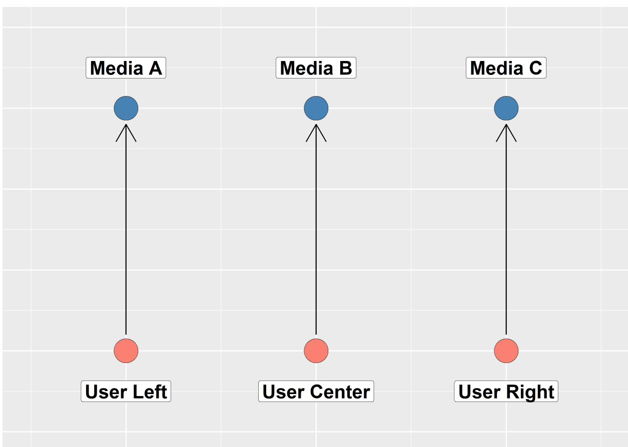


Figure 3: Effect of Ideological Congruence in a social media embeds. Users on the left, center, and right of the political spectrum embed content to media that is ideologically closer to them.

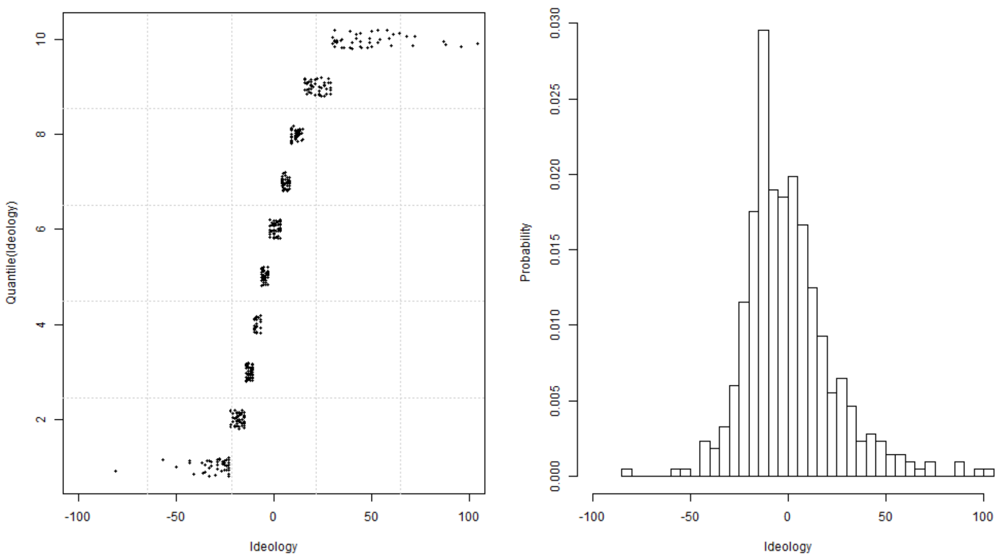


Figure 4: Binning the variable Ideology in equally sized quantiles

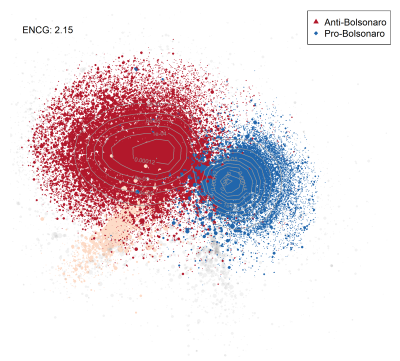


Figure 5: Visualization of all retweets in the #Bolsonaro network. Red color for anti-Bolsonaro network. Blue color for pro-Bolsonaro network. Fruchterman-Reingold layout estimated in Igraph/R. Size of nodes proportional to each user's log(in-degree)

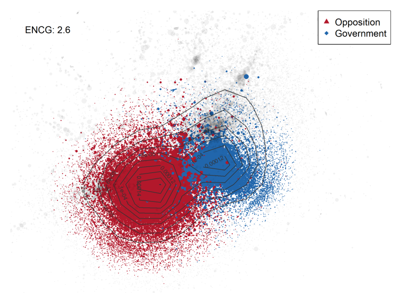


Figure 6: Visualization of all retweets in the #Maldonado network. Red color for opposition network. Blue color for pro-Government network. Fruchterman-Reingold layout estimated in Igraph/R. Size of nodes proportional to each user's log(in-degree)

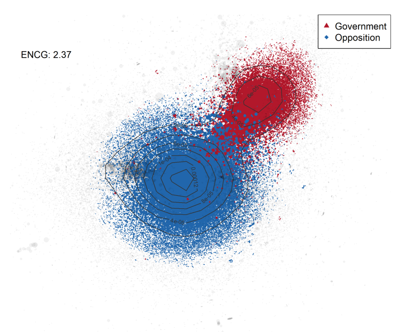


Figure 7: Visualization of all retweets in the #TravelBan network. Red color for anti-TravelBan network. Blue color for pro-TravelBan network. Fruchterman-Reingold layout estimated in Igraph/R. Size of nodes proportional to each user's log(in-degree).

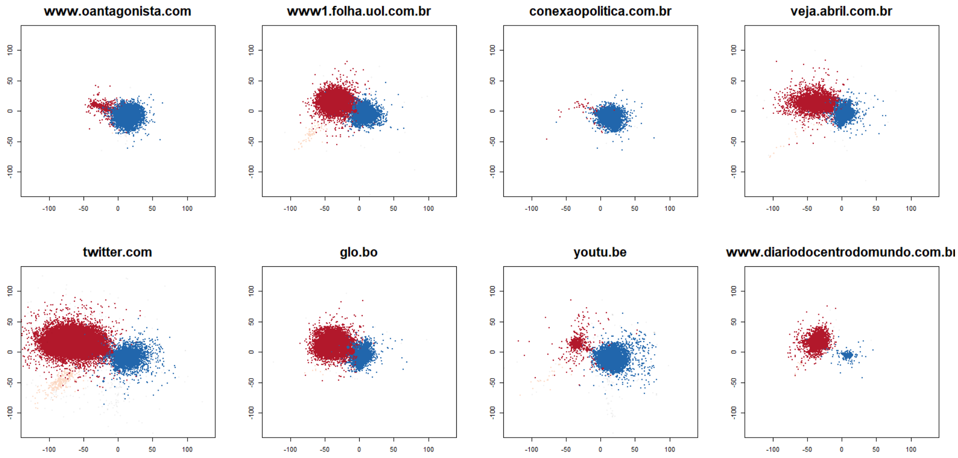


Figure 8: #Bolsonaro: size of nodes describe the log-count hyperlinks to each media

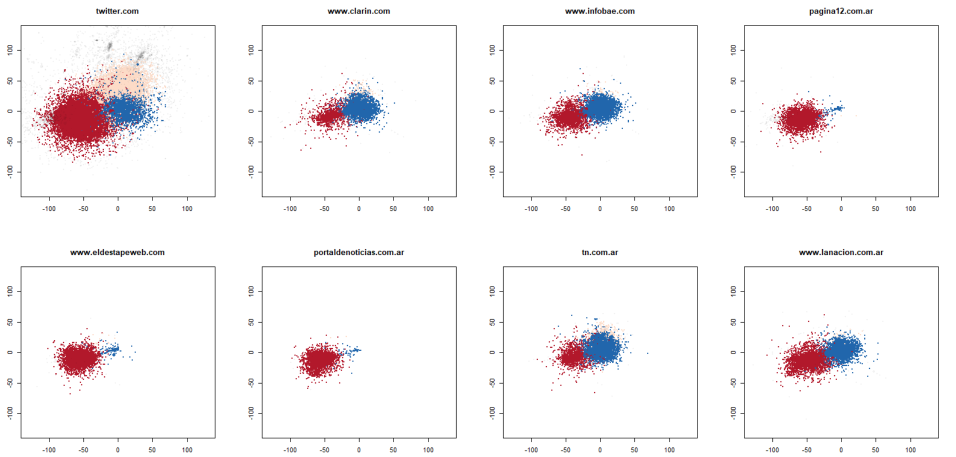


Figure 9: #Maldonado: size of nodes describe the log-count hyperlinks to each media

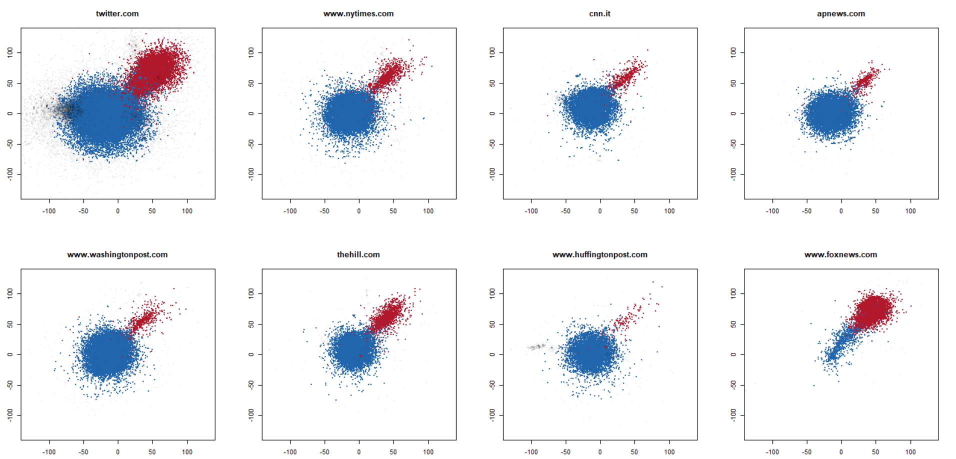


Figure 10: #TravelBan: size of nodes describe the log-count hyperlinks to each media

Figure 11: Ideology by quantile, all three networks

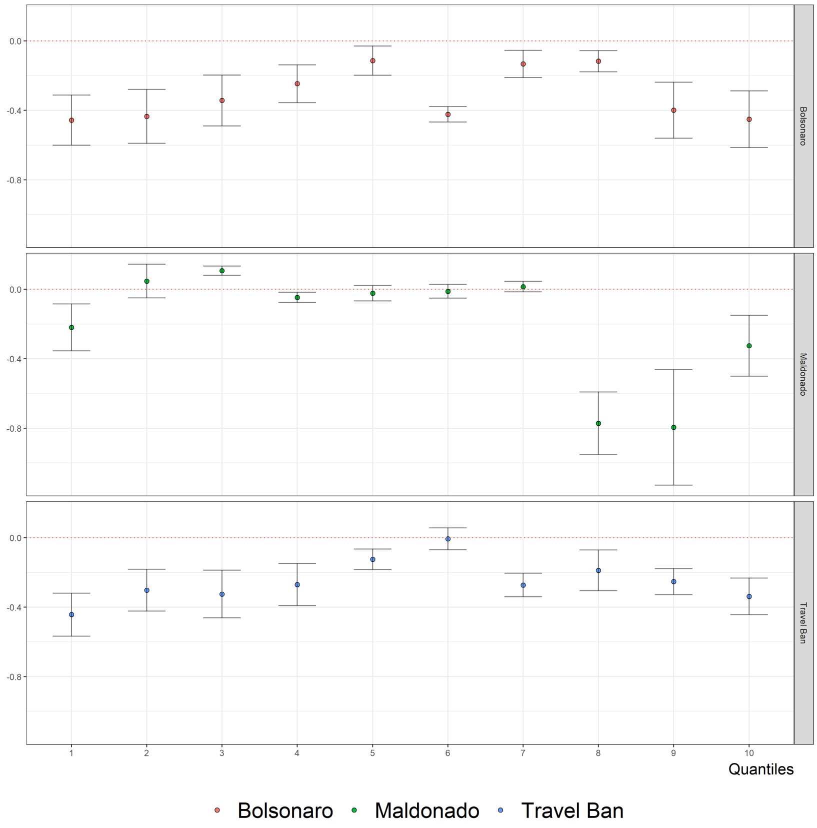
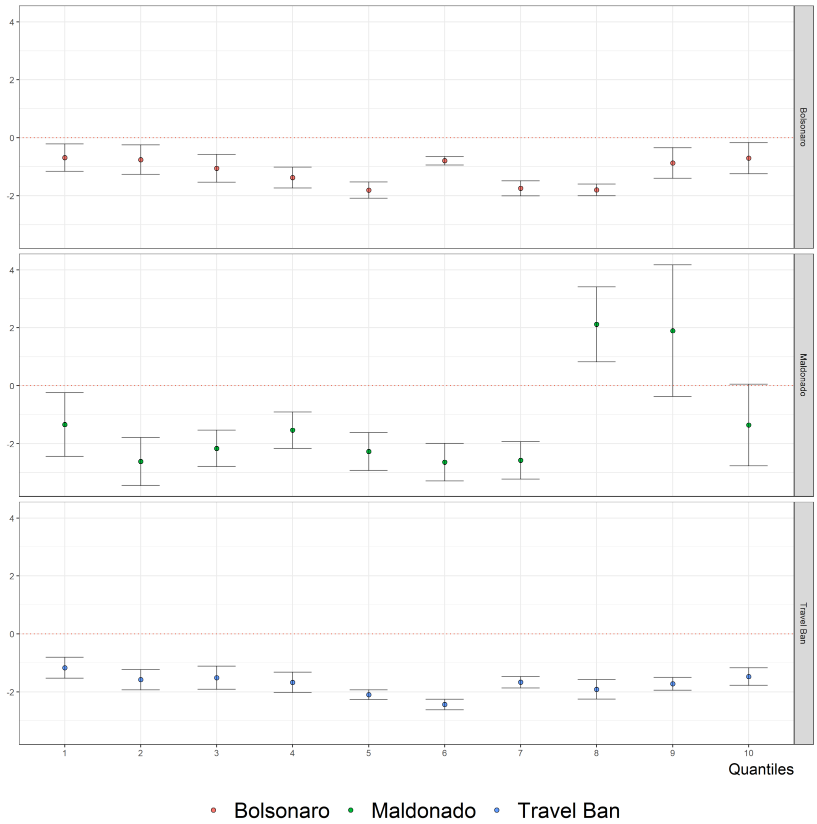


Figure 12: Attention by quantile, all three networks



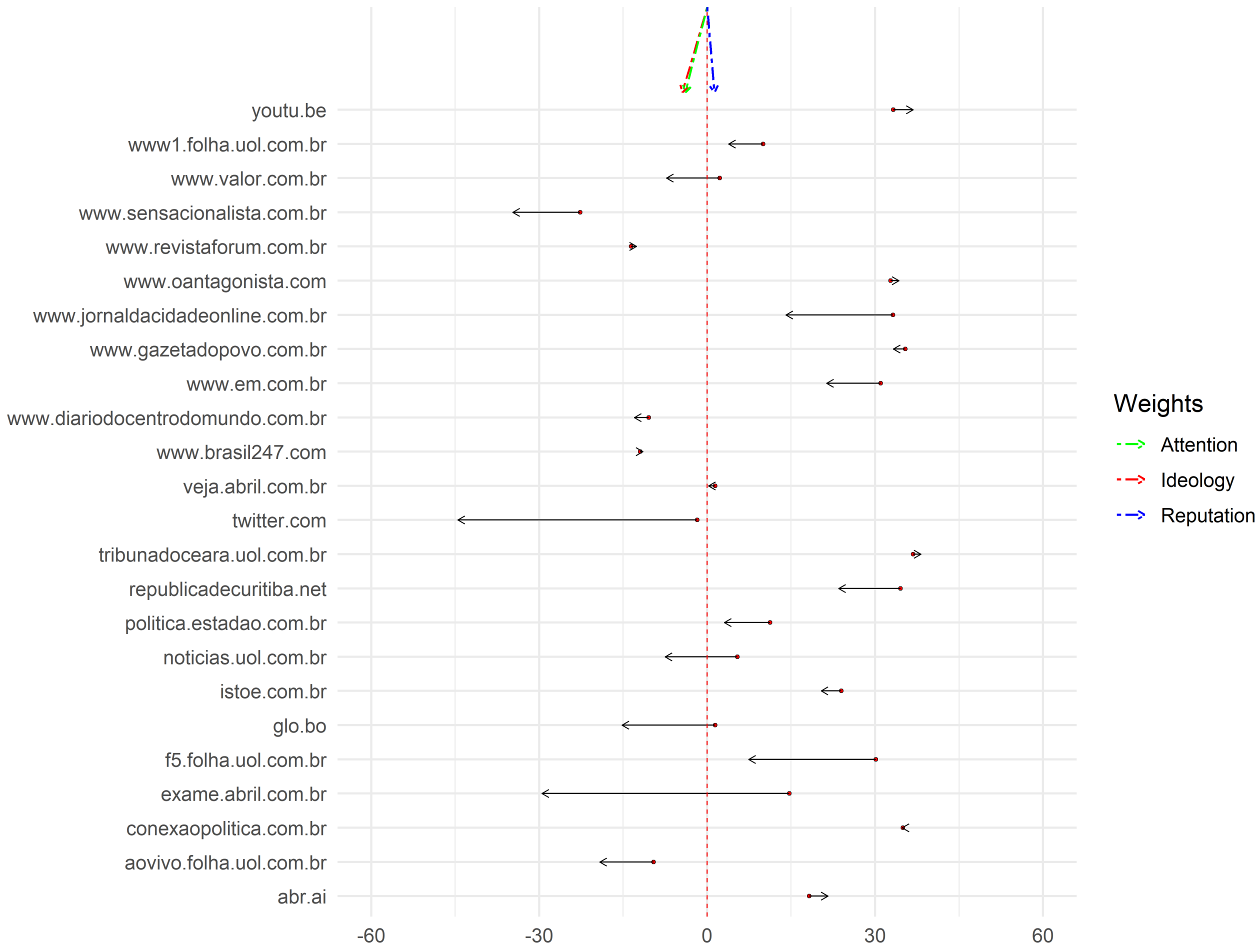


Figure 13: #Bolsonaro: Horizontal arrows describe the difference between the observe location and the optimal location of each news outlet. Vertical arrows describe the difference between the median voter and the weighted median voter by each parameter set, as described in equations (4), (5), and (6)

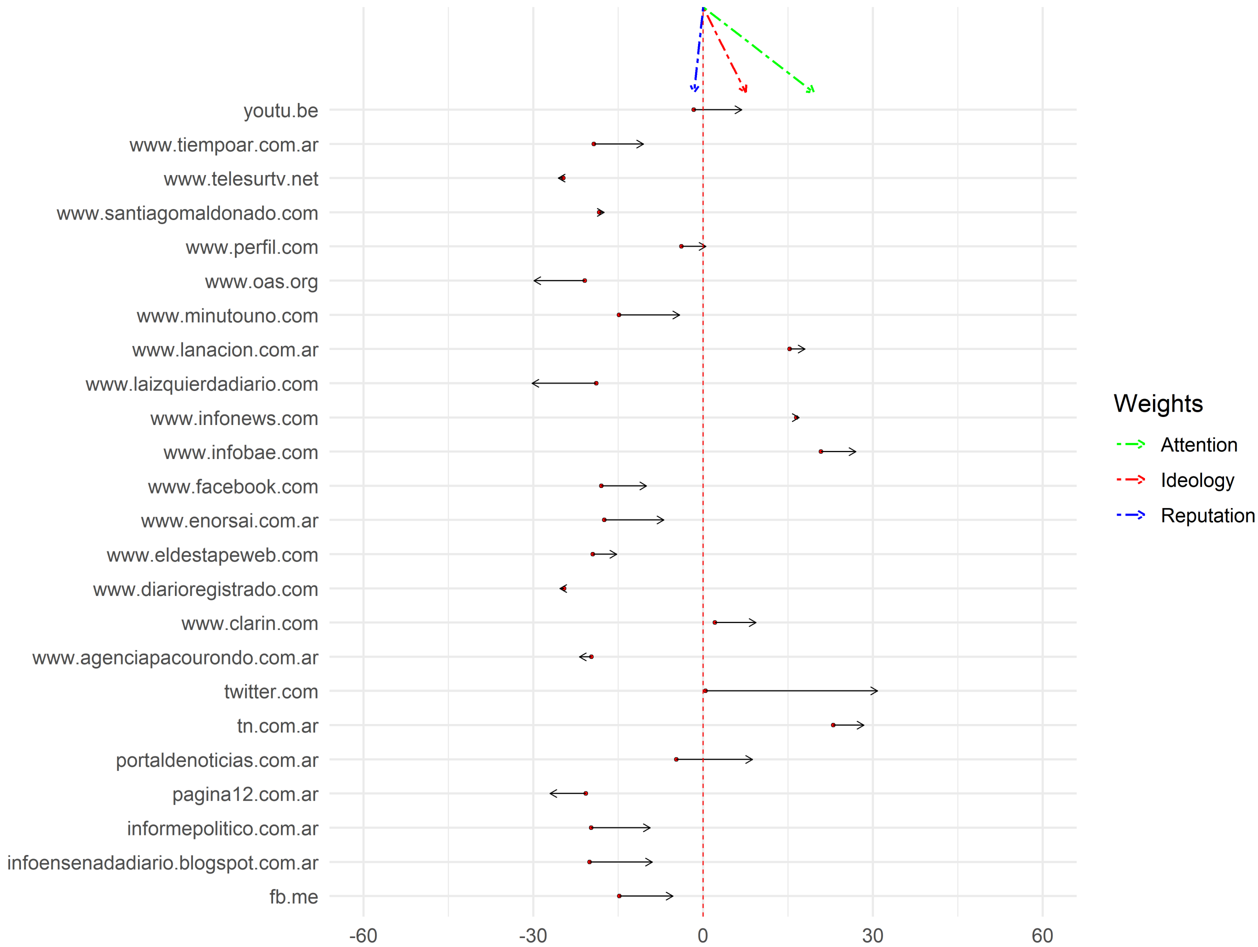


Figure 14: #Maldonado: Horizontal arrows describe the difference between the observe location and the optimal location of each news outlet. Vertical arrows describe the difference between the median voter and the weighted median voter by each parameter set, as described in equations (4), (5), and (6)

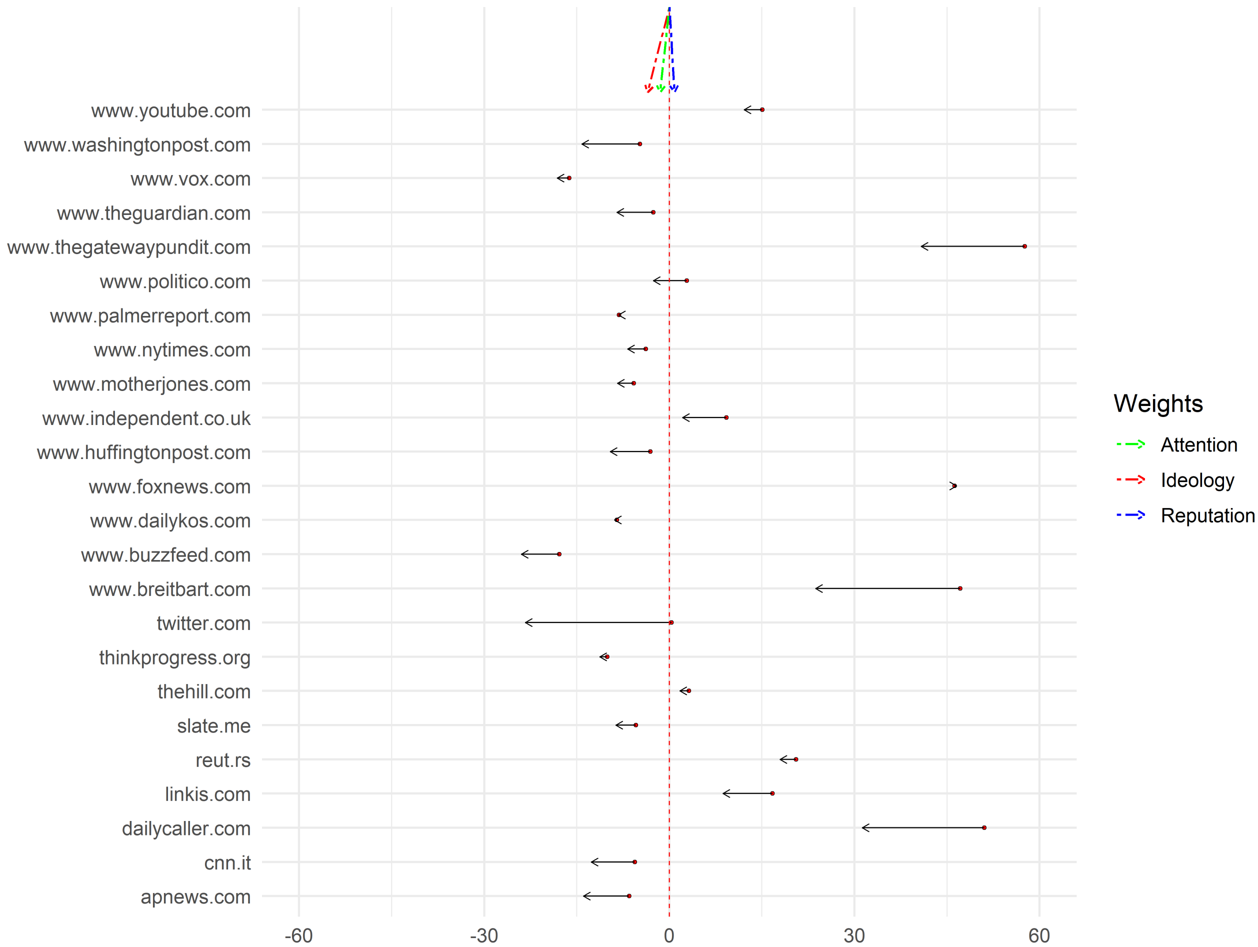


Figure 15: #TravelBan: Horizontal arrows describe the difference between the observe location and the optimal location of each news outlet. Vertical arrows describe the difference between the median voter and the weighted median voter by each parameter set, as described in equations (4), (5), and (6)