News by Popular Demand: Ideological Congruence, Issue Salience, and Media Reputation in News Sharing[1]

Natalia Aruguete[2] **Ernesto Calvo[3]** **Tiago Ventura[4]**

Forthcoming at the International Journal Press/Politics

Number of words: 7,995

May/2021

Abstract: Social media *news sharing* has become a central subject of scholarly research in communication studies. To test current theories, it is of the utmost importance to estimate meaningful parameters of *news sharing* behavior from observational data. In this article, we retrieve measures of *ideological congruence*, *issue salience*, and *media reputation* to explain *news sharing* in social media. We describe how the proposed statistical model connects to different strands of the news sharing literature. We then exemplify the usefulness of the model with an analysis of the relationship between *ideological congruence* and *issue salience*. Results show that if *ideology* and *salience* correlate with each other, the preferences of ideologues (i.e. users that give higher weight to ideological congruence) will be overrepresented in observational data. This will result in heightened perceptions of polarization. We test the performance of the model using data from Brazil, Argentina, and the US.

Keywords: News Sharing, Social Media, Issue Salience.

[1] We received invaluable feedback from the members of the interdisciplinary Lab for Computational Social Science (iLCSS-UMD) as well as important suggestions from Andy Baker, Martın Becerra, Fernando Guarnieri, Gabriel Kessler, Philip Kitzberger, Nadia Koziner, Liliana Mason, Vicky Murillo, Ivan Schuliaquer, Max Stabile, Ed Summers, Maria Laura Tagina, Marisa Van Bulow. Gabriel Vommaro, Esteban Zunino, Silvio Waisbord, and Yanina Welp. Earlier versions of this article were presented at the 2018 Meeting of the American Political Science Association Meeting, and the Second Annual Politics and Computational Social Science (PaCSS) conference.

[2] Universidad Nacional de Quilmes, Departamento de Ciencias Sociales, UNQ.

[3] University of Maryland, Government and Politics, UMD.

[4] University of Maryland, Government and Politics, UMD. Corresponding Author: venturat@umd.edu

1. Introduction

Why do users share news articles[5] with social media peers? How important are ideological considerations, issue salience, and the reputation of a news organization in the decision to embed and share news hyperlinks? News sharing is a complex phenomenon that combines attitudinal and reputational traits with the broad effects on the propagation of news, exposure to news, and on journalistic professional practices (Kümpel, Karnowski, and Keyling 2015; Choi and Lee 2015; Strömbäck, Djerf-Pierr, and Shehata 2013; Boehmer and Tandoc Jr 2015). As interest grows to explain the mechanisms underlying news sharing, scholars require workable empirical strategies to measure sharing behavior using observational data.

This article advances a statistical model to derive *news sharing* behavior from observational social media data. Our method decomposes news sharing in three sets of parameters: (i) user-specific *issue salience*, (ii) news media *reputation*; and (iii) *ideological congruence*. These three sets of parameters align with three different and well-established research areas in the communication's literature—described in Kümpel, Karnowski, and Keyling (2015)—which focus on the individual incentives to share content (**Type I**); the features of the content being shared (**Type II**); and group differences in network effects (**Type III**).[6] A similar classification was presented by Scheufele and Nisbet (2013), who write "(i)n many ways, selectively attending to

[5] In this article, we analyze the users' decisions to share hyperlinks to articles published by news organizations. We consider only formal embeds, where there are explicit links that redirect readers to articles published by these media organizations. The empirical strategy may be generalized to other content, such as the organization's handles (@nytimes), as well as to partial content such as images, text, or cited references.

[6] Kümpel et al. (2015: 4-5) describe the three different research areas in response to three questions: 1. "What motivates persons and organizations to share news in social media?"; 2. "what kind of news content is shared in social media (successfully)?"; and, 3. "how do general network structures influence news sharing?". This last one comprises a number of group and network effects of which we focus on one: group-specific ideological congruence.

some messages over others, based on perceptions of source credibility[7] [***Type II***], ideological congruence [***Type III***], or issue-specific interest [***Type I***], is what enables us to efficiently sift through large amounts of information" (Scheufele and Nisbet 2013: 45). In this article we show how these three distinct and rich literatures can be empirically evaluated as features of a common sharing matrix in observational data.

To exemplify the usefulness of our statistical approach, we use Twitter data and test the relationship between *ideological congruence* and *issue salience* (Weaver 1991). Researchers have documented a positive correlation between political position-taking and news sharing (Delli Carpini 2004; Strömbäck, Djerf-Pierre, and Shehata 2013), tested using survey and experimental data (Oosterhoff, Shook, and Ford 2018). If *ideological congruence* and user specific *issue-salience*[8] are positively correlated, ideologues would share news at higher rates than non-ideologues. The result would be that the preferences of ideologues would be overrepresented in observational data. We test for this relationship, finding support for a positive and statistically significant correlation between ideological congruence and issue salience in observational data. We test our statistical approach with data from the election of Bolsonaro in Brazil, the disappearance of activist Santiago Maldonado in Argentina, and the Travel Ban enacted by Donald Trump in the US.[9]

The organization of the article is as follows: in the next section, Section 2, we describe the importance of retrieving sensible behavior information from social media embeds to test current

[7] In studies of peer-to-peer networks, it is more frequent to use "reputation" as a construct that includes different features that make a website attractive. Credibility, for example, is one among a variety of attributes or properties that may make news organizations more attractive to users.

[8] It is worth noticing that issue salience is here described as an attribute of the user. Indeed, we will show in our empirical analysis that there is variation in mean sharing by groups of users.

[9] We use Twitter's *stream* and *search* APIs to collect to 5,3 million tweets in Argentina between August 1 and October 18, 2017; 2.9 million tweets in Brazil from September 26 through October 2, 2018; and 2 million tweets in the US in January 30 and 31 of 2017. Further details reported in section 4.

theories of *news sharing* using observational data. In Section 3, we present a formal description of the statistical model and exemplify how to interpret the parameters of interest. We then, in Section 4, discuss current research that argues why we should expect a positive correlation between *ideological congruence* and *issue salience*, a hypothesis that we can test using our estimation strategy. In Section 5, we describe our parameter estimates. Finally, in Section 6, we report support for the hypothesized relationship between ideology and issue salience. We show that ideologues are not only more likely to share content from a narrower set of news outlets but that they are also unconditionally more likely to share more news.

Our results contribute to the ongoing debate about the formation and dynamics of echo chambers in social media (Barbera 2020; Sikder et al. 2020; Guess 2021; Flaxman et al. 2016). The positive correlation between ideological congruence and issue salience explains why the content preferred by ideologues is overrepresented in social media networks. Even if partisan sorting is modest (Guess 2021), users may still perceive echo chambers when ideology and salience correlate with each other (Bail 2021).

2. Theory: A statistical description of news sharing in social media data

2.1 Why is it important to develop a statistical model of news sharing?

In a recent article, Kümpel, Karnowski, and Keyling (2015) conduct a meta-analysis of 461 articles on *news sharing* published between 2004 and 2014. The authors note that the number of *news sharing* articles published every year in peer-reviewed journals increased from 10 in 2004-2005 to over two hundred in 2013-2014. News sharing is today a key phenomenon that shapes journalistic and editorial practices (Blanchett Neheli 2018; Russell 2019); forges reciprocal ties among journalists (Hanusch and Nölleke 2019), and is an important source for journalistic content (Von Nordheim, Boczek, and Koppers 2018). Indeed, with the rise of social media, news sharing and news sharing behavior have become central topics in the communication's literature (Lee and Tandoc Jr 2017).

The emerging research on news sharing centers on three distinct but interrelated phenomena summarized by Kümpel, Karnowski, and Keyling (2015) in their review of the literature. The authors group different strands of news sharing research into three types or families: first, there are user-level traits that explain the users' preference for sharing content ("why do some users share more news than others?"). Second, Kümpel, Karnowski, and Keyling describe a burgeoning literature that focuses on content features that increase the likelihood of news being shared ("why is some content shared more frequently by users?"). Finally, a third strand of research focuses on group-specific features that segment news sharing ("why do these groups of users share these particular sets of news"?).

These different concerns yield a vast literature on the subjective, social, rational, and emotional factors that explain individual behavior (Kümpel 2019; Boehmer and Tandoc Jr 2015; Boyd et al. 2010; Rudat et al. 2014; Scheufele and Nisbet 2013); the content-level features that facilitate news sharing (Trilling et al. 2017; Suh et al. 2010; Wang et al. 2012; Karnowski et al. 2020; Macskassy and Michelson 2011); and the group differences (ideology, partisanship, and network structures) that segment the audiences (Barbera 2020). As we propose below, these three different agendas correspond to distinct features of a news-sharing matrix of observational social media embeds, which may be jointly estimated and serve communications scholars working on the subject.

2.2 Modeling news sharing behavior: an intuitive description

We begin with a formal description of the three sets of parameters and their connections to the existing literature. Consider an individual user u_i who observes a thread of social media posts with embedded hyperlinks to news articles. She has a natural "click" rate (i.e. a *trigger finger*) and wants to connect with her peers (Waruwu 2020; Weeks and Holbert 2013). Before sharing a post, however, she *asks*[10] the following three questions: (1) Is this content worthy of my attention?; (2) Is this news created by a reputable source?; and (3) do I agree with the content of this news? Each positive response increases the likelihood that she will share news while each negative response reduces this likelihood. The important question we hope to elucidate is, how much do each of these three incentives (i.e. respectively issue salience, reputation, and ideological congruence) matter for any given set of users that are active on social media and engaged during a particular time and on a given topic?

Much of the earlier research on news sharing focused on the individual level incentives to share news. The *Users and Gratification Approach* (UGA) is a prime example, later extended to more general theories describing the norms, motives, and attitudes of users (Karnowski, Leonhard, and Kümpel 2018). We define this dimension of news sharing as *issue salience*:

> **Def.1: *issue salient* is the utility of sharing content on an issue that the user considers more important (row feature).**

A different mechanism (and literature) explains news sharing by focusing on the characteristics that make some content more broadly liked (Type II literature). A prime

[10] We use the term *asking* loosely to denote a *fast* response to some stimuli. As in (Kahneman 2011), sharing is an automatic (system 1) response, where users are cognitively lazy and do not invest effort in their sharing response. This is usually the case in social media, as users browse through dozens or even hundreds of posts in a few minutes.

determinant of sharing is the reputation of the news organizations (Suh et al. 2010; Wang et al. 2012; Karnowski et al. 2020; Macskassy and Michelson 2011). In their recent article, "*Worth to Share?*", Karnowski et al. (2020) posit that "we need to learn more about the content characteristics that determine whether news is shared widely or is not disseminated at all" (pg. 60). In related work, Trilling et al. (2017) provide evidence of the effect that the importance of a news agency's reputation has on news sharing. The *reputation* dimension of news sharing, borrowing the term widely used in Peer-to-Peer (P2P) networks, explain differences due to the brand name of different organizations:

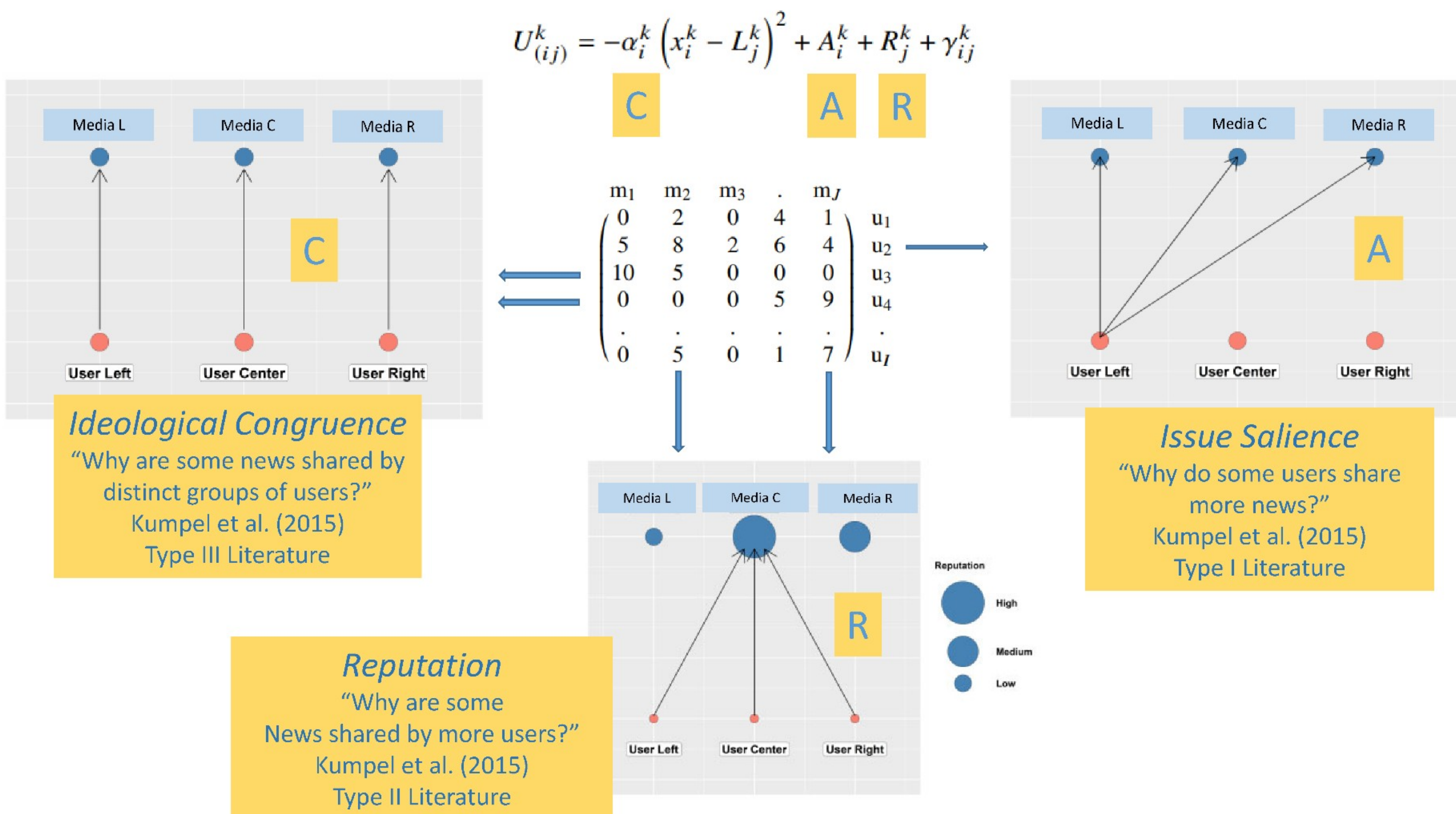
> **Def.2: *Reputation* is the utility of sharing content from news outlets recognized as higher quality by peers (column feature).**

Finally, a broad literature shows that different groups of users are more likely to share content that is ideologically congruent and from news organizations whose editorial lines are ideologically closer to them (Flaxman et al. 2016; Barberá 2020). We define this dimension of news sharing as ideological congruence:

> **Def.3: *Ideological Congruence* is the utility a user derives from sharing content that is consistent with her prior beliefs (cell feature).**

These three different *sharing* incentives remain unobserved for every user and news organization. However, we may obtain reasonable estimates by modeling aggregate responses from a large number of users. Figure 1 describes the underlying model, with each of the different sets of parameters described by row features (issue salience), column features (reputation), and cell features (ideological congruence). Table 1 presents a possible distribution of social media embeds to exemplify the different parameters of the model in observational data. The next section provides a more formal treatment.

Figure 1: Theoretical Model for News Sharing



$$
\begin{array}{cccccc}
m_1 & m_2 & m_3 & . & m_J & \\
\hline
0 & 2 & 0 & 4 & 1 & u_1 \\
5 & 8 & 2 & 6 & 4 & u_2 \\
10 & 5 & 0 & 0 & 0 & u_3 \\
0 & 0 & 0 & 5 & 9 & u_4 \\
. & . & . & . & . & . \\
0 & 5 & 0 & 1 & 7 & u_I
\end{array}
$$

Note: Table with counts of news embeds by user (rows) and media organizations (columns) is used to extract parameters issue salience, reputation, and ideological congruence.

2.3 How to interpret aggregate news sharing data

Consider a vector of social media users (rows) that embed hyperlinks to news published by media organizations (columns).[11] Table 1 of Figure 1 provides an example, with each user $u_i \in I$ sharing news published by media organizations, $m_j \in J$. For presentation purposes, to simplify the visualization of Table 1, let us assume that these media organizations are listed from left to right by conservatism, the dominant dimensions retrieved from the data in our case studies,[12] so that m_1 is less conservative than m_2, and $m_1 < m_2 < \ldots < m_j$. Sorting media organizations from left to right on table one allows the reader to easily interpret the column and cell features we are interested in.

In Table 1 of Figure 1, user u_2 shares a high number of news embeds (25) while user u_1 shares very few (7). We may consider this as an indication that user u_2 gives more weight to the issue reported in the news than u_1. Readers may also note that m_2 is shared by all but one of the users while m_3 is shared by only one user. That is, the news content published by m_2 is shared more broadly than that of m_3, which we consider an indication that m_2 is more reputable (higher quality, more investment, easier access, etc.). Finally, we may see that u_3 is sharing more progressive news (news published by outlets on the left) while u_4 is sharing more conservative news (news published by outlets on the right).

As political communication scholars, we value this information because knowing that some users really want to share content on an issue (rows) is conceptually different than knowing that

[11] Throughout the paper, we use *news sharing* and *news embeds* interchangeably.

[12] Multiple dimensions may distinguish news organizations, ideology (left-right) being but one possibility. When modeling congruence, we allow the data to characterize the underlying dimension. In the particular cases of Argentina, Brazil, and the US, political news are well described by a primary left-right dimension.

content from some outlets is widely shared (columns). This is also conceptually different than knowing that some users share information from only some outlets (cells). Demand for news on an issue by users and prevalence of news by a source are conceptually different (Kümpel et al. 2015; Karnowski et al. 2020; Trilling et al. 2017; Scheufele and Nisbet 2013).

3. The Statistical Model: nuts and bolts

We now provide a detailed description of the statistical model. Consider a utility function where each social media user u_i minimizes ideological 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 a news organization m_j. We define editorialized news as content that is posted by a news organization and that has an ideological charge readily observed by the user. News organization m_j may be a newspaper, 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 u_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 a social media post.

The utility of user u_i also increases with the perceived reputation R_j^k of news organization m_j. That is, users want to minimize cognitive dissonance when sharing news, but cognitive congruence is valuable insofar as information is trustworthy. For example, if the National Enquirer and Fox News publish the exact same article, we expect that users will value the National Enquire lower than Fox News. Therefore, we expect a higher likelihood of sharing the Fox News article.

By assumption, ideological dissonance is *negative* while reputations are *positive*. That is, users are less likely to share posts that disagree with their ideological beliefs and see a declining utility from news that is further removed or openly challenges those beliefs. While it is possible that users "ironically" share news that is cognitively dissonant, there is no empirical evidence that shows systematic sharing of dissonant content in social media. However, if users shared news that is cognitively dissonant (across the aisle), the model would describe such behavior.

Users also receive a *positive* utility for information they agree with if it is published by reputable news organizations. The reputation and the ideological leaning of news organizations may or may not be correlated with each other. Readers of Fox news, for example, may consider that its publications are of high reputation precisely because they minimize ideological dissonance.[13] 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.

[13] The "halo" effect has been extensively analyzed in political communication. See Kahneman (2011), Chapter 7, for an extensive discussion of this issue.

Issues may also be more or less important to each user. We identify this behavioral parameter as the variable A_i^k in user' utility function. Therefore, a Reader u_i will perceive a utility from sharing news on issue *k* by organization *j* as described in Equation (1):

$$U_{(ij)}^k = -\alpha_i^k \left(x_i^k - L_j^k\right)^2 + A_i^k + R_j^k + \gamma_{ij}^k \qquad \text{Eq. (1)}$$

In Equation (1), the quadratic term $-\alpha_i^k \left(x_i^k - L_j^k\right)^2$ describes the disutility of a publication by media m_j on issue k, with ideological leaning L that is further removed from the reader's preferred ideological position, x_i^k. For every unit of increase in ideological dissonance, the utility of user i declines by α. The parameter α also has a natural interpretation as the weight that a user 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, $\alpha_i^{Bieber} < \alpha_i^{Trump}$. As in Gelman et al. (2004), α, x, and L may be interpreted as latent and unobserved parameters.

Equation (1) also shows that news published by a more reputable actor, R_j^k, increase the utility of user u_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, $R_{ij}^{Bieber} < R_{ij}^{Trump}$. 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 salience to an issue, A_i^k, sharing a higher or lower average number of post with their social media peers. As in Delli Carpini (2004), issue salience is heterogeneous and varies by user and issue.

Equation (1) also includes an stochastic term that captures overdispersion, γ_{ij}^k, 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 users attach to ideological congruence, while the random intercepts describe the importance of issue salience and reputation.

Sharing news can take many forms, such as reading (clicking), liking, or sharing content (retweeting). For simplicity, we consider the number of times users share content as the dependent variable. Multilevel estimation of the proposed model proceeds as in Zheng, Salganik, and Gelman (2006). Readers may readily observe that equation (1) has a very large number of parameters and imposes significant computational demands. In the Supplemental Information File (SIF) we describe a strategy to reduce computational demands, binning *issue salience*—the row parameters—in quantiles.[14]

4. Testing the model: Are ideologues over represented in observational data?

Table 2 summarizes our parameter definitions and measurement strategy. Readers may immediately notice that interesting questions emerge when considering the relationship between the different parameter sets. For example, if the importance or weight that a user gives to ideology increases, will she also perceive the issue as more salient? Readers will immediately notice that if salience and ideological congruence are positively related to each other, ideologues would be more active and more readily observed in social media data.[15] Therefore, the content they share will be overrepresented, resulting in a network that appears more politicized (and likely polarized) than its average user. If ideologues preferences are over represented in news

[14] There is extensive research that retrieves the users' network position and uses this information as a proxy for her ideal point. This assumption is accepted practice in the emerging literature on social media preferences (Barberá 2015; Bond and Messing 2015; Conover et al. 2010). The assumption takes the decision to retweet as a signal of affinity, with more heterogeneous preferences taking on a more central location in the network.

[15] An interesting question is also considering if Reputation and Ideology are also related. This is an important question that directly touches on the literature of gatekeeping ("what editors decide to publish"). The "column" view of the problem, glanced in Figure 1 but not fully discussed, is a promising research extension that exceeds the goals of this article.

sharing data, therefore, the network will display heightened perceptions of polarization among the public. As it was described by Scheufele and Nisbet (2013): "Our social networks, that is, the people we are surrounded by in most of our daily activities, tend to be extremely like-minded and homogenous in their demographic and ideological makeup." (Scheufele and Nisbet 2013: 46). Such perception, however, may be the result of how different users differ in their behavior rather than being a feature of their social compositions and network prevalence.

Table 2: News Sharing Model - Parameters, Theory and Measurement.

Parameters	Theoretical Definition	Measurement Strategy
Issue Salience	The utility of sharing content more important or salient	Random Intercepts for each user (row) in the embeds matrix
Reputation	The utility of sharing content that is also shared by peers and, consequently, recognized as high quality	Random Intercepts for Media (column) in the embeds matrix
Ideological Congruence	The utility of sharing content that is consistent with prior beliefs	Distance in the first dimension of the network between user u_i and media m_j

Note: All the three parameters are estimated using a multilevel specification with random sloped by equally sized quantiles in the first dimension of the network. The use of random slopes reduces computational burden for the model and also provides measures to verify variation in the weights of each parameter in different parts of the network.

The over-representation of ideologues in social networks is a variation on the well-known *Friendship Paradox* (Feld 1991), with more connected and more active nodes resulting in subjective perceptions of polarization that differ from mean polarization. If more extreme users differ from less extreme users in how frequently they share news, the result will be observational data that is more polarized than its users.

A broad literature on affective polarization has shown that intense ideologues are also unconditionally more motivated to participate in politics and in social media (Mason 2018; Huddy et al. 2015; Barberá 2020, Guess 2021). Similar findings—this time related to sharing *fake* *news*—in a recent study by Osmundsen et al. (2021), show negative partisan effects yielding large increases in the likelihood of sharing news. If negative and positive evaluations of political events result in users seeking and delivering information that is consistent with their preferences, motivated users will be both more enthusiastic as well as more attuned to particular types of evidence, which will positively correlate ideological beliefs and issue salience (Weaver 1991). Therefore, the hypothesis to be tested states that:

> H_1***: More ideological users, who share news that aligns with their preferences, will also be unconditionally more issue-motivated than non-ideological users.***

The corollary to this hypothesis is that the relationship between ideology and issue salience will be multiplicative rather than additive. Therefore:

> H_2***: A positive correlation between ideological congruence and issue salience will bias the frequency with which researchers will observe high levels of polarization.***

5. 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.[16] 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 shifts in the status quo.

The data collection for the three cases followed similar procedures. We collected data using both APIs available on Twitter: the forward stream and the backward search. The streaming API collects live data, letting users capture a portion of tweets in real time. The search API allows users to access a repository of tweets published seven days prior to the query. Our search used

[16] For a related study of news sharing in these same countries see García-Perdomo et al. (2018).

both APIs with three search terms: Bolsonaro, Maldonado, and Travel Ban. By collecting data from both APIs we increase our sample (both in terms of size and source), therefore avoiding risks of messages' removals and algorithmic bias made arbitrarily by Twitter for each API (Timoneda 2018).

After collecting the data, we limit our sample to a network of all retweets from the original data, and retrieve information about the edges (retweet), the author of the tweet (authority) and the users who retweeted the original message (hubs). We then retained the largest connected cluster of the network, eliminating users with less than two retweets (out-degree>=2). With the thinned network, we draw users' [x,y] coordinates implementing the Fruchterman-Reingold algorithm in igraph-R (Csard 2006). We then ran the walk.trap algorithm to identify the main clusters in the network. We label these clusters as communities in our network, and we validate them with qualitative analysis of the main authorities in each cluster (see appendix E). We describe below these communities, and the data collection for each case.

First, we consider the *Bolsonaro* network in Brazil, using 2,943,993 tweets published by 162,107 high activity accounts the week prior to the election of President Jair Bolsonaro, from September 26 through October 2, 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 (Aruguete et al. 2021). Consequently, there are significant differences in the reputation of traditional media outlets and new extremely conservative ones.

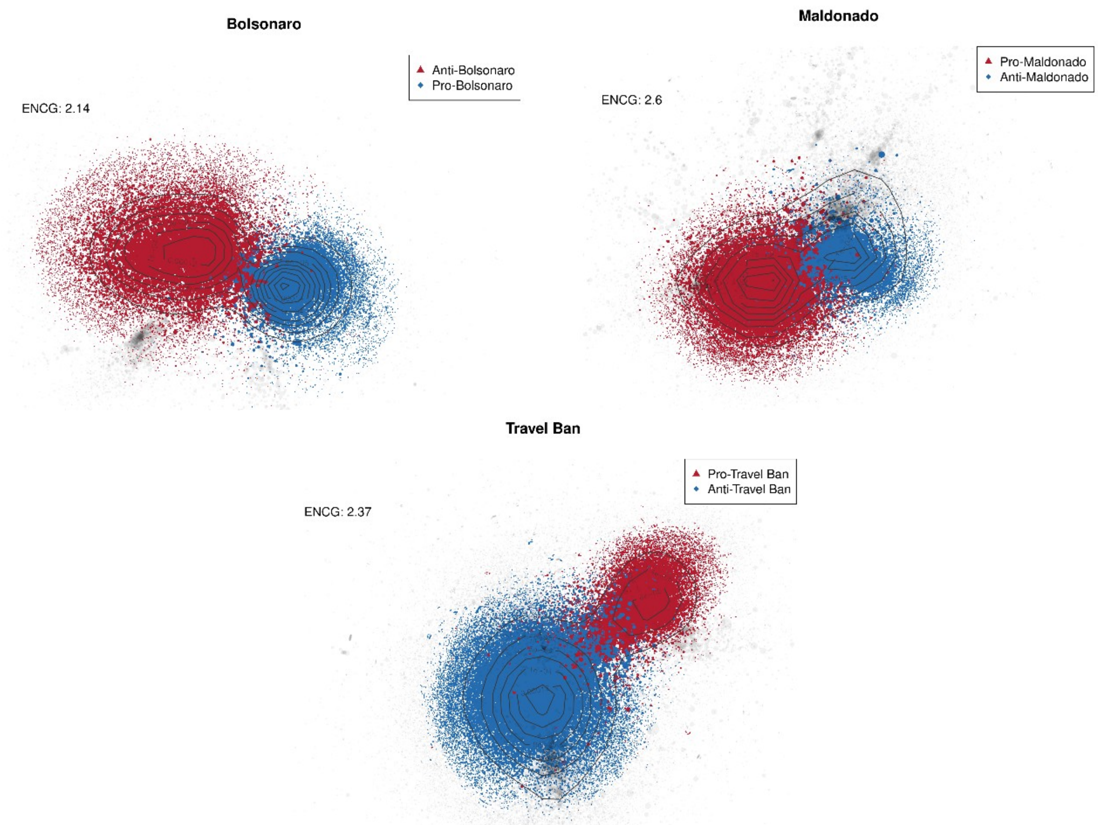
The upper-left plot of Figure 2 describes the basic layout of the *Bolsonaro* network, with pro-Bolsonaro users in blue and anti-Bolsonaro users in red. Of the more than 2.9 million retweets analyzed in the data, 432,591 (14.7%) included hyperlinks. The number declines to 387,841 (13.2%) if we do not consider hyperlinks to other tweets. The most frequent news outlet embedded in the data is the pro-Bolsonaro *Oantagonista*, which represents 63,862 hyperlinks, 14.7%, and is intensively retweeted by core supporters of the current president.

In the case of *Maldonado* in Argentina, the upper-right plot in Figure 2, 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, from August 1st to October 18th 2017. 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, lower-left plot of Figure 3, we analyze 2,031,518 retweets from 241,271 high activity accounts on January 30 and 31, 2017, following the decision of the Trump administration to restrict travel from seven majority Muslim countries. 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.[17]

[17] 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 embeds within each of these networks, as we discuss next.

Figure 2: Visualization of all retweets in the Bolsonaro network (upper-left), Maldonado (upper-right), and Travel Ban (lower).



Note: Layout (Fruchterman-Reingold) and Community detection (Random Walk) for the Bolsonaro, Maldonado, and Travel Ban networks. Igraph in R 3.6.

For each of the three networks we retrieve the matrix of users (rows) and media organizations (columns), keeping only the 24 most frequently embedded news outlets. We retrieve the first dimension value for each user (horizontal axes in Figures each plot of Figure 2), as a proxy for x_i^k in equation 3, and the quantile indices $q(i)$. To estimate the media locations, L_j^k, we borrow from the model developed by Bond and Messing (2015) and estimate the average x_i^k position for the user u_i, weighted by the number of media m_j embeds.[18]

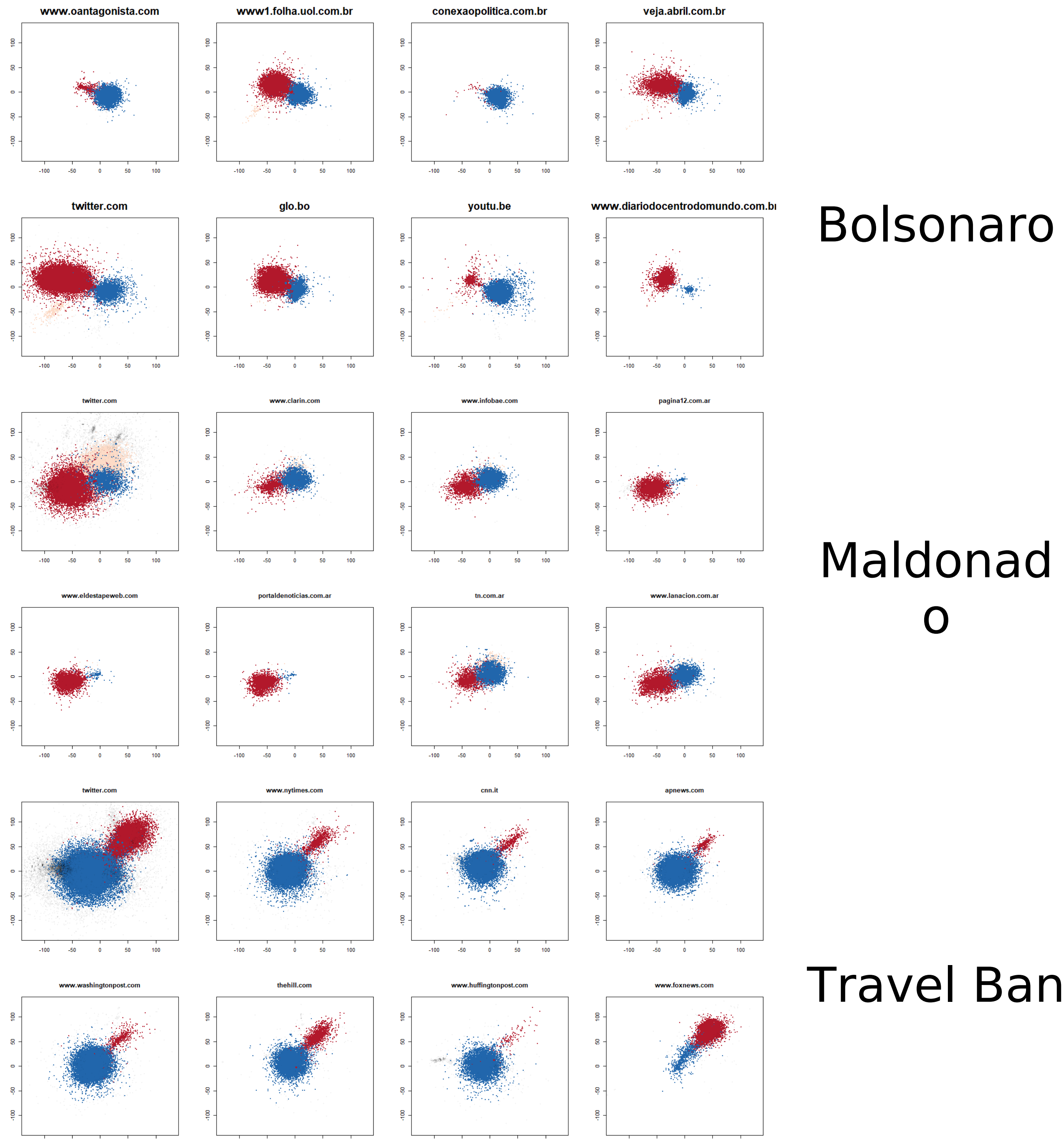
[18] We opt for this method for two main reasons. First, not all the media organizations had active account profiles on Twitter, therefore, we could not simply extract their positions from the network. Second, it is very uncommon for these official accounts to retweet other users, and for that reason, their position in a network of retweets is not the best representation of their ideological location.

5.1 A visual inspection of Media Embedding

Figure 3 presents 24 plots that describe the areas of the network activated by the top eight media outlets in *Bolsonaro*, *Maldonado*, and the *Travel Ban*. The other 48 media outlets can be found in the online supplemental file. In all three of our cases, the activated nodes describe the region of the network where news sharing is more active.

The top plot of Figure 3 shows activation in Brazil, with *Oantagonista* and *Conexaopolitica* activated more readily by users on the right of the political spectrum. The former was recently founded by three prominent 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.

Figure 3: News Sharing in Bolsonaro (top), Maldonado (middle), and Travel Ban (bottom)



Such is the result of the decision by Twitter to suspend accounts from the conservative group *Movimiento Brasil Libre*, who engaged from within YouTube on a very active campaign of misinformation. 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.

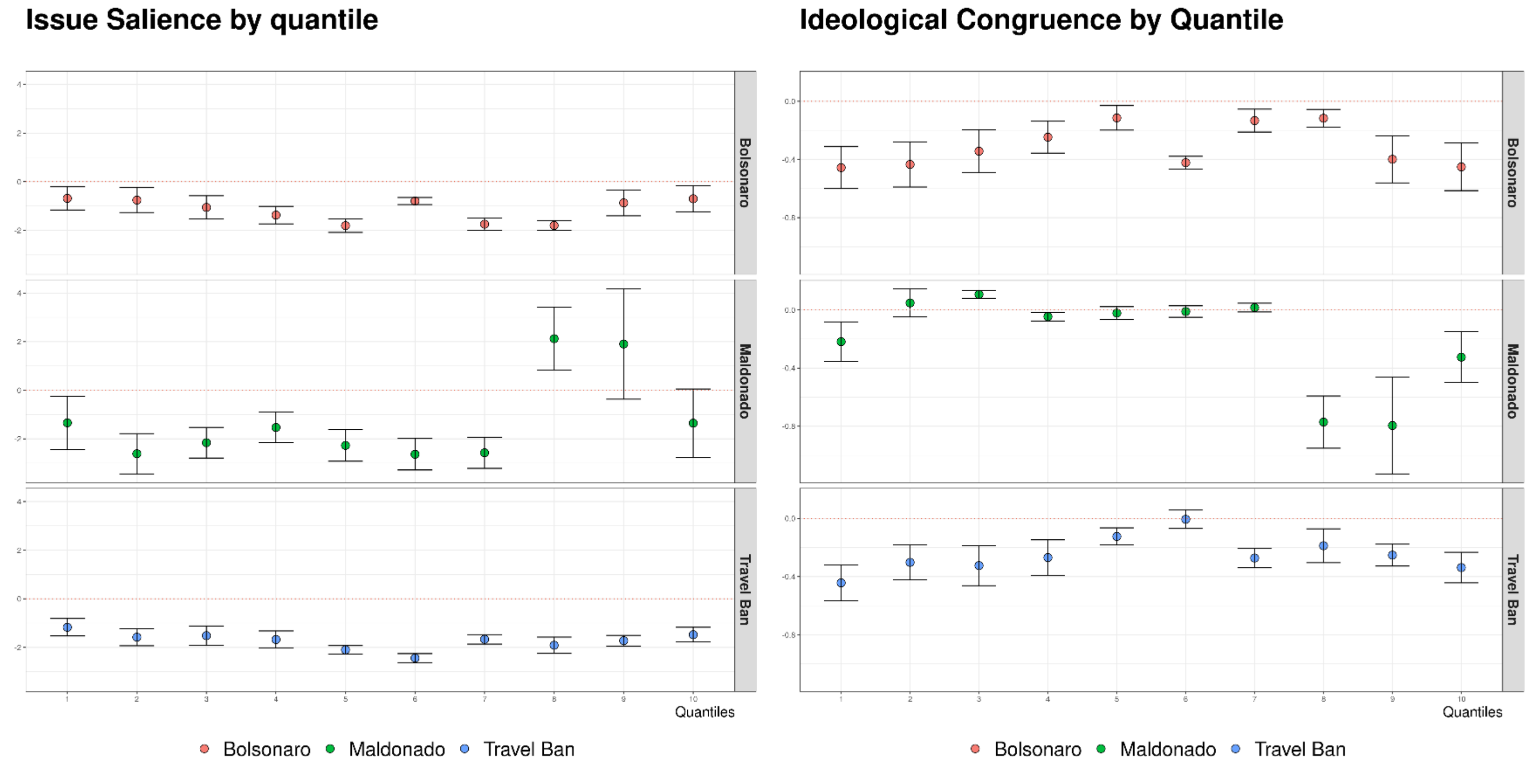
The middle plot in Figure 3 shows distinct activation by Argentine outlets on the left (Página/12) or right (TN) of the political spectrum. However, other outlets such as La Nación are embedded by most of the conservative users but also by a significant number of moderates on the center of the political spectrum. Finally, in the case of the *Travel Ban*, lower plot in Figure 4, we see outlets with 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 *Travel Ban* networks.

5.2 Ideological congruence, Issue Salience, and Reputation

Figures 2 and 3 describe the data captured in the three *RxC* matrices of news embeds of *Bolsonaro*, *Maldonado*, and the *Travel Ban*. Using equations 1 and 2, as well as the binning strategy proposed in section 3.1, we proceed to estimate all three sets of parameters (ideological congruence, issue salience, and reputation). Figure 4 provides a visual comparison of the ideology and issue salience parameters by quantile for each of our three cases.

Figure 4: Ideology by Quantile (left) and Issue Salience by Quantile (right)



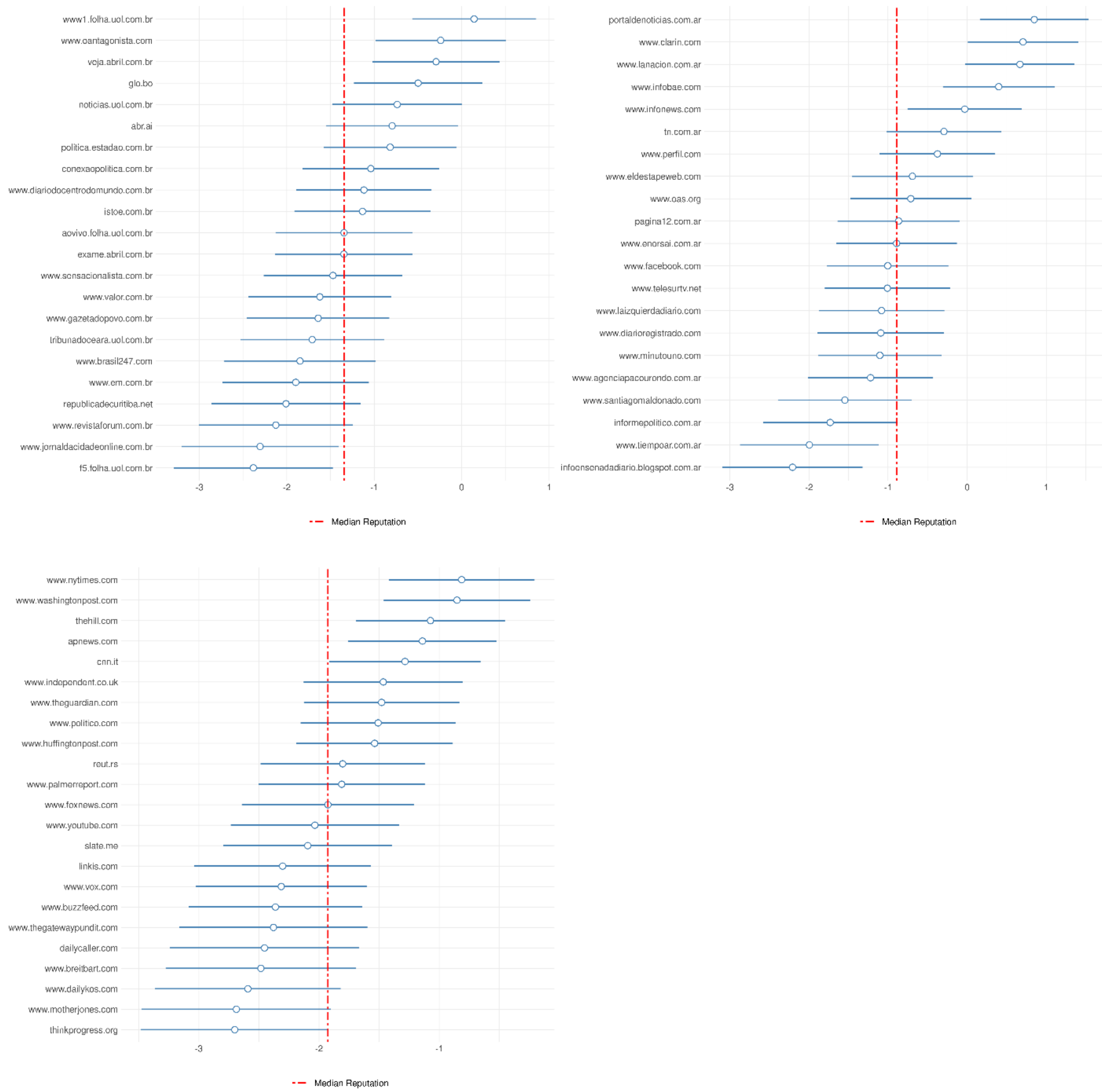
Note: Parameters of ideology and salience by quantile estimated from Equation (1).

Because ideological distance is cognitively costly, larger negative values indicate that ideological congruence 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 our results with similar estimates in survey data, we will see that ideological congruence has a significantly larger weight in these networks. In effect, most estimates of ideological congruence in survey data fall in the range of [-.05, -.12], four times smaller than the estimates in observational social media data (Calvo and Hellwig 2011).

The right plot in Figure 4 provides estimates of issue salience by quantiles, with larger values indicating a higher propensity to embed links on the collection terms we use for each case. 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 issue salience of users to the right of the political spectrum in #*Maldonado*, with activity that is orders of magnitude above the rest of the network. While users on the right of the political spectrum were fewer in number in the #*Maldonado* network, they still shared news on the issue at much higher rates.

Figure 5: Reputation parameters by news organization in Argentina, Brazil, and Mexico.



Note: Estimates describe the R_j^k parameters in equation (1). Higher parameter values describe media that is more broadly shared by users, holding *issue salience* and *ideology* constant.

Figure 5 provides estimates of the model's reputation parameter. As described in section three, these parameters capture that propensity of users to embed links to media organizations once we control for issue salience and ideology. As expected, readers can observe that niche organizations on the left and right of the political spectrum are at the bottom of the list, given that most of the news sharing incentives is explained by the users' ideological affinity with the media. By contrast, most traditional news organizations are more broadly shared once we control for the other factors.

In the next section we use the estimated parameters to evaluate the hypothesis that test for a positive correlation between ideology and issue salience.

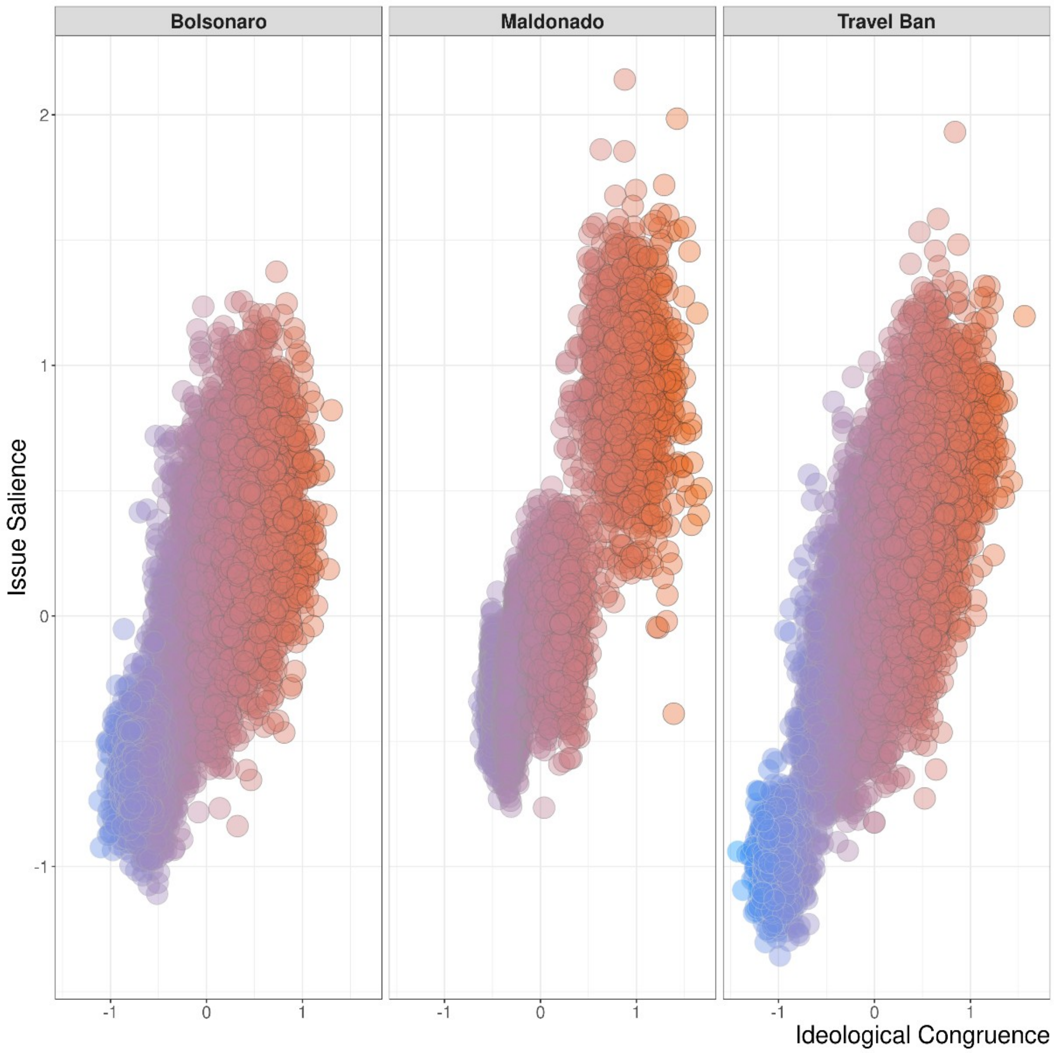
6. Are more intense users over represented in observational data?

Once we estimate the importance of ideology, issue salience, and reputation in observational data, we may use our parameter estimates to test the hypotheses in Section 4, which asked whether more ideological users were over-represented in observational data. As it was presented, if issue salience and ideology are positively correlated, then political news shared in social media will appear to be more polarized than they actually are. That is, the preferences of intense ideologues would be over-represented in observational data.

Visual inspection of Figure 4 in the previous section hinted of a relationship between ideological preference and issue salience. The *salience* parameters in *Bolsonaro*, *Maldonado*, and the *Travel Ban* closely align with the *ideology* parameters. It is worth reminding that while issue salience increases sharing behavior, ideological distance measures dissonance and, consequently, reduces sharing behavior. To simplify the interpretation of the results, we display α instead $-\alpha$, showing a positive correlation between the weight users attach to ideology and to issue salience.

Results displayed in Figure 6 strongly support Hypothesis 1, with a close fit between issue salience and ideology in all three countries. Country correlations of .82 for the US, .86 for Brazil, and .92 for Argentina validate the hypothesis for all three cases, showing clear support in observational social media data.

Figure 6: Relationship between the Ideology and Issue salience parameters, all three networks



Note: The figures use augmented data from the parameter estimates in equation (1). To provide a more intuitive understanding, we inverted the value of the ideology parameter; therefore, positive values for issue salience and ideological congruence means higher weights on both dimensions. The pearson correlation between ideology and issue salience using the augmented data is 0.84, 0.92 and 0.86 for each network, respectively.

The joint effect of ideology and issue salience in Figure 6 provide a clear mechanism to explain the appearance of high ideological sorting in social media data (Barberá 2020;

Osmundsen et al. 2021, Flaxman et al. 2016), which reflects the outsized weight of ideologues' news sharing behavior on politically salient issues.

Figure 6 provides support for H_1, showing that *ideology* and *issue salience* are closely connected. We may also evaluate the expected content shared by ideologues compared to non-ideologues by substituting proper values for the parameters in Equation (1). Consider for example a news organization that publishes posts exactly at the location of two users with $\alpha_1^k > \alpha_2^k$, so that user 1 is an ideologue and user 2 is not an ideologue. Given that $-\alpha_{1,2}^k (0)^2 + A_{1,2}^k$, we find that the content shared by user 1 is larger than the content shared by user 2 by exp $\left(A_1^k - A_2^k\right)$. Given that both A_1^k and A_2^k are linearly related to α_1^k and α_2^k, the content shared by ideologues is overrepresented in the observational data as expected by H_1.

7. Concluding Remarks

This paper describes a statistical strategy to study news sharing behavior using observational social media data. We propose a simple model that takes a matrix of embeds as inputs and estimates the importance of ideological congruence, issue salience, and reputation in social media data. Our model provides a path to test existing theories of news sharing using observation social media embeds. We exemplify this method with Twitter data from three major social media events in Argentina, Brazil, and the United States, although the model could use as input any *RxC* matrix of embeds.

The model allows researchers to estimate meaningful parameters of interest from observational data. While there have been extraordinary computational advances in the study of large social networks, designs that answer practical communication theory problems receive considerably less issue salience. Our analysis combines computational tools with multilevel modelling to fill this gap, focusing on the behavioral determinants of news sharing.

As it pertains to the cases of *Bolsonaro*, *Maldonado*, and the *Travel Ban*, results show that users on the left and right of the political spectrum are both more attentive to the issues and more likely to share ideologically congruent news. The proposed model allows us to test for a positive correlation between issue salience and ideology, which explains why ideologues are overrepresented in social media data.

The parameters retrieved from the matrix of news embeds may also be used to test other important communication problems, such as the editorial incentives to cater to extreme users (*gatekeeping*). Alternative specifications of the proposed model may also be used to explore the linkages between ideology and reputation. Finally, future extensions of this model may expand the matrix formulation to issue dimensions, with users (rows), media (columns), cognitive congruence (cells), and issues (layers). These are computationally tractable extensions that will allow researchers to work with flexible and scalable models to better understand news sharing behavior.

References

Aruguete, N., & Calvo, E. (2018). Time to# protest: Selective exposure, cascading activation, and framing in social media. *Journal of communication*, 68(3), 480-502.

Aruguete, N., Calvo, E., & Ventura, T. (2021). News sharing, gatekeeping, and polarization: A study of the# Bolsonaro Election. *Digital journalism*, 9(1), 1-23.

Bafumi, J., Gelman, A., Park, D. K., & Kaplan, N. (2005). Practical issues in implementing and understanding Bayesian ideal point estimation. *Political Analysis*, 13(2), 171-187.

Bail, C. (2021). *Breaking the Social Media Prism: How to Make Our Platforms Less Polarizing*. Princeton University Press.

Barberá, P. (2015). Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Political analysis*, 23(1), 76-91.

Blanchett Neheli, N. (2018). News by numbers: The evolution of analytics in journalism. *Digital Journalism*, 6(8), 1041-1051.

Boyd, D.M., Golder, S., & Lotan, G. (2010). Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter. In Proceedings of the 43rd Hawaii International Conference on System Sciences (HICSS) (pp. 1–10). Honolulu, HI: IEEE.

Bond, R., & Messing, S. (2015). Quantifying social media's political space: Estimating ideology from publicly revealed preferences on Facebook. *American Political Science Review*, 109(1), 62-78.

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

Choi, J., & Lee, J. K. (2015). Investigating the effects of news sharing and political interest on social media network heterogeneity. *Computers in Human Behavior*, 44, 258–266.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. InterJournal, complex systems, 1695(5), 1-9.

Delli Carpini, M. X. (2004). Mediating democratic engagement: The impact of communications on citizens' involvement in political and civic life. In L. L. Kaid (Ed.), *Handbook of political communication research* (pp. 395–434). Mahwah, NJ: Lawrence Erlbaum.

Duffy, A., Tandoc, E., & Ling, R. (2020). Too good to be true, too good not to share: the social utility of fake news. *Information, Communication & Society*, 23(13), 1965-1979.

Feld, S. L. (1991). Why your friends have more friends than you do. *American journal of sociology*, 96(6), 1464-1477.

Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public opinion quarterly*, *80*(S1), 298-320.

García-Perdomo, V., Salaverría, R., Kilgo, D. K., & 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.

Guess, A. M. (2021). (Almost) Everything in Moderation: New Evidence on Americans' Online Media Diets. *American Journal of Political Science*.

Huddy, L., Mason, L., & Aarøe, L. (2015). Expressive partisanship: Campaign involvement, political emotion, and partisan identity. *American Political Science Review*, 109(1), 1-17.

Jones, B. D., & Baumgartner, F. R. (2005). *The politics of attention: How government prioritizes problems*. University of Chicago Press.

Hanusch, F., & Nölleke, D. (2019). Journalistic homophily on social media: Exploring journalists' interactions with each other on Twitter. *Digital Journalism*, 7(1), 22-44.

Kahneman, D. (2011). *Thinking, Fast and Slow*. Macmillan.

Karnowski, V., Leonhard, L., & Kümpel, A. S. (2018). Why users share the news: A theory of reasoned action-based study on the antecedents of news-sharing behavior. *Communication Research Reports*, 35(2), 91-100.

Karnowski, V., Leiner, D. J., Sophie Kümpel, A., & Leonhard, L. (2020). Worth to Share? How Content Characteristics and Article Competitiveness Influence News Sharing on Social Network Sites. *Journalism & Mass Communication Quarterly*, 1077699020940340.

Kümpel, A. S., Karnowski, V., & 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.

Lee, Eun-Ju, and Edson C Tandoc Jr. 2017. When News Meets the Audience: How Audience Feedback Online Affects News Production and Consumption. *Human Communication Research* 43 (4): 436–49.

Luo, M., Hancock, J. T., & Markowitz, D. M. (2020). Credibility perceptions and detection accuracy of fake news headlines on social media: Effects of truth-bias and endorsement cues. *Communication Research*, 0093650220921321.

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

Mason, L. (2018). *Uncivil agreement: How politics became our identity.* University of Chicago Press.

Oosterhoff, B., Shook, N. J., & Ford, C. (2018). Is that disgust I see? Political ideology and biased visual attention. *Behavioral brain research*, 336, 227-235.

Osmundsen, M., Bor, A., Vahlstrup, P. B., Bechmann, A., & Petersen, M. B. (2020). Partisan polarization is the primary psychological motivation behind “fake news” sharing on Twitter. American Political Science Review, 115(3), 999-1015.

Russell, F. M. (2019). Twitter and news gatekeeping: Interactivity, reciprocity, and promotion in news organizations' tweets. D*igital Journalism,* 7(1), 80-99.

Scheufele, D. A., & Nisbet, M. C. (2013). Commentary: Online news and the demise of political disagreement. *Annals of the International Communication Association*, 36(1), 45-53.

Sikder, O., Smith, R. E., Vivo, P., & Livan, G. (2020). A minimalistic model of bias, polarization and misinformation in social networks. *Scientific reports*, 10(1), 1-11.

Strömbäck, J., Djerf-Pierre, M., & Shehata, A. (2013). The dynamics of political interest and news media consumption: A longitudinal perspective. *International journal of public opinion research*, 25(4), 414-435.

Suh, B., Hong, L., Pirolli, P., & Chi, E. H. (2010, August). Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In 2010 IEEE second international conference on social computing (pp. 177-184). IEEE.

Timoneda, J. C. 2018. Where in the World Is My Tweet: Detecting Irregular Removal Patterns on Twitter. *PloS One* 13 (9): e0203104.

Von Nordheim, G., Boczek, K., & Koppers, L. (2018). Sourcing the Sources: An analysis of the use of Twitter and Facebook as a journalistic source over 10 years in The New York Times, The Guardian, and Süddeutsche Zeitung. *Digital journalism*, 6(7), 807-828.

Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S. (2012, July). A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In Proceedings of the ACL 2012 system demonstrations (pp. 115-120).

Waruwu, B. K., Tandoc Jr, E. C., Duffy, A., Kim, N., & Ling, R. (2020). Telling lies together? Sharing news as a form of social authentication. *New Media & Society*, 1461444820931017.

Weaver, D. 1991. Issue Salience and Public Opinion: Are There Consequences of Agenda-Setting? *International Journal of Public Opinion Research* 3 (1): 53–68.

Weeks, B. E., & Holbert, R. L. (2013). "Predicting dissemination of news content in social media: A focus on reception, friending, and partisanship." *Journal & Mass Communication Quarterly*, 90(2), 212-232.

Zheng, T., Salganik, M. J., & 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), 409-423.