Network Activated Frames:

Content Sharing and Perceived Polarization in Social Media

Natalia Aruguete* Ernesto Calvo † Tiago Ventura ‡

October 27, 2021

Abstract

Our paper describes how the sharing behavior of interconnected users alters the frequencies of content observed by social media peers. Changes in the frequency of distinct frame elements, in different regions of a social network, shape how individuals interpret, classify, and define situations and events. We label this process Network Activated Frames (NAF). We test the mechanisms behind NAF with an original image-based conjoint design that replicates network activation in three surveys. Results show that partisans share more content than non-partisans and that their preferences become over-represented. Our findings show that a network of peers with cross-cutting ideological preferences will still amplify disproportionate numbers of partisan frames. Beginning with fully randomized probabilities, the output from our experiments is more extreme than the preference of the median users, as partisans activate more frame elements than non-partisans. We implement the survey experiments in Argentina, Brazil, and Mexico.

Number of words: 8,112

Keywords: Framing, Social Media, Conjoint Experiments

*Universidad Nacional de Quilmes, Departamento de Ciencias Sociales, UNQ. Address: Castro Barros 981. Ciudad Autónoma de Buenos Aires, Argentina. Email: nataliaaruguete@gmail.com. Webpage: http://unq.academia.edu/nataliaaruguete

†University of Maryland, Government and Politics, UMD. Address: 3140 Tydings Hall, College Park, MD 20742, USA. Email: ecalvo@umd.edu. Webpage: http://gvptsites.umd.edu/calvo/

‡University of Maryland, Government and Politics, UMD. Address: 4118 Chiconteague, College Park, MD 20742, USA. Email: venturat@umd.edu. Webpage: http://tiagoventura.rbind.io/

1 Introduction

In today's social media environment, the activation and propagation of content requires users to share posts published by their peers. As users share posts, they make content available to a wider public. In this paper, we describe how the sharing behavior of interconnected users alters the frequencies of the texts, images, and endorsements available to peers (Aruguete and Calvo, 2018; Entman and Usher, 2018). Changes in the frequency of frame elements in distinct regions of a social network, shape how individuals interpret, classify, and define situations and events. Understanding how users activate social media content, and the content users expect to see activated by their friends, is critical for explaining framing in social media. We name this behavioral phenomena Network Activated Frames (NAF).

We describe Network Activated Frames on two key dimensions: i) behavior – how individuals activate content they agree with; and ii) expectations – which content users expect that friends will activate. Empirically, we use a novel image-based conjoint experiment to fully described NAF behavior and NAF expectation. We implement three conjoints with pairs of tweets that rotate the author, text, images, and number of likes of competing posts. As we randomize frame elements, we measure changes in the frequencies of the authorities, texts, and images made available to peers.

After exposure, we ask respondents which Tweet they are likely to share, allowing the options *both* and *neither*. We also ask which tweet they expect their friends to share; and which tweet they expect to see first in their favorite news show. The design allows us to compare network activation by oneself and expectations of activation by peers. We implement the conjoint experiments in Argentina, Brazil, and Mexico.

The decision to share content (self) and the expectation of content shared by others (peers),

allows us to explain the subjective perception of *social media bubbles* (Alipourfard et al., 2020; Lee et al., 2019; Jackson, 2019). That is, the perception of a highly polarized social media environment that is the result of more active partisan nodes being overrepresented in the data. Our experiment clarifies the source of conflicting evidence on the existence (or not) of social media bubbles (Barberá et al., 2015; Barberá, 2020; Bakshy et al., 2015). Findings in this article converge with recent research that explores the topological effects of confirmation bias in social networks (Sikder et al., 2020). Findings also align with new studies in social network that distinguish between the connectivity effects (high in-degree) and activation effects (high frequency) of users in networks (Saxena and Kumar, 2019). Findings show statistically significant differences in the frequency and type of frame elements activated by partisans and non-partisans in a fully randomized experiment.

The organization of this article is the following. The first section positions the articles' contributions in relation to the literature on social media bubble, polarization and echo-chambers formation. The second section revisits current models of frame activation in social media and describe the statistical connections between frame activation and the *Friendship Paradox* in social networks (Feld, 1991; Sikder et al., 2020). Third section describes the design of the conjoint experiments and their implementation, while the following section introduces our hypothesis comparing activation by users (*self*) and by their friends (*peers*). The fifth section describes the results in Argentina, Brazil, Mexico. The conclusion summarizes the experimental results of the "friendship paradox" in all three countries. Results show that users expect their friends to be more active than they are and that ideologues expect their friends to be more extreme than they are. We conclude in the sixth section.

2 Network Activated Frames, Social Media Bubbles, and Polarization

Few phenomena are as characteristic of our times as social media and political polarization. Studies about the causes and the consequences of contemporaneous levels of political polarization have abounded, both in the US context (Mason, 2018; Iyengar et al., 2012; Fiorina and Abrams, 2008; Abramowitz and Saunders, 2008) and in a comparative perspective (Gidron et al., 2019). In the vibrant scholarship that developed on the causes of polarization, a frequent question asks whether social media and the emergence of new digital technologies contributed to the heightened levels of polarization of the last two decades (Lelkes et al., 2017; Stroud, 2010; Settle, 2018; Sunstein and Sunstein, 2018)

A popular hypothesis that connects rising social media usage and mass polarization focuses on the ability of new technologies to generate echo chambers or filter bubbles. That is, to deliver content that reinforces existing partisan animosity through motivated reasoning and sorting. The argument was forcefully advanced by Carl Sustein (2018), who argues that the high possibilities of sorting on the user side of with whom to build connections and which information one prefers to consume increases homophily on online networks. In this environment, citizens are mostly exposed to pro-attitudinal opinions, reinforcing their online, and contact with counter-attitudinal messages, and users are therefore severely restricted. In line with a long tradition of deliberative aspects of democracies (Habermas, 1991; Mansbridge, 1983; Fishkin, 1991), polarization rises as a consequence of the lack of cross-contamination on citizens living in their online echo chamber and filter bubbles (Sunstein and Sunstein, 2018; Mason, 2018).

Even though this argument has achieved the status of conventional wisdom, more recent empirical investigations offer more mixed findings of echo chambers' formation on social media. These more nuanced views provide robust empirical evidence that the levels of online segregation

detected in early studies (Bakshy et al., 2015; Conover et al., 2011) are not particularly distinct when compared to offline media consumption (Gentzkow and Shapiro, 2011). In addition, other studies show that most of the homophily on users' networks occurs in interactions related to political events and political discussions (Barberá et al., 2015; Wojcieszak and Mutz, 2009). Finally, considering online media consumption, at least in the United States, users' media habits are considerably less segregated, and with a strong salience of more centrist, highly reputable sources (Guess et al., 2021). Although homogeneity on social media networks is the norm, cross-contamination occurs at more frequent rates that assumed be early studies.

This recent scholarship on exposure to political information on social media, therefore, reveals an interesting puzzle. Most users are embedded in diverse social networks, yet widespread perceptions of polarization on social media are strong among users, experts, and policymakers (Barberá, 2020). Previous studies tackled this puzzle by showing that actually, contrary to more deliberative normative expectations, heterogeneity on users' networks might actually be the driver of polarization. Bail et al. (2018) shows how users' exposure to counter-attitudinal information might increase levels of affective polarization, while others have shown similar findings when focusing on exposure to social media content framed in an uncivil, negative way (Bail et al., 2018; Banks et al., 2021; Suhay et al., 2018). We build upon these previous studies to advance a new explanation for the formation of "compositional" informational bubbles. Using a novel experimental design implemented across three countries, we show that a moderate, balanced users' network on social media does not impede the formation of echo chambers or filter bubbles. Instead of focusing on sorting, we moved the debat to what we consider being the crucial element on how content propagates in social media: how the sharing behavior of interconnected users alters the frequencies of the texts, images, and endorsements available to

peers.

Perceptions of high ideological congruence belong to the family of phenomena knows as the *friendship or class size paradox*, where if there is: "any variation in college class sizes, then more students experience the average class size as larger than the mean. They experience a higher average class size than exists for the college because many students experience the large classes, while few students experience the small classes" (Feld, 1991, pp. 1475). Similar mechanisms explain why changes in the relative frequencies of content activated by ideologues heighten the users' subjective perceptions of social media bubbles. As shown by Sikder et al. (2020), once confirmation bias is formally linked to social connectivity it is enough for a "small group of individuals to generate permanent opinion polarization and cascade dynamics" (Sikder et al., 2020, pg. 1). We provide conclusive evidence that this effect will also occur when the sharing rates of partisans and non-partisans differ from each other.

Because content on social media depends fundamentally on the users' decision to propagate messages, to account for the subjective perception of bubbles in a network it is only necessary to show that partisanship and frame activation are positively correlated (Aruguete et al., 2021). That is, proving that, first, (i) partisans share content with a higher frequency than non-partisans, so that the content they prefer is over-represented in observational data. Second, (ii) proving that partisans share content that is different from that of non-partisans. The result will be that individuals in a network will observe larger doses of partisan content. This general result does not depend on how diverse is our network of peers, which explains why it is possible to both have a diverse network of peers and to see partisan content overrepresented in our social media feeds (Aruguete, 2019; Barberá et al., 2015; Barberá, 2020; Bakshy et al., 2015).

Our experiments confirm both mechanisms that support the perception of social media bub-

bles: beginning with identical probabilities for all frame elements via a conjoint design that randomizes all content, we show that (i) partisans and ideologues are more likely to share content they agree with; and (ii) the preferred content of partisans and ideologues is different from the content preferred by non-partisans. We confirm these conditions when analyzing sharing (self) and sharing expectations (peers).

Support for the "generalized friendship paradox" is theoretically and empirically relevant (Fotouhi et al., 2014; Jo and Eom, 2014; Feld, 1991; Benevenuto et al., 2016; Eom and Jo, 2014). First, while limited topological sorting in networks is one of the reasons that researchers have challenged the existence of social media bubbles, evidence is overwhelming that users are more likely to share content that is ideologically congruent and to perceive that the content shared by peers is highly partisan (Del Vicario et al., 2016; Barberá et al., 2015). Frequency differences in the content shared by our friends will result in sharing probabilities that do not reflect the proportions of ideologues among our friends but rather their degree centrality and their sharing behavior (Saxena and Kumar, 2019).

Second, because network activated frames depend on the *frequency* of activation, rather than on the number of users in the population, frame elements will be weighted towards the preferences of the most connected and engaged *local* users (Barberá, 2020). As the density of partisans increases, the content variance will decline locally and "bubble" like frames will heighten perceptions of polarization among our friends. It is not necessary to follow like-minded friends to be exposed to higher doses of like-minded content and to see congruent content coming from different local regions of a network. Conflicting evidence on cross-cutting ideological connections is not inconsistent with observational data that "looks" like a bubble (Bakshy et al., 2015).

Third, our results explain that partisan voters will increasingly perceive that their friends are

more partisan than they are. By contrast, independent voters in Argentina, Brazil, and Mexico do not report that their friends share more partisan frames in social media. This is consistent with the central tenants of the *generalized friendship paradox*, with "bubble" like content being more prevalent only among those who share the selected trait (i.e. partisanship).

3 Network Activated Frames

Entman coined the term cascading activation (Entman, 2004) to describe how traditional media organizations render visible only some of the frames proposed by elites, preventing some content from reaching the public. Cascading activation, as a faulty Rube Goldberg machine, allows only a subset of the falling pieces to activate source content, altering the frequency of the frames observed by readers. In Entman (2004), however, frame elements are never amplified but rather filtered by the traditional media[1]; framing and frame are two sides of the same coin. The former refers to the integral and active process of production, circulation, and reproduction of socially shared and persistent meanings over time (Reese et al., 2001, Pg.11). The latter is present in the different stages of the communication process. In social media, framing is the result of how content is created and posted by users (production) and how this content is activated by peers (reproduction and circulation).

To address the problem of amplification in social networks, Entman and Usher (2018) generalize the concept of activation as a process that produces, distributes, assimilates, and activates information. The new media scenario prompt them to revise the initial theory in favor of a

[1]Entman (2004) notes that "The metaphor of the cascade was chosen in part to emphasize that the ability to promote the spread of frames is stratified" (Entman 2004, p. 9). They start in the governments, go through the network of nonadministrative elites and follow their course through the news companies and their texts to stay in the public perception schemes. Entman asks if the frames expressed in the highest stratum of that system do manage to arrive intact to the social base or if, instead, alternative interpretations from the bottom level back up to policymakers to challenge the governmental frame.

Cascading Network Activation model, which describes the characteristics of digitization on the symbolic relations of power between elites, traditional media, and citizens.

The concept of Network Activated Frames (Aruguete and Calvo, 2018) seeks to update the notion of integrality of the framing process, taking into account the dynamics of content propagation in a digital media and virtual social networks have a prime role in the circulation of endorsements, texts, and images that structure the social world. The *Network Activated Frames* extends notions of activation in Entman and Usher (2018) to describe the framing effect of frequency changes in content that is amplified by algorithms, users, and the media. Aruguete and Calvo (2018) describe this change in the frequencies of the content shared by users as a "selection effect". Meanwhile, the aggregate frames observed by the users are described as a "compositional effect", with different aggregate interpretations of a phenomena in each region of the social media network.

Behavior that amplifies some frame elements rather than others, the result of a higher weight given to cognitive congruence or the result of more attention to an issue, increases the probability of sharing particular frame element (selection effect). The activation on content by peers, and the aggregation of distinct frames in a given location, yield locally consistent network activated frames ("compositional effect").

In the next section we describe an experiment to measure the "selection effect" of frame elements by partisans. Observed from the consumption side, the end result are local frames that provide a meaningful interpretation of locally important events, with partisans contributing to local frames at a higher rate than non-partisans (i.e. bubbles).

4 Using Conjoint Experiments to Measure Network Activated Frames

The objective of our conjoint experiments is to theoretically relate *activation*, *framing*, and partisanship in social media. After Hainmueller et al. (2014), conjoint designs have become a prominent methodological tool within the fields of communication, political science, and economics, covering many different types of phenomena, such as immigration preferences (Hainmueller and Hopkins, 2015), bureaucratic behavior (Oliveros and Schuster, 2018), corruption (Mares and Visconti, 2020; Klašnja et al., 2021), and vote choice (Franchino and Zucchini, 2015; Kirkland and Coppock, 2018). Conjoint experiments provide treated individuals with two competing profiles with randomized traits (conjoint profiles). After exposure, they ask the subjects to select the profile they prefer or, in our case, the social media post they would like to share.

Different from the traditional conjoint, our experiment adapts this design to compare frame elements that are embedded in social media posts and measure changes in the frequency of the different frame elements. As important, we do not force the selection of one of the two frames, allowing respondents not to share traits. This allows us to observe differences in the activation rate of different frame elements by partisans and non-partisans.

The experiment uses a factorial design that creates on-the-fly Tweets. During the survey, each respondent is exposed to pairs of edited tweets created solely for the experiment. The messages replicates news media content on issues such as public security (Mexico and Brazil) and COVID-19 (Argentina). The messages vary on four dimensions: the author of the tweet (endorsement), the text of the message (positive and negative frames), an associated image (partisan, collaborative, and neutral), and high or low numbers of 'likes' and 'retweets' (public support). In the appendix, we present the full sets of frame elements and examples of the paired tweets that are randomly created. While the frame elements used in the conjoint vary

by country, the design and questions are identical. Therefore, all three experiments test exactly the same two mechanisms: (i) partisans share content at a higher rate than non-partisans and (ii) the content shared by partisans is different than that of non-partisans.

4.1 Conjoint Design

Each of the frame components varies as follows. First, (1) the authors of the tweet randomly display Liberal and Conservative media outlets, creating four possible combinations: Lib-Lib, Lib-Cons, Cons-Lib, Cons-Cons. Second, (2) the text of the tweet offers competing positive and negative attributions of responsibility for the event (COVID-19 in Argentina and security in Brazil and Mexico). Respondents are exposed to one of four possible combinations, introducing small variations to the wording of the positive and negative messages to ensure they are not strictly identical. These small variations minimize experimental detection by respondents. Third, we (3) randomize images that reinforce or undermine the partisan interpretation of text of the Tweets. Three pictures are rotated to ensure that pairs of Tweets always display different images: Congruent-Incongruent, Congruent-Placebo, Incongruent-Placebo. Finally, (iv) we randomize the numbers of likes and retweets at the bottom of the message to indicate high or low support by peers: high-High, High-Low, Low-High, Low-Low.

Since all attributes are independently randomized, causal effects are jointly estimated using simple OLS regression models (Hainmueller et al., 2014). We provide full details in the Appendix A of the online SIF. The frame elements (four variables) are combined into 144 different random pairings of tweets. Power requirements for the experiment are not stringent, given that the four frame element rotate a total of 15 attributes. The survey samples have 2,442 respondents in Argentina; 2,417 in Brazil; and 2,373 in Mexico, comfortably exceeing power requirements.

Our survey includes a variety of instruments on political attitudes and vote preferences. In

particular, we measure vote intention, partisanship, and a variety of social media engagement questions that explain differences in the rate of content activation.

4.2 Hypothesis

The experimental design randomly rotates the frame elements and measures differences in sharing behavior among respondents. While the *input* frequencies are uniform (equal probability), the *output* frequencies of the frame elements are modulated by the preferences of the respondents. We expect ideologues and partisans to share more content ("selection effect"). We expect the preferred content of partisans to be over represented in the experimental data ("composition effect"). Finally, we expect the content shared by partisans to be different from that of non-partisans (Fotouhi et al., 2014; Jo and Eom, 2014; Feld, 1991; Benevenuto et al., 2016; Eom and Jo, 2014).

The first hypothesis of our study measures whether partisans share more content than non-partisans. This hypothesis has been amply discussed in the literature (Weaver, 1991). If activation (attention) and partisanship are positively correlated, then content shared in social media will appear to be more polarized than it actually is. The preferences of intense ideologues would be over represented in the data and contribute to heightened perceptions of polarization. A test of this finding using observational data was reported by Aruguete et al. (2021).

The literature on affect and polarization shows that partisans and ideologues are unconditionally more motivated to participate in politics and in social media (Barberá, 2020; Slothuus and De Vreese, 2010; Guess et al., 2021). A recent study by Osmundsen et al. (2021) also describes partisan effects in fake news sharing, with larger increases in the likelihood of social media news among respondents that are more attentive to issues raised by their parties. Indeed, motivated reasoning that seeks to validate negative and positive evaluations of political events

among partisans not only increase attention to particular types of evidence but will also be more enthusiastic in communicating this information. A recent study of observational social media data in Argentina, Brazil, and the United States by Aruguete et al. (2021) supports that ideology and attention are highly correlated in observational social media data, with the preferred content of ideologues more frequent than the content of non-ideologues. Accordingly, we expect partisanship and social media sharing to be closely connected. Therefore, the hypothesis to be tested states that:

The expected correlation between ideological preferences and attention to issues is predicated of differences in motivated reasoning and hot cognition (Slothuus and De Vreese, 2010; Lelkes et al., 2017), where information that validates existing beliefs is more readily searched and shared by ideologues. If negative and positive evaluations of political events result in voters seeking and delivering information that is consistent with their preferences, motivated voters will be both more enthusiastic as well as more attuned to particular types of evidence, which will positively correlate ideological beliefs and issue attention (Weaver, 1991). Stronger partisan priming will also result in faster memory retrieval (Kahneman, 2011), which is another marker of *hot cognition* that is expected to increase sharing. Therefore, the hypothesis to be tested states that:

H_1*: Partisan users will be unconditionally more issue motivated than non-partisan voters to share cognitively congruent political content.*

The first hypothesis, H_1, expects partisan content to be more readily shared in social media and, consequently, over represented in observational data. The second hypothesis connects partisan respondents with frame elements. In effect, bubbles are expected because partisans share more and because they have distinct taste for the type of content they share. Conservative

voters are more likely to share content from conservative newspapers, such as Fox News (USA), La Nacion (Argentina), OAntagonista (Brazil), or Reforma (México). As the conjoint experiment guarantees that there is no topological sorting, differences in sharing distinct partisan content could only be explained by differences in the sharing behavior of these attentive partisans. The hypothesis aligns with evidence that explains social media bubbles by changes in the frequency of content shared by partisans (Del Vicario et al., 2016; Barberá and Rivero, 2015). Similar results are shown in signal processing and machine learning, where the amplification of weak signals reduces total variance in what is known as "boosting". Therefore, our second hypothesis:

H_2: Users will share congruent content that aligns politically with the preferences of their co-partisans (in-group cognitive congruence), reducing the stochastic variance in the initial frame elements.

Our final hypothesis is derived from the family of phenomena knows as the *friendship or class size paradox*, with an increase in the variance of partisan content resulting in subjective perceptions of partisanship that are larger than its overall frequency. In Aruguete and Calvo (2018) this is described as the "compositional effect" of activation from the viewpoint of the observer. In terms of Feld (Feld, 1991, pp.1475), the average users experiences more partisan content than the prevalence of partisans. The experimental results align with the formal treatment by Sikder et al. (2020), where is enough a "small group of individuals to generate permanent opinion polarization and cascade dynamics" (Sikder et al., 2020, pg. 1) once confirmation bias is formally linked to social connectivity. A similar discussion is presented by Saxena et al. (2019) when considering the level of activity of a node. They present the question as a thought experiment on an organization that is interested in the value of a highly connected node that is not very active and a less connected node that is very active. They ask whether the best choice to prop-

agate a message is "a node with higher degree but lower activity level or a node with lower degree but higher activity level?" (Saxena and Kumar, 2019, pp.40). In the political arena, partisans should observe higher levels of partisanship among their friends than non-partisans if their partisan friends are more active. Therefore, our third hypothesis:

H_3: Partisan users expect their friends to share more partisan frames.

Having summarized the theory behind all three hypothesis, we will now describe the experimental design. As explained before, we expect H_1 and H_2 to increase the sharing of partisan content in our experimental design (and in social networks) while H_3 increases the expectations of observing partisan content among our networks of friends.

4.3 Variables

The main dependent variables measure the decision to share each of the paired tweets by a respondent (self) and the expectation that the friends of the respondent will share each of the tweets (friends). For the first question, *self*, the variable takes the value of 1 if the respondent indicates is preference to share a tweet and 0 otherwise. Respondents can share both tweets, tweet 1, tweet 2, or neither.[2] The second variable takes the value of 1 if the respondent expects friend to share a tweet and zero otherwise.

In addition to our conjoint features, we separate our results between partisans and non-partisans users. We measure partisanship in two different ways. First, we measure the self reported partisan identification for the major parties in Argentina, Brazil, and Mexico. Second, we measure vote choice "if the election were to take place next week", allowing for the option to

[2]A separate question asks which tweet you would be less likely to share, forcing the choice of a single one of those tweets. This would be the traditional design in a conjoint experiment, but forcing all respondents to select one of the two tweets would not allow us to measure the frequency of activation. Therefore, our design allows for both tweets and neither to be shared, allowing a direct measure of the frequency of activation of distinct frame elements

vote blank. Therefore, we are able to compare both the difference in frequencies among individuals that report a partisan identification and also for individuals that vote the different parties. We consider partisans users who reported prefer/vote for any of the political parties listed in our survey, and non-partisans respondents who reported not having a partisan preference, or voting blanc in the last presidential election in Argentina, Brazil, and Mexico.

Given the fully randomize nature of our experiment, and proper balance as described in the SIF file, results are conclusive and do not require further controls. However, the supplemental information file presents models that include a variety of controls for readers interested in the effect of socio-demographic covariates that alter the likelihood of sharing tweets.

5 Results

In this section we present two different and critical results. First, we show that partisans are more likely to share tweets (self) and that partisans are more likely to expect that their friends will share tweets (friends). Therefore, the first set of results validate H_1 and shows that in experimental data the preferences of partisans are more broadly shared.

Second, we present results which shows that partisans share different frames than non-partisans H_2 and that they expect their friends to also share partisan frames H_3.

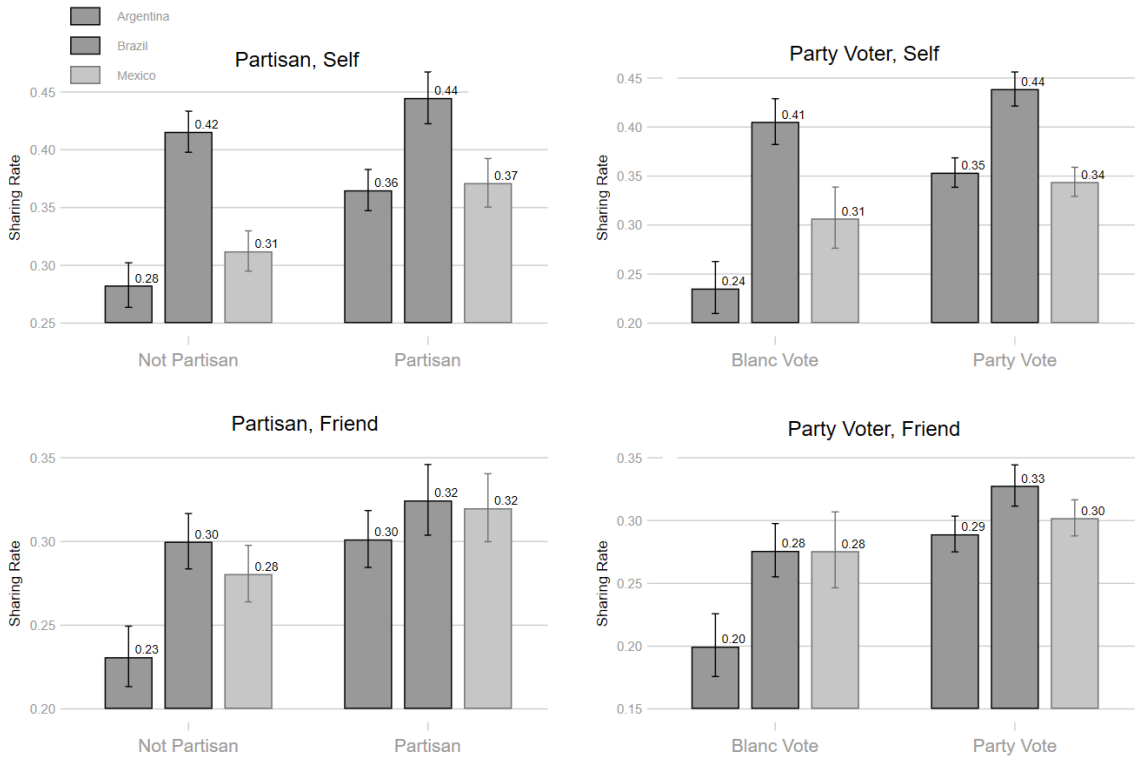
5.1 Higher Activation: Test of H_1

In Figure 1 we present results that test for the difference in overall activation by partisans and non-partisans as well as for party voters and blanc voters. Figure 1(a) reports the likelihood of sharing tweets for partisans and non-partisans in all three countries. Figure 1(b) reports the likelihood of sharing tweets for the voters of parties rather than those that voted blanc. Figure

1(c) reports the expectation that the friends of partisans and non-partisans will share these tweets. Finally, Figure 1(d) reports the expectation that the friends of the different party voters or those who vote blanc will share tweets.

In all four plots results show higher rates of sharing among partisans and their friends as well as higher rates of sharing among those respondents who voted for a party and their friends. Approximately 36% of partisans in Argentina indicated their preference to share tweets compared to 28% of non-partisans, a statistically significant increase of 8 points. Similarly, the expectation of content being shared by the friends of partisans is 7 points larger, increasing from 23% to 30%. Differences among voters are even larger, a total of 11 points for self and 9 points for friends.

Figure 1 Effect of Partisan and Voter on Sharing



Note: Figure 1(a) describes partisans on self. Figure 1(b) describes partisans on Friends. Figure 1(c) describes party voters on self. Figure 1(d) describes party voters on friends.

Results are more modest but also statistically significant for all comparisons in Brazil and Mexico. In Brazil, the increase in sharing is only of 2 and 3 points respectively for partisans and their friends. However, the differences are statistically significant. Higher sharing is also observed among party voters, with a 3 point difference for self, and 7 points expected for the party voters' friends. In Mexico, sharing increases are 6 points and 4 points for partisans and their friends, and 3 points and 2 points for party voters and their friends. This last coefficient, the 2 point difference observed among the friends of party voters in Mexico, is the only one that fails to reach statistical significance.

In the appendix, we present numerical results for statistical models regressing our partisanship variables and the decision to share and the expectation about the respondents' friends. Across all the models, partisans show higher rates of sharing and also more pronounced expectation about the sharing behavior of their friends.

5.2 Congruent Partisan Sharing, H2 and H3

In this section, we present results showing that partisans share different frames than non-partisans and also expect their friends to also share partisan frames. To make the presentation easier, we focus only on the variable vote choice to identify partisans in the three countries. Our quantity of interest is the difference in marginal means for every feature in our three conjoint experiments between partisans and non-partisans. We focus on the marginal means, instead of the more heavily used Average Interactive Component Effect (Hainmueller et al., 2014), because these quantities are more appropriate to identify heterogeneous, subgroup effects when dealing with conjoint designs (Leeper et al., 2020). In addition, we separate the results between leftists and conservatives partisans using the vote choice independent variable [3]. This decision allows

[3]The question asks: 'if the election were to take place next week, who would you vote for?"

us also to observe the directional effects of the frames between distinct partisan groups.

Figures 2 presents the results for the respondents decision to share. The point estimates in each figure indicate the difference in sharing rate between leftists/conservatives and non-partisans in Argentina, Brazil, and Mexico. The x-axis contains the features (frame elements) that are embedded in our social media image-based experiment. Positive point estimates show that leftists/conservative respondents have a higher propensity to share one particular frame element averaging across all the other features. This quantity of interest uses the non-partisan groups as a baseline.

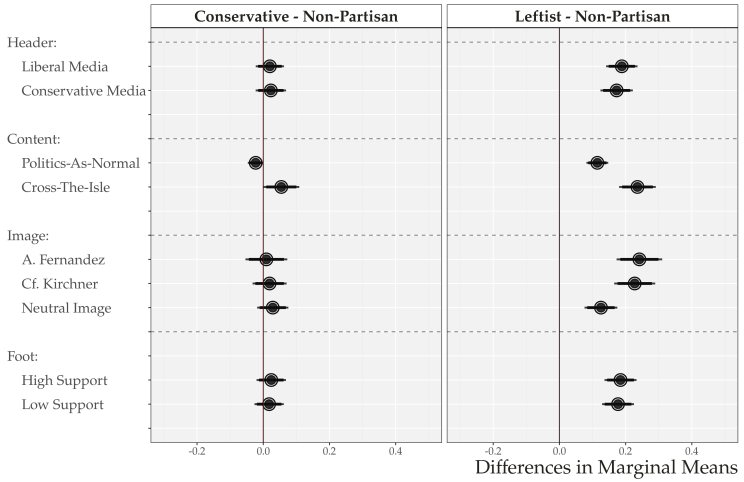
All the models show partisans, from the left and right, are more likely to share congruent social media messages when compared to non-partisans. In Argentina, conservative respondents are 5% more likely to share a tweet where the actual Argentinean President Fernandez (from the left) sends a message crossing-the-isle and signalling to the opposition about a national front to fight the Covid-19 crises. Meanwhile, the leftists voters are more likely to share both contents, one with the cross-the-isle message, but also a message blaming the previous government for the health crises in Argentina.

Similar patterns, when considering only the content of the social media messages, appear in Brazil and Mexico. Conservative Brazilians, who support President Bolsonaro, are more likely to share in-group messages, which calls for a more punitive security policies to reduce crime in Brazil, and less likely to share messages calling for more welfare policies. The oppose trend appear among leftists voters. Meanwhile, leftists Mexicans, supporters of the actual Incumbent, show higher propensity to share a general framing message about the crime issue, and lower propensity on messages blaming the actual administration for the rise in violence in the country.

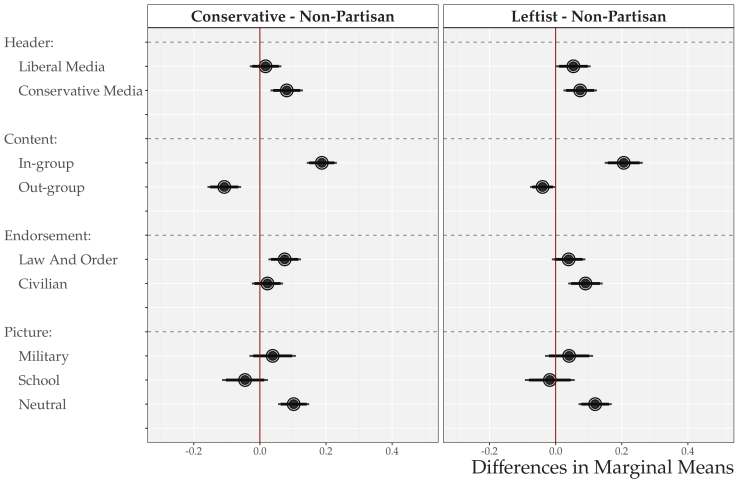
Importantly, the reference groups for all these models are independent, non-partisan voters.

Therefore, these differences in the propensity to share show how congruent frames make partisans more active in social media, when compared to moderate, independent voters. These robust findings across three different countries provide conclusive experimental evidence for the formation of bubbles on sharing behavior, and explain the representation of partisan content on social media (Del Vicario et al., 2016; Barberá and Rivero, 2015).

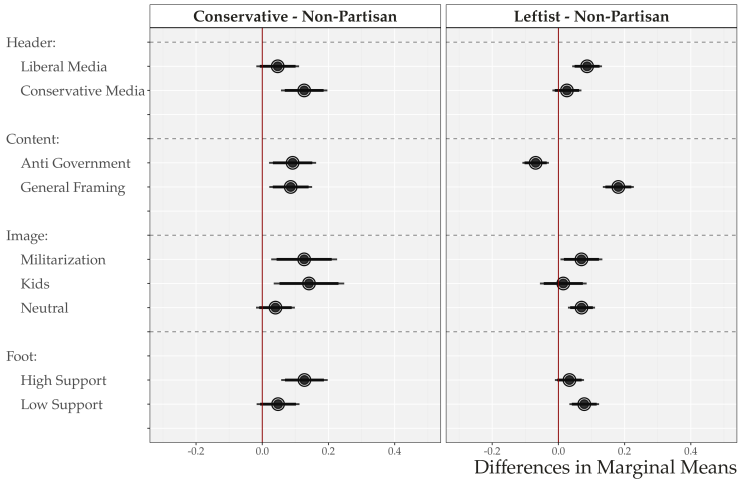
Figure 2 Congruent Partisan Sharing: Network Activated Frames



a) Argentina



b) Brazil



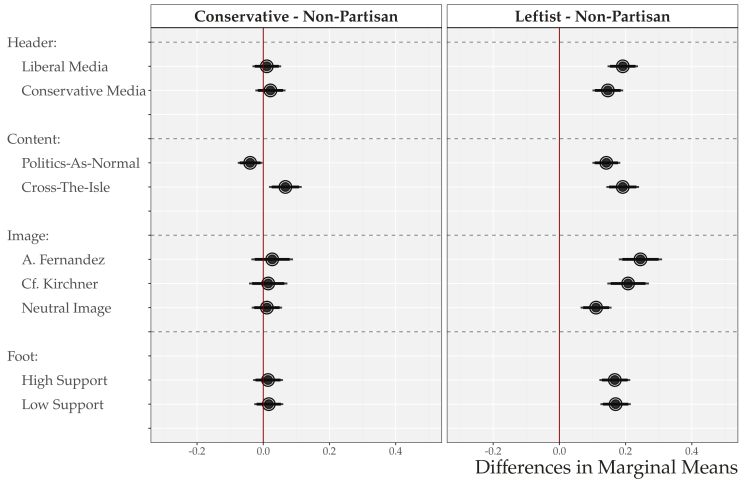
c) Mexico

We now present results for the respondents' expectation about the sharing behavior of their friends behavior. Our results provide strong evidence for the NAF behavior on social media. Figure 3 provides a similar interpretation as in figure 2; conjoint features (frame elements) for the three countries are presented in the x-axis and differences in sharing rate (marginal means) are presented on the y-axis with point estimates and confidence intervals. However, instead of focusing on behavior, figure 3 focuses on respondents' expectations about sharing behavior of their friends regarding each tweet. Positive point-estimates in figure 3 indicates leftists/conservatives expect their friends to, on average, share more a particular frame element than non-partisans expect their friends to do.

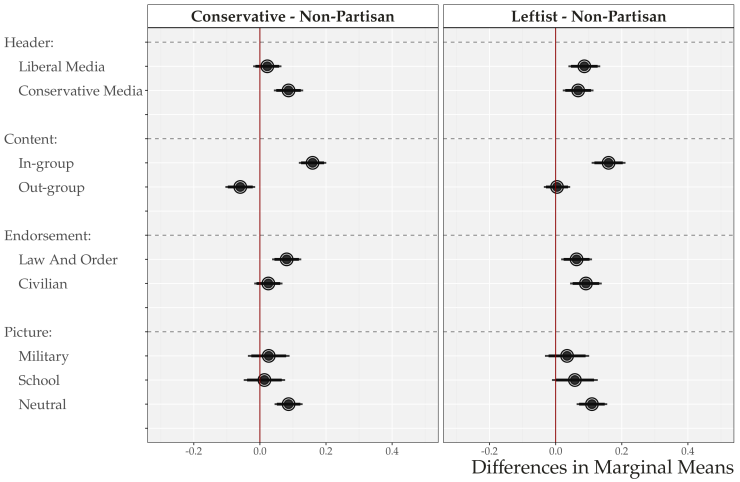
Across all the three cases, partisan respondents expect higher levels of partisanship among their friends than non-partisans do. Now, instead of looking directly at the content of the conjoint features, let us consider the effect of the feature Header, which shows a more liberal and more conservative news media as the author of the tweet.

In all the three cases, partisanship aligns closely to the expectation about which outlet ones' friend would share. Leftists in Brazil, Argentina and Mexico believe their friends show higher sharing rates for social media messages sent by in-group media outlets, while conservative, for most cases, have the exact opposite expectation. As before, differences in the marginal means for the contents follow closely to the ones discussed in figure 2.

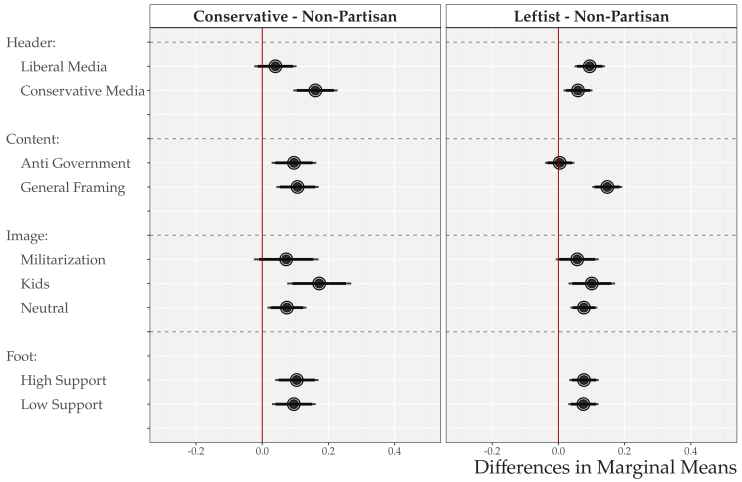
Figure 3 Friends Congruent Sharing: Network Activated Frames



a) Argentina



b) Brazil



c) Mexico

In the appendix, we compare estimates between sharing behavior and expectation about the respondents friends. For the majority of the cases, there are meaningful statistical differences between the respondents' behavior and what They expect from their friends.

6 Conclusion

How does the sharing behavior of interconnected users frames political events? How do partisans and non-partisans alter the frequencies of the texts, images, and endorsements we observe in social media? This article provides clear experimental evidence that a social network with input frame elements with random uniform probabilities will output local frames that will over-represent the preferences of partisan respondents. In doing so, we provide conclusive evidence that a network with cross-cutting ideological friends will still produce "bubbles".

The only two conditions that are required is for partisans to share at higher rates than non-partisans and for partisans to have different preferences than those of non-partisans. Beginning with uninformative probabilities, changes in the frequency of distinct frame elements, in different regions of a social network, shape how individuals interpret, classify, and define situations and events. The proposed conjoint experiments describe the mechanisms behind Network Activated Frames (NAF) replicating observational findings of network activation in three distinct surveys. The results of this study also show similar effects when analyzing sharing behavior (self) as well as the expectation of activation by peers (friends). Our partisan self shares more content than non-partisans and the preferences of our partisan friends are expected to be over-represented in social media.

Our study provides several important contributions. We develop a novel theoretical explanation for the mixed-finding relating the formation of echo chambers in social media and the

lack of empirical evidence of users' sorting on social networks and media diets (Barberá et al., 2015; Guess et al., 2021; Bakshy et al., 2015). Instead of focusing on sorting, our theory focuses on how social media bubbles emerge from different propensities of partisan users' share content, and their expectations about homophily on their personal networks. This explanation helps our understanding about the reasons why perceptions of polarization are so heightened among social media users, even though the levels of segregation on users' social media networks and media consumption are not particularly high.

Second, we implement a novel research design that serves as a methodological contribution to future studies on social media effects. We show how we can easily combine the methodological advantages of factorial experiments with an image-based implementation that provides high-ecological validity for social media studies. By construction, these designs are more flexible when compared to commonly deployed social media framing experiments, since researchers can manipulate several theoretically relevant features (Hainmueller et al., 2014). In addition, the image-based design provides a more realistic environment in which users interact with "real" tweets resembling their normal daily activities, and providing potential gains on ecological validity for our experiment (Vecchiato and Munger, 2021; Thal, 2020).

Third, our study expands the literature on social media, polarization, and the formation of social media bubbles to a comparative perspective. The lack of empirical studies about this topic outside of the US context has long been warned by the scholarship (Barberá, 2020; Tucker et al., 2018). As noted recently by Mitchelstein and Boczkowski (Mitchelstein and Boczkowski, 2021), the dominance of the Global North on communications and social media studies has pernicious consequences, and in their words: "not only reproduces and reinforces inequalities but also results in inferior scholarship" (pp. 132). To the best of our knowledge, we are the

first to implement three similar conjoint experimental designs on a cross-national perspective with a focus on social media filter bubbles and content activation. Our results are robust across the three countries, provide high external validity for our initial hypotheses, and contribute with high-quality, and more diverse empirical evidence for a topic that has received considerable attention in the last years by political communication scholars.

Our results are not unique to social media networks. Decades of research in neurobiology research show that neural networks encode information by increasing or decreasing their firing rates (Humphries, 2021). While it may be true that "neurons that wire together, fire together", the topology of a network provides limited information on how memories are formed or decisions are made. In social networks, an over emphasis on connectivity ("who is in our network") does not necessarily inform on activation ("what is shared") or how individuals interpret, classify, and define situations and events ("how is this framed"). An experimental understanding of how users activate frame elements is a required step to understand how social networks compose local frames. Consequently, we expect readers to interpret the experimental results in this article as an important stepping stone to model activation in complex networks. That is, in a context in which the initial probabilities of observing and sharing a frame element are not simply at random.

References

Abramowitz, A. I. and Saunders, K. L. (2008). Is polarization a myth? *The Journal of Politics*, 70(2):542–555.

Alipourfard, N., Nettasinghe, B., Abeliuk, A., Krishnamurthy, V., and Lerman, K. (2020). Friendship paradox biases perceptions in directed networks. *Nature communications*, 11(1):1–9.

Aruguete, N. (2019). Network-activated frames (naf), redefining framing in a new digital era. *Encyclopedia of Educational Innovation, Springer Nature: Singapore*.

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.

Aruguete, N., Calvo, E., and Ventura, T. (2021). News by popular demand: Ideological congruence, issue salience, and media reputation in news sharing.

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. F., Lee, J., Mann, M., Merhout, F., and Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37):9216–9221.

Bakshy, E., Messing, S., and Adamic, L. (2015). Exposure to ideologically diverse news and opinion on facebook. *science*, 348(6239):1130–1132.

Banks, A., Calvo, E., Karol, D., and Telhami, S. (2021). # polarizedfeeds: Three experiments on polarization, framing, and social media. *The International Journal of Press/Politics*, 26(3):609–634.

Barberá, P. (2020). Social media, echo chambers, and political polarization. *Social Media and Democracy: The State of the Field, Prospects for Reform*, page 34.

Barberá, P. and Rivero, G. (2015). Understanding the political representativeness of twitter users. *Social Science Computer Review*, 33(6):712–729.

Barberá, P., Wang, N., Bonneau, R., Jost, J. T., Nagler, J., Tucker, J., and González-Bailón, S. (2015). The Critical Periphery in the Growth of Social Protests. *PLOS ONE*, 10(11):e0143611.

Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., and Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological science*, 26(10):1531–1542.

Benevenuto, F., Laender, A. H., and Alves, B. L. (2016). The h-index paradox: your coauthors have a higher h-index than you do. *Scientometrics*, 106(1):469–474.

Conover, M. D., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., and Flammini, A. (2011). Political polarization on twitter. In *Fifth international AAAI conference on weblogs and social media*.

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H. E., and Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3):554–559.

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

Entman, R. M. and Usher, N. (2018). Framing in a fractured democracy: Impacts of digital

technology on ideology, power and cascading network activation. *Journal of Communication*, 68(2):298–308.

Eom, Y.-H. and Jo, H.-H. (2014). Generalized friendship paradox in complex networks: The case of scientific collaboration. *Scientific reports*, 4(1):1–6.

Feld, S. L. (1991). Why your friends have more friends than you do. *American Journal of Sociology*, 96(6):1464–1477.

Fiorina, M. P. and Abrams, S. J. (2008). Political polarization in the american public. *Annu. Rev. Polit. Sci.*, 11:563–588.

Fishkin, J. S. (1991). *Democracy and deliberation: New directions for democratic reform*. Yale University Press.

Fotouhi, B., Momeni, N., and Rabbat, M. G. (2014). *Generalized Friendship Paradox: An Analytical Approach*, pages 339–352. Springer.

Franchino, F. and Zucchini, F. (2015). Voting in a multi-dimensional space: A conjoint analysis employing valence and ideology attributes of candidates. *Political Science Research and Methods*, 3(2):221.

Gentzkow, M. and Shapiro, J. M. (2011). Ideological segregation online and offline. *The Quarterly Journal of Economics*, 126(4):1799–1839.

Gidron, N., Adams, J., and Horne, W. (2019). Toward a comparative research agenda on affective polarization in mass publics. *APSA Comparative Politics Newsletter*, 29:30–36.

Guess, A. M., Barberá, P., Munzert, S., and Yang, J. (2021). The consequences of online partisan media. *Proceedings of the National Academy of Sciences*, 118(14).

Habermas, J. (1991). *The structural transformation of the public sphere: An inquiry into a category of bourgeois society*. MIT press.

Hainmueller, J. and Hopkins, D. J. (2015). The hidden american immigration consensus: A conjoint analysis of attitudes toward immigrants. *American Journal of Political Science*, 59(3):529–548.

Hainmueller, J., Hopkins, D. J., and Yamamoto, T. (2014). Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Political analysis*, 22(1):1–30.

Humphries, M. (2021). *The Spike: An Epic Journey Through the Brain in 2.1 Seconds*. Princeton University Press.

Iyengar, S., Sood, G., and Lelkes, Y. (2012). Affect, not ideologya social identity perspective on polarization. *Public opinion quarterly*, 76(3):405–431.

Jackson, M. O. (2019). The friendship paradox and systematic biases in perceptions and social norms. *Journal of Political Economy*, 127(2):777–818.

Jo, H.-H. and Eom, Y.-H. (2014). Generalized friendship paradox in networks with tunable degree-attribute correlation. *Physical Review E*, 90(2):022809.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kirkland, P. A. and Coppock, A. (2018). Candidate choice without party labels. *Political Behavior*, 40(3):571–591.

Klašnja, M., Lupu, N., and Tucker, J. A. (2021). When do voters sanction corrupt politicians?–corrigendum. *Journal of Experimental Political Science*, pages 1–2.

Lee, E., Karimi, F., Wagner, C., Jo, H.-H., Strohmaier, M., and Galesic, M. (2019). Homophily and minority-group size explain perception biases in social networks. *Nature human behaviour*, 3(10):1078–1087.

Leeper, T. J., Hobolt, S. B., and Tilley, J. (2020). Measuring subgroup preferences in conjoint experiments. *Political Analysis*, 28(2):207–221.

Lelkes, Y., Sood, G., and Iyengar, S. (2017). The hostile audience: The effect of access to broadband internet on partisan affect. *American Journal of Political Science*, 61(1):5–20.

Mansbridge, J. J. (1983). *Beyond adversary democracy*. University of Chicago Press.

Mares, I. and Visconti, G. (2020). Voting for the lesser evil: evidence from a conjoint experiment in romania. *Political Science Research and Methods*, 8(2):315–328.

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

Mitchelstein, E. and Boczkowski, P. J. (2021). What a special issue on latin america teaches us about some key limitations in the field of digital journalism. *Digital Journalism*, 9(2):130–135.

Oliveros, V. and Schuster, C. (2018). Merit, tenure, and bureaucratic behavior: Evidence from a conjoint experiment in the dominican republic. *Comparative Political Studies*, 51(6):759–792.

Osmundsen, M., Bor, A., Vahlstrup, P. B., Bechmann, A., and Petersen, M. B. (2021). Partisan polarization is the primary psychological motivation behind political fake news sharing on twitter. *American Political Science Review*, pages 1–17.

Reese, S. D., Gandy Jr, O. H., and Grant, A. E. (2001). *Framing public life: Perspectives on media and our understanding of the social world*. Routledge.

Saxena, B. and Kumar, P. (2019). A node activity and connectivity-based model for influence maximization in social networks. *Social Network Analysis and Mining*, 9(1):1–16.

Settle, J. E. (2018). *Frenemies: How social media polarizes America*. Cambridge University Press.

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

Slothuus, R. and De Vreese, C. H. (2010). Political parties, motivated reasoning, and issue framing effects. *The Journal of Politics*, 72(3):630–645.

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

Suhay, E., Bello-Pardo, E., and Maurer, B. (2018). The polarizing effects of online partisan criticism: Evidence from two experiments. *The International Journal of Press/Politics*, 23(1):95–115.

Sunstein, C. and Sunstein, C. R. (2018). *# Republic*. Princeton university press.

Thal, A. (2020). The desire for social status and economic conservatism among affluent americans. *American Political Science Review*, 114(2):426–442.

Tucker, J. A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D., and Nyhan, B. (2018). Social media, political polarization, and political disinformation: A review of the scientific literature. *Political Polarization, and Political Disinformation: A Review of the Scientific Literature (March 19, 2018)*.

Vecchiato, A. and Munger, K. (2021). Visual conjoints: Twitter profiles.

Weaver, D. (1991). Issue salience and public opinion: Are there consequences of agenda-setting? *International Journal of Public Opinion Research*, 3(1):53–68.

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