# Network Activated Frames:
# Content Sharing and Perceived Polarization in Social Media

Natalia Arugute
Ernesto Calvo
Tiago Ventura

September 21, 2022

***Forthcoming at the Journal of Communication***

**Abstract**

Our paper describes how users' decisions to share content alter the frequencies of the frame elements observed by social media peers. Changes in the frequency of distinct frame elements 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 differ from those of non-partisans. Our findings show that a network of peers with cross-cutting ideological preferences may be perceived as a bubble if partisans amplify content they like at higher rates. Beginning with fully randomized probabilities, the output from our experiments is more extreme than the preferences of the median users, as partisans activate more and different frame elements than non-partisans. We implement the experiments in Argentina, Brazil, and Mexico.

**Number of words**: 7,942

**Keywords: Framing, Social Media, Polarization, Conjoint Experiments**

# Network Activated Frames: Content Sharing and Perceived Polarization in Social Media

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 observed in a network (Aruguete & Calvo, 2018; Entman & Usher, 2018). We define this process as *Network Activated Frames* (NAF), whereby users *frame events* by *sharing content* in social networks.

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 users' decisions to share content and the content users expect to see activated by their friends is critical for explaining framing in social media.

Most importantly, differences in the user's decision to share content (self) and in the content user's expect to see shared by others (peers), allow us to explain the subjective perceptions of *social media bubbles* (Alipourfard et al., 2020; Jackson, 2019; Lee et al., 2019), where the preferences of more active partisan nodes are overrepresented in the data. In this paper we show that if partisans share *more* frames and if partisans share *different* frames, users will perceive social media networks as highly partisan. We provide evidence of this effect through a fully randomized experiment, where *input* frames are presented to respondents with equal probability. *Output* frames, however, overrepresent the preferences of partisan respondents.

Our findings clarify the source of conflicting evidence on the existence of social media bubbles (Bakshy et al., 2015; Barberá et al., 2015; Barberá, 2020). They also converge with recent research that explores the topological effects of confirmation bias in social networks (Sikder et al., 2020), which only require a small group of very active nodes to dominate data in networks. Findings also align with new studies that distinguish between the connectivity effects (high in-degree) and activation effects (high frequency sharing) of users in networks (Saxena & Kumar, 2019). We document statistically significant differences in the frequency ("partisans share more") and in the type of frame elements ("partisans share different") activated by partisan and nonpartisan users in a fully randomized experiment.

We design our novel image-based conjoint experiment to test for quantitative and qualitative differences in the frames shared by partisan and nonpartisan respondents. We implement three conjoints with pairs of tweets that rotate the different frame elements: *author*, *text*, *images*, and number of *likes* of paired posts. 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 on 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 first section of this article describes how NAF behavior explains the subjective perception of social media bubbles, echo-chambers, and polarization. The second section revisits alternative models of frame activation in social media and describes the statistical connections between frame activation and the *Friendship Paradox* in social networks

(Feld, 1991; Sikder et al., 2020). The third section describes our experimental design and hypotheses, highlighting the difference between activation by users (*self*) and the expected activation by friends (*peers*). The fourth section describes the results of our experiments in Argentina, Brazil, and Mexico. Findings validate three key hypotheses: first, partisan users share *more* content; second, partisan users share *different* frames than non-partisans; and, third, partisans expect their friends to share partisan content. As a result, partisans' preferences are overrepresented in the *output* data. We conclude with a discussion of why previous studies reported mixed findings when describing social media bubbles. We also suggest promising future extensions of this research.

## Network Activated Frames, Social Media Bubbles, and Polarization

### Bubbles

Few phenomena are as characteristic of our times as political polarization (Abramowitz & Saunders, 2008; Gidron et al., 2019; Iyengar et al., 2012; Mason, 2018). In the scholarship that developed to explain the causes of polarization, researchers frequently ask whether social media and the emergent digital technologies contribute to heightened levels of perceived polarization (Bail et al. 2018; Lelkes et al., 2017; Settle, 2018; Stroud, 2010; Sunstein, 2018).

A popular hypothesis that connects social media usage and mass polarization focuses on the potential of new technologies to generate echo chambers or filter bubbles. These phenomena involve delivering content that reinforces existing partisan animosity through motivated reasoning and online sorting. Cass Sustein (2018) forcefully advanced this theory when he argued that we are more likely to connect with users that consume

information we like and users that share our beliefs. The result is *online sorting* and *network homophily*, whereby citizens are exposed to higher doses of pro-attitudinal opinions that reinforce their existing preferences and attitudes. In contrast, people are less likely to observe counter attitudinal, screening out users who do not share their opinions. In line with a long tradition of deliberative democracy (Fishkin, 1991; Habermas, 1991; Mansbridge, 1983), polarization is a consequence of the lack of ideological diversity that characterizes life in online echo chambers and filter bubbles (Mason, 2018; Sunstein, 2018).

While the argument remains popular, recent research finds little support for the *sorting* model of echo chambers in social media. There is robust empirical evidence that the levels of online segregation detected in early studies (Bakshy et al., 2015; Conover et al., 2011) are not particularly different from offline media consumption (Gentzkow & Shapiro, 2011). Recent studies also show that homophily in networks is limited to interactions related to political events (Barberá et al., 2015; Wojcieszak & Mutz, 2009). Finally, evidence is mounting that online news consumption, at least in the United States, is less segregated than previously thought, with users actively accessing centrist and high reputation news organizations that do not necessarily align with the user's ideology (Guess et al., 2021). Although homogeneity in social media networks is prevalent, cross-contamination is not infrequent.

Current research studying exposure to political information in social media presents researchers with an interesting puzzle. Most users are embedded in diverse social networks, yet they routinely describe subjective perceptions of widespread polarization and high doses of partisan content. This subjective experience also fits descriptions by experts and

policymakers (Barberá, 2020). Previous studies tackled this puzzle by showing that, contrary to more deliberative normative expectations, heterogeneity and exposure to uncivil content in users' networks might actually be one of the drivers of polarization (Bail et al., 2018; Banks et al., 2021; Suhay et al., 2018).

**Content Activation and the Friendship Paradox**

We advance an alternative explanation that is compatible with socially heterogeneous yet highly partisan social media experiences. We focus on whether diverse input content may still yield a highly partisan output. Our experiment shows that a network with a large number of moderate social media users and diverse input content is consistent with a social media experience that includes echo chambers. Instead of focusing on sorting, we consider differences in the activation of partisan and nonpartisan content. If partisans share *more* frames and partisans share *different* frames, their preferences and attitudes will be overrepresented in a social media network.[1]

Perceptions of high ideological congruence belong to the family of phenomena known 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, p. 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" (p. 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, first, it is necessary to show that (i) partisans share content with a higher frequency than non-partisans and their preferences are over-represented in observational data. Second, (ii) that partisans share content that is different from that of non-partisans. Whether individuals in a network will observe larger doses of partisan content does not depend on the diversity of the peer network. Therefore, it is possible to have a diverse network of peers and to see partisan content overrepresented in our social media feeds (Aruguete, 2019; Bakshy et al., 2015; Barberá, 2020; Barberá et al., 2015).

Our empirical work confirms both mechanisms. 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 with which they agree; and (ii) the preferred content of partisans and ideologues is different from the content preferred by non-partisans. We confirm these conditions when we analyze: i) sharing (self), and ii) sharing expectations (peers).

Support for the "generalized friendship paradox" is theoretically and empirically relevant (Benevenuto et al., 2016; Eom & Jo, 2014; Feld, 1991; Fotouhi et al., 2014; Jo & Eom, 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 (Barberá et al., 2015; Del Vicario et al., 2016). 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 & 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 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 as partisan as they are. By contrast, independent voters in Argentina, Brazil, and Mexico do not report that their friends share more partisan frames on social media. This is consistent with the central tenets of the *generalized friendship paradox*, with "bubble" like content being more prevalent only among those who share the selected trait (i.e. partisanship).

## 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, similar to 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.[2] 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., 2001:11). The latter is present in the different stages of the communication process. In social media, framing is the result of how users create and post content (production) and how peers activate content (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 prompts them to revise the initial model in favor of a 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, 2019; Aruguete & Calvo, 2018) seeks to update the notion of the integrality of the framing process, taking into account the dynamics of content propagation in a digital media and that virtual social networks have a prime role in the circulation of endorsements, texts, and images that structure the social world. *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 users, algorithms, 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 users are described as a "compositional effect," with different aggregate interpretations of 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, increase the probability of sharing a particular frame element ("selection effect"). Meanwhile, selective activation of congruent content by social media peers and the accumulation of certain frame elements at a given network location yield locally homogeneous frames, forming what we know as social media bubbles ("composition effect").

In the next section, we describe an experiment to measure the "selection effect" of frame elements by partisans. On the consumption side, this phenomenon produces 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).

## 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 across many distinct fields. Conjoint experiments provide treated individuals with two competing profiles with randomized traits (conjoint profiles). After exposure, subjects 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. It is important to note that we do not force the

selection of one of the two frames. We instead allow respondents not to share traits. This option 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 observes pairs of edited tweets created solely for the experiment. The messages replicate 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 randomly-created, paired tweets. While the frame elements and the issues in the conjoints 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.

Image-based conjoint experiments offer several advantages for measuring the connections between framing, social media activation, and partisanship. First, the fully randomized nature of conjoint experiments allows researchers to remove sorting effects that might endogenously contaminate research relying on behavioral social media data. Second, as argued by Hainmueller et al. (2014), conjoint designs allow researchers to manipulate many different features and identify treatment effects simultaneously. Because social media effects can emerge from many factors (author of a tweet, text, images, social support, among others), the flexibility of this design makes it ideal for understanding media

effects. Finally, when incorporated together with real social media images, visual rotations of the conjoint provide a more realistic environment in which participants make decisions. Previous research has shown important gains in ecological validity when experiments provide more realistic environments (Horiuch et al., 2021; Thal, 2020; Vecchiato & Munger, 2021).

### 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 observe one of four possible combinations, introducing small wording variations of the positive and negative messages to ensure they are not 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.

**<<Figure 1, Conjoint Example>>**

We provide full details in Appendix A of the online SIF. The survey samples have 2,442 respondents in Argentina; 2,417 in Brazil; and 2,373 in Mexico.

## Hypotheses

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 overrepresented in the experimental data ("composition effect"). Finally, we expect the content shared by partisans to be different from that of non-partisans (Benevenuto et al., 2016; Eom & Jo, 2014; Feld, 1991; Fotouhi et al., 2014; Jo & Eom, 2014).

The first hypothesis of our study measures whether partisans share more content than non-partisans. If activation (attention) and partisanship are positively correlated, then content shared on social media will appear to be more polarized than it actually is. The preferences of intense ideologues would be overrepresented in the data and contribute to heightened perceptions of polarization. Aruguete et al. (2021) report this finding using observational data.

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; Guess et al., 2021; Mason, 2018; Slothuus & De Vreese, 2010; Törnberg, 2018). 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 causes increases in attention to particular types of evidence, but also causes more enthusiasm in communicating the 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.

The expected correlation between ideological preferences and attention to issues is predicated on differences in motivated reasoning and hot cognition (Lelkes et al., 2017; Slothuus & De Vreese, 2010), where ideologues more readily search for and share information that validates existing beliefs. If negative and positive evaluations of political events result in voters seeking and delivering information consistent with their preferences, motivated voters will be both more enthusiastic and more attuned to particular types of evidence, which will positively correlate ideological beliefs and issue attention (Weaver, 1991). Partisan priming will also elicit affective, automatic, and fast memory retrieval (Kahneman, 2011), which are markers of hot cognition that we expect to increase sharing. Therefore, we the following hypothesis:

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

The first hypothesis, $H_1$, expects partisan content to be more readily shared in social media and, consequently, overrepresented in observational data. The second hypothesis connects partisan respondents with frame elements. We expect bubbles because partisans share more and because they have a distinct taste for the type of content they share. Conservative voters are more likely to share content from conservative media sources, such

as Fox News (USA), La Nación (Argentina), OAntagonista (Brazil), or Reforma (Mexico). 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 (Barbera & Rivero, 2015; Del Vicario et al., 2016). 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:

*H2: 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 comes from the family of phenomena known as the *friendship or class size paradox*, with an increase in the variance of partisan content resulting in subjective perceptions of bubble that are larger than its actual frequency. In Aruguete and Calvo (2018) this is described as the *composition effect* of activation from the viewpoint of the observer. In terms of Feld (Feld, 1991, p. 1475), the average user experiences more partisan content than the prevalence of partisans. The experimental results align with the formal treatment by Sikder et al. (2020), where it is enough for a "small group of individuals to generate permanent opinion polarization and cascade dynamics" (Sikder et al., 2020, p. 1) once confirmation bias is formally linked to social connectivity. A similar discussion is presented by Saxena and Kumar (2019) when considering the level of activity of a node. The authors present the question as a thought experiment: is it better to secure a highly connected node that is not very active or a less connected node that is very active? (Saxena

& Kumar, 2019: 40). Therefore, we test a third hypothesis:

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

Having summarized the theory behind all three hypotheses, we 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 we expect $H_3$ to increase perceptions of partisan content among networks of friends.

**Variables**

The main dependent variables measure the decision to share each of the paired tweets by a respondent (self) and the expectation that 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 their preference to share a tweet and 0 otherwise. Respondents can share both tweets, tweet 1, tweet 2, or neither.[3] The second variable takes the value of 1 if the respondent expects a friend to share a tweet and zero otherwise.

In addition to our conjoint features, we separate our results between partisans and non-partisan users. We measure partisanship in two different ways. First, we measure 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 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 define partisans as users who reported that they prefer/vote for any of the political parties listed in our survey. We defined non-partisans as respondents who reported not having a partisan preference, or voting blank in the last

presidential election in Argentina, Brazil, and Mexico.

Given the fully randomized 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 on sharing behavior. We added controls for age, gender, income, and education. The effects of these socio-demographic variables vary across countries. As expected, the inclusion of these controls does not alter the direction or significance of the estimates presented here.

## Results

In this section, we present two 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 validates $H_1$ and shows that in experimental data the preferences of partisans are more broadly shared. Second, we present results demonstrating that partisans share different frames compared to non-partisans, which support $H_2$, and that they expect their friends to also share partisan frames, which supports $H_3$.

### Higher Activation: Test of $H_1$

In Figure 2, we present results that test for differences in activation by partisans and non-partisans as well as for party voters and blank voters. Figure 2(a) reports the likelihood of sharing tweets for partisans and non-partisans in all three countries. Figure 2(b) reports the likelihood of sharing tweets for the voters of parties rather than those that voted blank.

Figure 2(c) reports the expectation that friends of partisans and non-partisans will share these tweets. Finally, Figure 2(d) reports the expectation that friends of the different party voters or those who vote blank will share tweets.

**<<Insert Figure 2>>**

In all four plots, results show higher rates of sharing among partisans and their friends and higher rates of sharing among those respondents who voted for a party and their friends compared to respondents who voted blank. 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.

Results are a bit more modest but also statistically significant for all comparisons in Brazil and Mexico. In Brazil, the increase in sharing is 2 and 3 points respectively for partisans and their friends. However, the differences are statistically significant. We observe higher sharing 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 appendix C, we present model results regressing our partisanship variables and the decision to share (*self* and *friends*). In addition to the two previous partisanship variables, we also include three models using ideology extremism (distance between respondents' ideology and the center of the ideology scale) as an explanatory variable. Summary effects

in Figure 3 show that the increases in content sharing by partisans are significant and of similar magnitude across all three countries. Similarly, the positive effect of ideological extremism on sharing is quite robust, with comparable effects in all experiments.

**<<Insert Figure 3>>**

### Congruent Partisan Sharing, H2, and H3

In this section, we present results showing that partisans share different frames compared to non-partisans and also expect their friends to share partisan frames. To make the presentation simpler, we focus 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 leftist and conservative partisans using the vote choice independent variable. This decision allows us also to observe the directional effects of the frames between distinct partisan groups.

**<< Insert Figure 4>>**

Figure 4 presents the results of 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 leftist/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 group 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 current Argentinian President Fernández (from the left) sends a message crossing-the-isle and signaling to the opposition about a national front to fight the Covid-19 crisis. Meanwhile, leftist voters are more likely to share both contents, one with the cross-the-isle message and the other with 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 call for more punitive security policies to reduce crime in Brazil, and less likely to share messages calling for more welfare policies. The opposite trend appears among leftists voters. Meanwhile, leftist Mexicans, supporters of the incumbent president, show a higher propensity to share a general framing message about the crime issue and a lower propensity to share messages blaming the current administration for the rise in violence in the country.

Contrary to our expectations, there is lower than expected discrimination between the pro- and anti-government frames shared by conservative respondents in Mexico. Results show that conservative voters who support the PRI and the PAN share "*more*" of both frames (the anti-government and the cross-the-aisle messages). Consequently, while conservative respondents contribute more information to the Mexican network, as

expected, they do not necessarily contribute information that increases polarization.

The reference groups for all 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 independent voters. These robust findings across three different countries provide conclusive experimental evidence for the formation of bubbles from sharing behavior. Results also explain the over-representation of partisan content on social media.

**<<Insert Figure 5>>**

We now present results for the respondents' expectations about their friends' sharing behavior. Our results provide strong evidence for NAF behavior on social media. Figure 5 provides a similar interpretation as in Figure 4; 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 4 focuses on respondents' expectations about sharing behavior of their friends regarding each tweet. Positive point estimates in Figure 4 indicate leftists/conservatives expect their friends to, on average, share more of a particular frame element compared to non-partisans.

Across all three cases, partisan respondents expect higher levels of partisanship among their friends compared to non-partisans. Now, instead of looking directly at the content of the conjoint features, we 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 three cases, partisanship aligns closely with the expectation about which outlet our 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 conservatives, in most cases, have the exact opposite expectation. As before, differences in the marginal means for the contents follow closely the ones discussed in Figure 4. In the appendix, we compare estimates between sharing behavior and expectation about the respondents' friends. For the majority of the cases, there are no meaningful statistical differences between the respondents' behavior and what they expect from their friends.

## Conclusion

How does the sharing behavior of interconnected users frame 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 randomized input frame elements will output local frames that 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. Only two conditions are required for this conclusion: partisans sharing at higher rates than non-partisans and sharing different frame elements than non-partisans. The proposed conjoint experiments describe these mechanisms replicating observational findings of network activation (Aruguete, 2019; Aruguete & Calvo, 2018; Aruguete et al., 2021) in three distinct surveys. The results of this study measure substantially similar partisan effects when analyzing sharing behavior (self) and the expectation of activation by peers (friends). Partisans share more content than non-partisans, their preferences are over-represented in the output of our conjoint experiments, and they expect their friends to be over-represented in social media data.

Our study provides several important contributions. We develop a novel theoretical explanation for the mixed finding about the formation of echo chambers in social media and the lack of empirical evidence of users' sorting on social networks and media diets (Bakshy et al., 2015; Barberá et al., 2015; Guess et al., 2021). Instead of focusing on sorting, our theory focuses on how social media bubbles emerge as users with different partisan loyalties to share content differently and expect different behavior from their personal networks. Our results explain heightened perceptions of polarization among social media users, even if user segregation across social media networks is not particularly high.

Second, our novel research design contributes methodologically to future research on social media effects. We show how to easily combine the methodological advantages of factorial experiments with an image-based implementation that provides high-ecological validity for social media studies. These designs are more flexible compared to commonly deployed social media framing experiments, since researchers can manipulate several theoretically relevant features (Hainmueller et al., 2014). This method also provides gains considering the ecological validity of survey experiments (Horiuchi et al., 2020; Thal, 2020; Vecchiato & Munger, 2021). Combined with recent research showing robustness of survey methods to measure sharing behavior on social media (Mosleh et al., 2020), our design alleviates concerns about external validity of survey experiments.

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 (Barbera, 2020). As noted recently by 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" (p. 132). To the best of our knowledge, we are the first to implement three similar conjoint experimental designs from 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, more diverse empirical evidence for a topic that has received considerable attention in recent years from political communication scholars.

Our results are not unique to social media networks. Decades of neurobiology research show that neural networks encode information by increasing or decreasing neuron's firing rates (Humphries, 2021). In social networks, an over emphasis on connectivity ("who is in our network") does not necessarily inform on activation ("what is shared").

The results of our experiment show that a set of randomized input frames may still yield partisan output frames (i.e. information bubbles). However, it does not provide insight into the subjective experience of a bubble, how *it feels* to be in a bubble. Additional research is also needed to determine peer effects –what do we do when our friends share partisan content. Two promising extensions of our factorial design should bring networks *back in*, modeling how peer effects alter our behavior as described by the integrated behavioral model (Montano & Kasprzyk, 2015) and how peers sharing decisions modify our own sharing behavior, as described by the "intuitive politician" model (Tetlock, 1991). While randomized experiments may tell us how to make bubbles, further research is needed to explore in greater detail the subjective experience of observing bubbles.

## References

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

Alipourfard, N., Nettasinghe, B., Abeliuk, A., Krishnamurthy, V., & 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. https://doi.org/10.1007/978-981-13-2262-4_55-1

Aruguete, N., & Calvo, E. (2018). Time to #Protest: Selective Exposure, Cascading Activation, and Framing in Social Media. *Journal of Communication, 68*(3), 480–502. https://doi.org/10.1093/joc/jqy007

Aruguete, N., Calvo, E., & Ventura, T. (2021). News by popular demand: Ideological congruence, issue salience, and media reputation in news sharing. *The International Journal of Press/Politics*, https://doi.org/10.1177/19401612211057068

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. F., Lee, J., Mann, M., Merhout, F., & 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. https://doi.org/10.1073/pnas.1804840115

Bakshy, E., Messing, S., & Adamic, L. (2015). Exposure to ideologically diverse news and opinion on facebook. *Science, 348*(6239), 1130–1132. https://doi.org/10.1126/science.aaa1160

Banks, A., Calvo, E., Karol, D., & Telhami, S. (2021). #polarizedfeeds: Three experiments on polarization, framing, and social media. *The International Journal of Press/Politics*, 26(3), 609–634. https://doi.org/10.1177/1940161220940964

Barberá, P. (2020). Social media, echo chambers, and political polarization. *Social Media and Democracy: The State of the Field, Prospects for Reform*, p. 34-55. https://doi.org/10.1017/9781108890960

Barberá, P., & Rivero, G. (2015). Understanding the political representativeness of twitter users. *Social Science Computer Review, 33*(6), 712–729. https://doi.org/10.1177/0894439314558836

Barberá, P., Wang, N., Bonneau, R., Jost, J. T., Nagler, J., Tucker, J., & González-Bailón, S. (2015). The Critical Periphery in the Growth of Social Protests. *PLoS one, 10*(11), e0143611. https://doi.org/10.1371/journal.pone.0143611

Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science, 26*(10), 1531–1542. https://doi.org/10.1177/0956797615594620

Benevenuto, F., Laender, A. H., & Alves, B. L. (2016). The h-index paradox: your coauthors have a higher h-index than you do. *Scientometrics, 106*(1), 469–474. https://doi.org/10.48550/arXiv.1510.04767

Conover, M. D., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., & 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., & Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of*

*the National Academy of Sciences, 113*(3), 554–559. https://doi.org/10.1073/pnas.1517441113

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

Entman, R. M., & 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. https://doi.org/10.1093/joc/jqx019

Eom, Y-H., & Jo, H-H. (2014). Generalized friendship paradox in complex networks: The case of scientific collaboration. *Scientific Reports, 4*(1), 1–6. https://doi.org/10.1038/srep04603

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

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

Fotouhi, B., Momeni, N., & Rabbat, M. G. (2014, November). Generalized friendship paradox: An analytical approach. In *International Conference on Social Informatics* (pp. 339-352). Springer, Cham. https://doi.org/10.48550/arXiv.1410.0586

Gentzkow, M., & Shapiro, J. M. (2011). Ideological segregation online and offline. *The Quarterly Journal of Economics, 126*(4), 1799–1839. https://doi.org/10.3386/w15916

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

Greene, K. F., & Sánchez-Talanquer, M. (2018) Latin America's Shifting Politics: Mexico's

Party System Under Stress. *Journal of Democracy, 29*(4), 31–42.

Guess, A. M., Barberá, P., Munzert, S., & Yang, J. (2021). The consequences of online partisan media. *Proceedings of the National Academy of Sciences, 118*(14), e2013464118. https://doi.org/10.1073/pnas.2013464118

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

Hainmueller, J., Hopkins, D. J., & Yamamoto, T. (2014). Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Political Analysis, 22*(1), 1–30. https://doi.org/10.1093/pan/mpt024

Horiuchi, Y., Markovich, Z., & Yamamoto, T. (2021). Does conjoint analysis mitigate social desirability bias? *Political Analysis,* 1-15. https://doi.org/10.1017/pan.2021.30

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

Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, Not Ideology: A Social Identity Perspective on Polarization. *Public Opinion Quarterly, 76*(3), 405–431. https://doi.org/10.1093/poq/nfs038

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., & Eom, Y. H. (2014). Generalized friendship paradox in networks with tunable degree-attribute correlation. *Physical Review E, 90*(2), 022809. https://doi.org/10.1103/PhysRevE.90.022809

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

Lee, E., Karimi, F., Wagner, C., Jo, H.-H., Strohmaier, M., & Galesic, M. (2019). Homophily and minority-group size explain perception biases in social networks. *Nature Human Behaviour, 3*(10), 1078–1087. https://doi.org/10.1038/s41562-019-0677-4

Leeper, T. J., Hobolt, S. B., & Tilley, J. (2020). Measuring subgroup preferences in conjoint experiments. *Political Analysis, 28*(2), 207–221. https://doi.org/10.1017/pan.2019.30

Lelkes, Y., Sood, G., & 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. https://doi.org/10.1111/ajps.12237

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

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

Mitchelstein, E., & 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. https://doi.org/10.1080/21670811.2021.1873813

Mosleh, M., Pennycook, G., & Rand, D. G. (2020). Self-reported willingness to share political news articles in online surveys correlates with actual sharing on Twitter. *PLoS one* 15(2), e0228882. https://doi.org/10.1371/journal.pone.0228882

Montano, D. E., & Kasprzyk, D. (2015). Theory of reasoned action, theory of planned behavior, and the integrated behavioral model. *Health Behavior: Theory, Research and Practice, 70*(4), 231.

Prud'homme, J. F. (2020). Parties and the party system in the 2018 mexican elections. *Foro Internacional, 60*(2), 397–450. https://doi.org/10.24201/fi.v60i2.2730

Osmundsen, M., Bor, A., Vahlstrup, P. B., Bechmann, A., & Petersen, M. B. (2021).

Partisan polarization is the primary psychological motivation behind political fake news sharing on twitter. *American Political Science Review, 115*(3), 1–17. https://doi.org/10.1017/S0003055421000290

Prior, M. (2019). *Hooked: How politics captures people's interest*. Cambridge University Press.

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

Saxena, B., & 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. https://doi.org/10.1007/s13278-019-0586-6 https://doi.org/10.1017/S002238161000006X

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

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. https://doi.org/10.1038/s41598-020-62085-w

Slothuus, R., & De Vreese, C. H. (2010). Political parties, motivated reasoning, and issue framing effects. *The Journal of Politics, 72*(3), 630–645. https://doi.org/10.1017/S002238161000006X

Stroud, N. J. (2010). Polarization and partisan selective exposure. *Journal of Communication, 60*(3), 556–576. https://doi.org/10.1111/j.1460-2466.2010.01497.x

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

*23*(1), 95–115. https://doi.org/10.1177/1940161217740697

Sunstein, C. R. (2018). *#Republic*. Princeton University Press.

Tetlock, P. E. (1991). An alternative metaphor in the study of judgment and choice: People as politicians. Theory & Psychology, 1(4), 451-475.

Thal, A. (2020). The desire for social status and economic conservatism among affluent americans. *American Political Science Review, 114*(2), 426–442. https://doi.org/10.1017/S0003055419000893

Tornberg, P. (2018). Echo chambers and viral misinformation: Modeling fake news as complex contagion. *PLoS one, 13*(9), e0203958. https://doi.org/10.1371/journal.pone.0203958

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

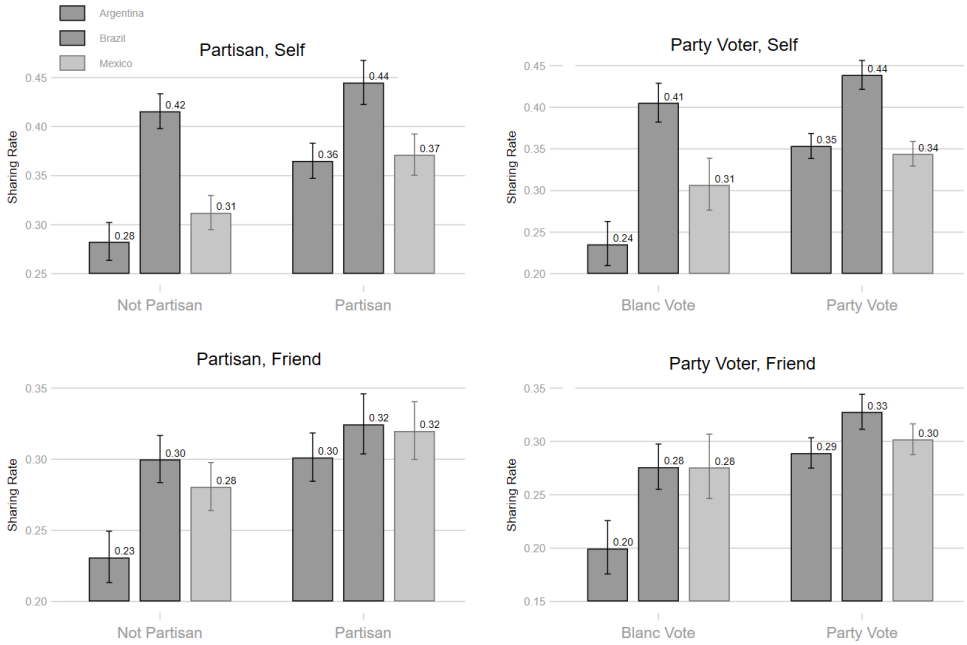
Weaver, D. (1991). Issue salience and public opinion: Are there consequences of agenda-setting?. International Journal of Public Opinion Research, 3(1), 53-68.

Wojcieszak, M. E., & 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. https://doi.org/10.1111/j.1460-2466.2008.01403.x

**Figure 1 Example of Image-Based Conjoint**


***Note*: Figure 1 describes different sample frame elements (left) and how they are on-the-fly combined to produce two possible tweets. Each respondent receives a different combination of frame elements.**

**Figure 2** Effect of Partisan and Voter on Sharing




Note: Figure 2(a) describes partisans on self. Figure 2(b) describes partisans on Friends. Figure 2(c) describes party voters themselves. Figure 2(d) describes party voters on friends.

**Figure 3:** Linear point estimates for Partisanship on sharing

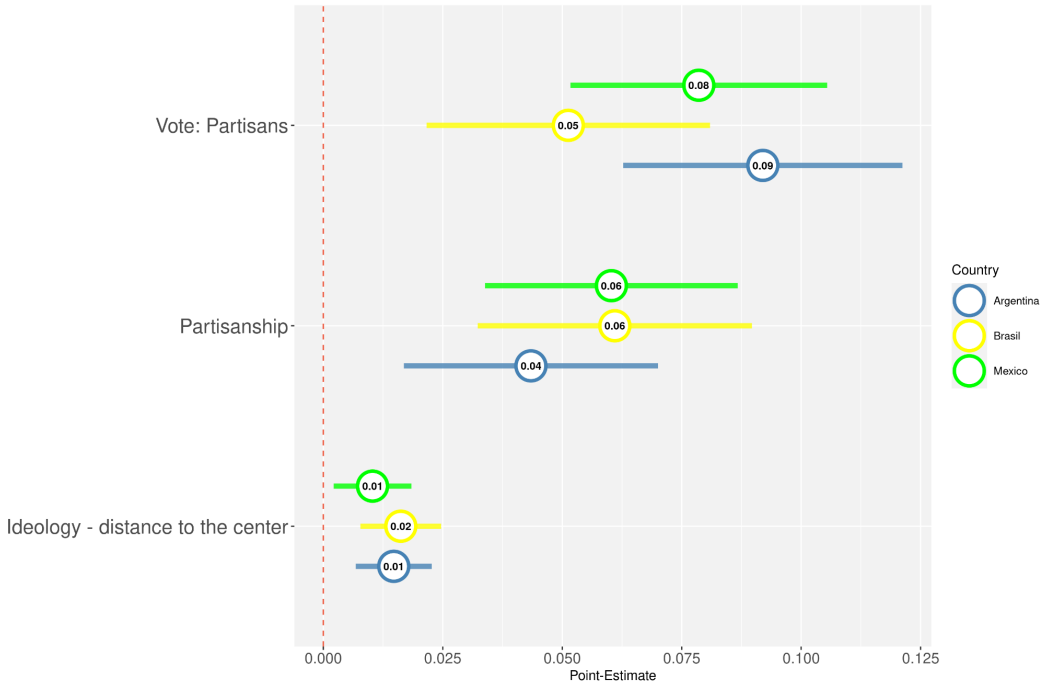
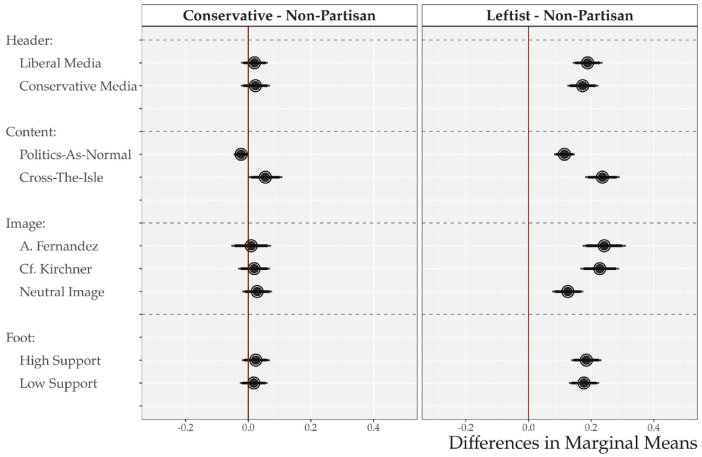
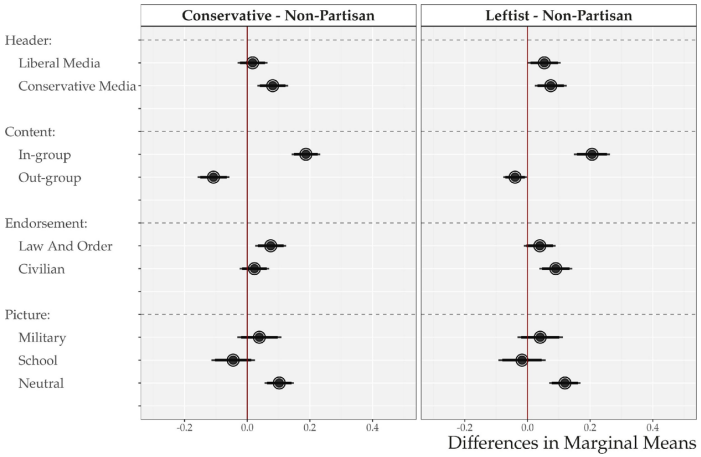


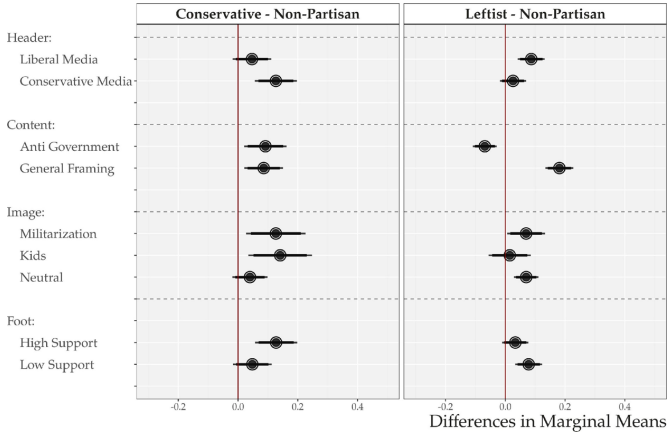
**Figure 4** Congruent Partisan Sharing: Network Activated Frames




a) Argentina

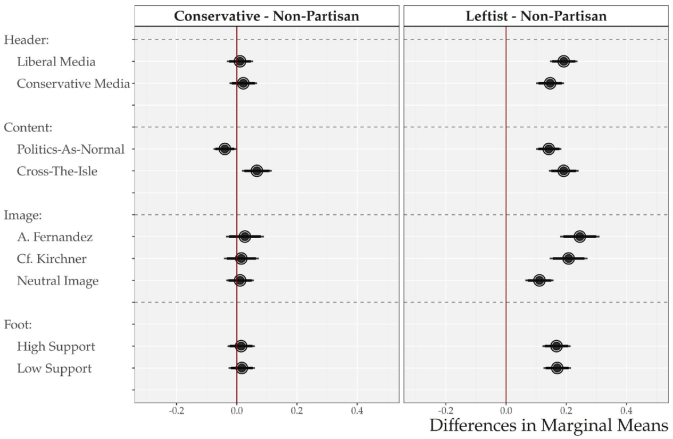



b) Brazil

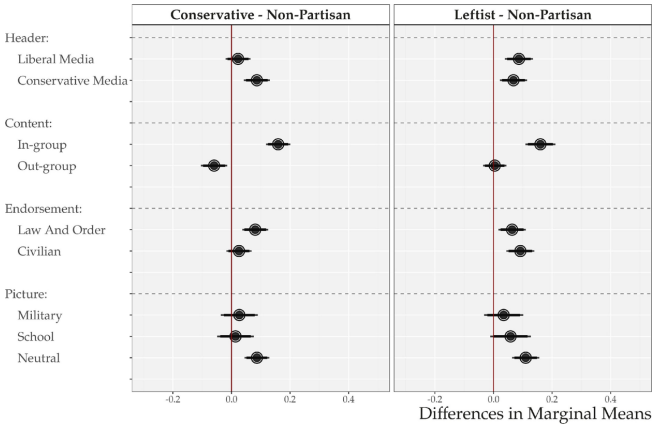



c) México

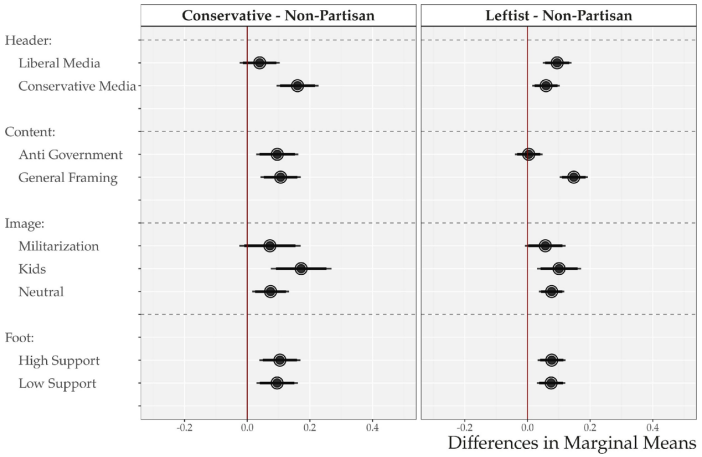
**Figure 5** Friends Congruent Sharing: Network Activated Frames




a) Argentina




b) Brazil




c) México

[1] While we test for the effect of partisanship on sharing, we do not sort out the potential reasons that explain why partisans share *more* and *different* content. An extensive discussion that relates partisanship and dispositional political interest is thoroughly discussed in Prior (2019).

[2] 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 non-administrative 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.

[3] 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