Friends Don't Let Friends Free Ride

Nicholas Eubank[*] and Dorothy Kronick[‡]

April 2, 2019

Please click here for most recent version.

Abstract

Theory predicts that social sanctioning can solve the collective action problem inherent to political participation, but only when people find out whether their peers participate. We test this prediction using data from the near-universe of cell-phone subscribers in Venezuela. Those whose behavior is more easily observed by peers are much more likely to protest and much more likely to sign a political petition than otherwise similar people in less-visible social network positions. Together with qualitative data, we interpret this finding as evidence that social network structure can facilitate (or frustrate) social sanctioning as a solution to the collective action problem.

[*]Center for the Study of Democratic Institutions, Vanderbilt University, nick@nickeubank.com

[†]Department of Political Science, University of Pennsylvania, kronick@sas.upenn.edu

[‡]We are grateful to Seth Bromberger and to contributors to open source projects like Julia, LightGraphs, iGraph, and Python. For comments, we thank Guy Grossman, Yue Hou, Jenn Larson, Michele Margolis, Zachary Steinert-Threlkeld, Dawn Teele, Stephanie Zonszein, Liz Zechmeister, and participants at the NYU Methods Workshop, and NEWEPS. **Disclosure:** This work was made possible by a data-sharing agreement between the authors and a telecommunications firm. In exchange for access to the data, the authors provided training in coding and data analysis. At no time did the authors provide the company with any financial compensation, nor did the company provide the authors with any financial compensation.

Fifty years ago, Mancur Olson (1965) posed a question that has endured. Why do people protest, or lobby, or strike, or otherwise take part in collective action, given that doing so requires effort and risk, while everyone—participant and free-rider alike—reaps the spoils of victory?

One influential explanation is that free riding is rarely free: non-participants often suffer from peer sanctioning (Ostrom, 1990), while participants enjoy social approbation. Empirical work in political science, psychology, sociology, and economics confirms that peers affect political participation, but we know little about why certain communities are better able to wield social influence, or why certain individuals are more susceptible to it.[1]

In this paper, we study the role of social network structure. Any system of community enforcement—that is, any agreement to shun free riders or praise participants—requires that people find out about pro- or anti-social behavior. If the authors could sit out a protest without anyone hearing about it, we would not fear social sanction; if everyone would hear about it, we might be more likely to attend, wary of the disapproval of friends, family, or coworkers.

Information diffusion is thus a prerequisite for using social influence to solve the collective action problem. Specifically, theorists studying other types of pro-social behavior (such as law abidance) have argued that people with higher *network exposure*—that is, people about whom information spreads quickly through the network—will be more susceptible to peer sanctioning and thus less likely to free ride (e.g. Wolitzky, 2012; Larson, 2017; Jackson et al., 2017).

We evaluate whether network exposure drives political participation. To measure network exposure, we use meta-data from the near-universe of cell-phone subscribers in Venezuela. Because summary statistics such as number of immediate connections do not capture the speed of information diffusion (Newman et al., 2006), we use simulations to estimate how quickly information about an individual is likely to spread through her social network – the quantity that matters in theoretical models.[2] To measure political participation, we observe (a) whether each person signs a petition demanding a referendum on recalling Venezuelan President Nicolás Maduro and (b) whether each person participates in a protest aimed at convincing the government to honor the petition (measured using the location of her cell phone during the protest).

Both activities are costly, not only in time and effort, but also in the risk

[1] This is a vast literature that we do not attempt to summarize. In political science, see, e.g., Gerber et al. (2008, 2016); Panagopoulos (2010); Anoll (2018).

[2] Our measure can be applied to any network data; we contributed the code to the open-source `LightGraphs.jl` library, and it is available here. Our own full implementation will be posted upon publication of this paper, and is available by request in the meantime.

of political reprisal. Participants had died at previous protests in Venezuela, and signers of an earlier political political petition were fired from government posts and barred from certain welfare programs (Hsieh et al., 2011). And because the success of either effort would entail benefits for participants and non-participants alike, both activities faced the collective action problem.

To study the relationship between network exposure and political participation, we construct samples of participants and non-participants who have similar observable characteristics. For each person in a random sample of protesters, and for each person in a random sample of petition signers, we find a non-protester or a non-signer who votes at the same polling place, has the same party registration, is of the same gender and of a similar age, and has a similar level of geographic mobility.

Within this matched sample, we estimate (politically) large and (statistically) significant effects of network exposure on political participation: for example, a (within-pair) one-standard deviation increase in exposure predicts a 2.8-percentage-point increase in the probability of signing the petition. We then consider the political consequences of shifts in the network exposure of the mobilizable population (that is, the population with characteristics similar to those of actual protesters), finding that even subtle changes in the social network can generate politically meaningful changes in participation.

Given the particulars of the Venezuelan context (Section 1), and based on responses to original survey questions, we interpret these findings as evidence that social network exposure facilitates social sanctioning, which helps solve the collective action problem.

Our ability to observe participation in *both* the protest and the petition signature drive conveys inferential advantages beyond external validity. A protest is an inherently social activity, usually attended in the company of friends; thus, protesters might have higher network exposure than non-protesters for reasons unrelated to social sanctioning: protesters might simply be more sociable. In other words, they might also be more likely to attend (say) dinner parties. Indeed, the relationship between network exposure and participation is stronger for protest than for petition signing, suggesting that the protest results are driven in part by sociability. This comparison has implications for other results in the literature (c.f. Gonzalez, 2017; Larson et al., 2019).

One concern is that, despite our attempt to match on observables, these results are driven by unobserved characteristics that are correlated with both network exposure and political participation. Placebo tests suggest that the results are not driven by unobserved characteristics that are unbalanced across groups.

This supports our interpretation of the data: that network exposure facilitates social sanctioning and thereby political participation.

This analysis contributes to the literature on social networks and political participation. First, we introduce a direct empirical measure of network exposure. Second, ours is the first large-scale quantitative study to relate offline social networks to offline political activity using behavioral (rather than self-reported) measures; Larson et al. (2019) relates Twitter networks to street protest activity, but, as the authors acknowledge, online social networks differ from offline social networks. In particular, online social networks—and self-reported information on social ties—may censor certain types of connections, such as inter-class or inter-ethnic relationships, or relationships with people who are not on social media.

Third, because cell-phone penetration in Venezuela is above 90%, we study the relationship between social networks and protest in a sample that is more representative than most convenience samples (such as geo-tagged Twitter data). Fourth, because we are able to link the cell-phone meta-data to the electoral registry (at the individual level) and to census data (at the tract level), we compare protesters with non-protesters who are more observationally similar than the comparison sets used in past work.

And finally, because of the particulars of the specific incidents of political participation being studied, we are able to rule out explanations other than social sanctioning for why network exposure matters with greater confidence than would be possible in other contexts.

Our results also speak to the literature on political participation in Latin America. In descriptive analysis using a representative sample of six large Venezuelan states,[3] we find that network exposure is a strong predictor of participation—even compared to the factors most discussed in the literature, like socio-economic status (e.g. Booth and Seligson, 2006). Our findings thus complement the qualitative literature that emphasizes the role of informal social networks in facilitating participation (Valesco, 2015).

[3]We restrict this part of the analysis to these six states because they are the only ones for which we have census-tract maps. The census-tract maps are used to link neighborhood characteristics (like education) to electoral precincts (and thus to the voters in our data).

1 Context: Collective Action in Venezuela

In the two years after Nicolás Maduro took office as president of Venezuela in April of 2013, the country suffered the worst recession in its recorded history (Kronick, 2015).[4] Maduro's approval ratings dipped below 25% (Datanalisis, 2016), and, in April of 2016, a coalition of opposition parties decided to collect signatures in support of holding a recall referendum: a national yes-or-no vote on whether to recall Maduro from office. Under the Venezuelan constitution, written during the presidency of Maduro's predecessor and mentor Hugo Chávez, the signatures of 1% of registered voters would suffice to begin the recall referendum process.[5]

The coalition of opposition parties set up petition-signing stations throughout the country, where registered voters could add their signatures and thumbprints to the petition (see Noticia al Día, 2016, for photos and a description). The stations were open for two days (April 27 and April 28, 2016); on May 2, opposition leaders announced that they had collected and submitted to the National Electoral Council 1.8 million signatures—nearly ten times the required 1% (Toro, 2016).

The Electoral Council claimed that the signatures were not valid. In the view of economist Francisco Rodríguez, "There was just nothing even resembling a normally coherent argument about why it was that the referendum was stopped. The government alleges that there was fraud in the collection of signatures, but ... there were enough signatures, even excluding the presumed fraudulent signatures, to get the process to go forward. But nevertheless the government stopped it" (2017).

Opposition leaders responded by calling for a large demonstration, the *Toma de Caracas* ("taking of Caracas"), and they scheduled it for September 1. The time and location of the march were publicized on Twitter and in major media outlets, and the march itself was covered in the national and international press. The protest organizers claimed that more than one million people participated in the march; an independent measure based on photographs estimated participation at 700,000 (Rodriguez, 2016).

We use (a) signing of the recall referendum petition and (b) presence at the

[4]Since then, the recession has become the worst in recorded Latin American history.

[5]Technically, it is an electoral regulation (not the Constitution) that requires the signatures of 1% of registered voters to begin the process. The regulation specifies that, with signatures of 1% of the electorate in hand, the Electoral Council would supervise the collection of signatures of the 20% of the electorate required by the Constitution in order to hold the recall vote.

Toma de Caracas as our measures of participation in collective political activities. (See Section 3 for measurement details.) Both were costly activities subject to the collective action problem. In addition to the time and transportation costs required to sign the petition, signatories had reason to expect retribution: signatories of an earlier, similar petition had been subsequently discriminated against in the labor market (Hsieh et al., 2011) and in applications for government benefits (Albertus, 2015). Likewise, in addition to the time and transportation costs required to attend the protest, participants had reason to fear violence: participants in earlier protests had suffered injuries, arrest, torture, and even fatalities (Toro, 2014).

2 Theory: Social Networks, Information Diffusion, and Political Participation

The free-rider problem plagues protests, petitions, and and other collective political activities: why spend time (and risk reprisals) participating when you could instead stay home and still benefit from any political gains?

One answer is social sanctioning. A general community agreement to shun, chastise, or otherwise punish shirkers provides powerful incentives against free-riding (Ostrom, 1990; Kandori, 1992; Fearon and Laitin, 1996). Likewise, social recognition of people who do "do their civic duty" increases the returns to participation (Sinclair, 2012)

Jackson et al. (2017, p. 75) explain why information diffusion underpins any system of community enforcement:

> If friends and neighbors are made quickly aware of the individual's behavior, then they can react quickly. This is important for providing the right incentives, as the threat of a punishment in the near future will have the greatest scope for disciplining behavior. If instead, the setting is such that it takes a long time for one's friends to learn of misbehavior then it becomes difficult to provide incentives for individuals to behave according to some desired social norm.

Theorists have formalized this intuition. Larson (2017), for example, presents a model in which the threat of collective sanction of defectors sustains cooperation in one-shot bilateral prisoner's dilemma games (played between random pairs of individuals) (à la Kandori, 1992). In the equilibrium of interest, cooperation can be sustained only among individuals whose network positions ensure that, if they defect, word will spread quickly enough to ensure a credible threat of sanction by future partners. This excludes people in isolated network positions, for whom the threat of collective sanction is weak.

A similar dynamic emerges in a repeated public goods game in Wolitzky (2012). In the maximally cooperative equilibrium, a person's contribution to the public good is monotonically increasing in the speed with which other contributors would hear about it if she under-contributed. The more quickly other contributors would hear, the greater the threat of sanction, and thus the greater the contribution in equilibrium.

These theories yield similar comparative statics: people in more visible network positions—that is, network positions that facilitate the spread of information about their actions—are more likely to participate, because they face stronger potential social sanction for staying home.

Of course, information diffusion through social networks might enable collective action through means other than enabling social sanctioning. Most obviously, it might increase public awareness of a protest or other political activity, or it might facilitate private coordination among protest organizers seeking to avoid government disruption if the details of protests become known (Little, 2015; Christensen and Garfias, 2018; Enikopolov et al., 2017). In our context, however, there is little evidence that organizers were forced to rely on private social networks to coordinate and promote events to avoid government repression; to the contrary, the petition signing opportunities and the *Toma de Caracas* protest we study were announced on social media and covered extensively in all major press outlets. Consequently, we do not believe private communications along social networks were necessary for coordination in the manner they likely were in, for example, Arab Spring protests. Indeed, it is the very public nature of these activities motivates our focus on social sanctioning (or social approbation) as the mechanism of interest.

Similarly, the presence of (at the time) reliable public opinion polls motivates our choice to focus on free riding rather than preference falsification as the impediment to collective action. In many authoritarian contexts, preference falsification generates the distorted impression of widespread support for the regime, which then hinders protest participation (Kuran, 1991; Steinert-Threlkeld, 2018; Little, 2015).[6] In Venezuela, in contrast, media outlets regu-

[6] If we were to focus on preference falsification, we would have used a different network model to motivate our analysis, such as complex contagion (Centola, 2018; Centola and Macy, 2007) or social context (Siegel, 2009). In these models, participation snowballs: individuals choose to protest as they observe peers protesting, because these observations lead them to update about the level of support for the government in the population. Our

larly reported the president's unpopularity: in the summer of 2016, just before the *Toma de Caracas* protest, major pollsters placed Maduro's approval ratings between 22% and 31%. There was no public perception of widespread support for the government.

To be clear, we do not study the direct influence of one person's observed participation her friends' participation decisions (as in, e.g., Rolfe, 2012; Siegel, 2009; Banerjee et al., 2014; Eubank et al., 2018; Ferrali et al., 2018). Rather, our interest lies in how social network structure "determines the forms of cooperative behavior that can be maintained" (Jackson et al., 2017, p. 5) by shaping the *anticipated* social cost of non-participation—which exists even (or especially) when sanctioning never actually occurs in equilibrium. Our empirical analysis is thus more in line with Jackson et al. (2012a, p. 1882–1883) or Larson et al. (2019), both of which compare the network positions of cooperative and non-cooperative agents.

2.1 Networks, Social Capital, and Civic Duty

While the focus of this paper is on network theory, it is worth noting an important connection to both the social capital and civic duty literatures.

As noted by Jackson et al. (2012b), the literature on social capital can be broadly divided into two models of social capital. The first suggests that communities with high social capital are communities in which individuals care more about the well-being of others (in the language of economics, the well-being of others enters directly into their utility function). Consequently, we observe pro-social behavior in these communities because individuals enjoy doing good for the community (they internalizing the externalities of their actions). Social capital, in other words, is a social-psychological phenomenon.

But a second set of models suggests that communities with high social capital are communities that are better able to police poor behavior, leading people to act in a pro-social manner not because they care more about their peers than those in low social capital communities, but because they fear sanction if they do not.[7] In this view, the pro-social behavior associated with social capital is best viewed as *epiphenomenal*.

A similar distinction can also be found in work on norms and political partici-

focus instead on free riding motivates our choice of Larson (2017) and Wolitzky (2012) as theoretical foundations for the analysis.

[7]This distinction has not alway been recognized in the social capital literature, and work on social capital often includes elements of both mechanisms (e.g. Putnam et al. (1993)).

pation (?Anoll, 2018). Within this literature, norms are sometimes suggested to drive political participation by giving participants a warm-fuzzy feeling when they vote (intrinsic motivation), while other times (or in addition) they motivate participation by ensuring participations will garner social approbation for their behavior (extrinsic motivation).

In both cases, the latter of these mechanisms (epiphenomenal social capital and extrinsic social motivation) is closely related to the theory being tested here. As such, while our paper is largely framed as a test of a test of a specific theory about the role of networks in moderating capacity to apply peer pressure, our results also have important implications for the literature on both social capital and civic duty.

3 Measuring Network Exposure and Political Participation

The theoretical literature summarized in Section 2 predicts that, all else equal, people with more network exposure are more susceptible to peer pressure and therefore more likely to participate in collective action. To evaluate this prediction, we develop an empirical measure of network exposure—or, in other words, a measure of how quickly information about a person diffuses through her social network.

3.1 Mapping the Social Network

Before measuring network exposure, we must map the social network through which information diffuses. We accomplish this using cell-phone meta-data from a major Venezuelan telecommunications company ("Partner Telecom").[8]

The data include records of all SMS (text) and voice transactions sent or received by subscribers of the Partner Telecom from June 2016 to February 2017, totaling approximately 30 billion transactions. Because we observe transactions sent *or* received by Partner Telecom subscribers, our data include information about other providers' clients—not just Partner Telecom clients. Each record includes type (voice or text), identifiers for both caller and receiver, date and time, duration, and GPS coordinates of the antenna tower through which all calls were placed (we do not observe antenna towers for text messages).[9]

[8]The telecommunications company requested anonymity.

[9]All identified data is stored on a computer with no physical means of connecting to the

Within this data, we classify two users as *connected* if they (a) call each other at least twice *or* text each other at least twelve times in (b) at least two of the eight months in our data.[10] We also present results using call thresholds of four voice calls or twenty-four texts; results are similar for six voice calls or thirty-six texts. In our baseline network specification, the median person has nine immediate connections, and the average person has nineteen immediate connections.

In principle, we could use call frequency to measure the strength of ties (rather than classifying each pair of people as *connected* or *not connected*). In practice, we would not trust this measure. First, it is unclear whether call frequency is strongly correlated with the social importance of a relationship.[11] Second, phone calls are only one of several modes of communication. Even if overall communication frequency were indicative of social importance, weighting by *call* frequency may under-estimate the importance of relationships among individuals who live nearby and primarily communicate face-to-face—or those who communicate via email or WhatsApp.[12] For more on our motivation for using a dichotomous rather than a continuous measure of social ties, see Appendix A.

Our mapping of the social network is imperfect. First, while we do observe communications between Partner Telecom subscribers and those using a different service provider, we do not observe connections among pairs of cell-phone users when neither of them is a Partner Telecom subscriber. (To preserve anonymity, we cannot report their market share.) Second, our data may generate false positives: a person might make many calls to a local business, for example, without being socially connected to that business (although by requiring multiple communications in at least two of our eight months of data, we do filter out one-off business communications, and by dropping the 1% of users with the largest number of connections, we also filter out very large companies and businesses). Likewise, we cannot avoid false negatives even among Partner Telecom subscribers; family members or others who see each other daily may rarely call or text, especially with the growth of WhatsApp.

In our view, the advantages of our data outweigh these drawbacks. Many

internet (an air-gapped computer), in an access-controlled room, on encrypted hard drives; the data protection protocols were approved by two Institutional Review Boards.

[10]The value of twelve texts reflects the fact that messages are about six times more common than voice calls.

[11]As one author's mother will attest, calls to socially important family members are not always as frequent as less-important communications with co-workers.

[12]WhatApp looks like data usage to telecommunications firms, so our data excludes information on WhatsApp communication.

empirical studies map social networks by asking individuals for the names of their closest friends in a geographically bounded area (e.g. Larson and Lewis, 2017; Banerjee et al., 2014; Rojo and Wibbels, 2014; Fafchamps and Vicente, 2013; Dionne, 2015). Relative to these self-reported friendship networks, our data provide more breadth: in our primary specification, we observe 27 million unique phone lines, or the near universe of lines in the country.[13] Our data also provide more depth, capturing even weak social ties that may be censored or forgotten when networks are measured by asking people to list a finite number (often five) of their closest friends—weak ties that sociologists have found to be important (Granovetter, 1973). Third, unlike data from self-reports or social media, our data measure actual, non-public, offline communication patterns. They therefore reveal many connections—like inter-class or inter-party ties—that might be under-reported in surveys or absent online (perhaps due to fears of prejudice).

3.2 Measuring Network Exposure

With this mapping of the social network in place, we measure a person's *network exposure* by simulating an information diffusion process. Calculating network exposure requires simulation because to analytically calculate the speed information would diffuse through a network, "one needs to account for all the possible paths that information might take, and some end up overlapping, producing correlation in the chance that information makes it from one node to another" (Jackson, 2018, p. 8).

For each individual i, we simulate a diffusion process starting at i and record the number of i's peers who are reached (that is, informed) by the diffusion process at each time t.[14] The simulation proceeds as follows:

1. At time $t = 0$, the subject vertex v_i is endowed with a unique piece of knowledge. In other words, she is *informed*. All other vertices are ignorant. Letting $\mathbb{I}_{i,t}$ denote the set of vertices who have been informed by time t in a diffusion process beginning at v_i, this implies that $\mathbb{I}_{i,0} = \{v_i\}$.

[13] The Venezuelan population was approximately 31 million in 2016; the World Bank estimates 93 cell lines per 100 people.

[14] In some theoretical models, information about a person *not* participating would have to spread without her help—i.e., in the network $g \backslash i$—while she might help spread information if she *did* participate. We abstract away from this distinction, allowing the information diffusion process to begin at person i whether or not she participated. In our view, it is plausible that a person's peers could learn about her non-participation directly from her (for example, by asking, or indirectly through general awareness of a friend's schedule).

2. At time $t = 1$, information spreads to each neighbor of v_i with probability $p/log(degree(v_i) + 1) \in (0, 1)$. Our decision to normalize the probability of diffusion by the log of i's number of neighbors—that is, by $log(degree(v_i)+1)$ —is motivated by an empirical regularity: people with more friends do make more calls overall than users with fewer friends, but this increase in sub-linear in the number of friends (Miritello et al., 2013), likely due to time constraints. The idea that overall frequency of interactions is limited by time constraints is also consistent with the findings of (Larson and Lewis, 2017).[15]

3. From each vertex v_j informed at $t = 1$ (that is, from each $v_j \in \mathbb{I}_{i,1}$), the information spreads to v_j's neighbors with probability $p/log(degree(v_j)+1)$ in $t = 2$, creating a new set of informed vertices $\mathbb{I}_{i,2}$.

4. For each subsequent time period t, information spreads from each informed vertex $v_j \in \mathbb{I}_{t-1}$ to each of v_j's neighbors with probability $p/log(degree(v_j) + 1)$, creating a new set of informed vertices $\mathbb{I}_{i,t}$.

We run this simulation 1,000 times for each person (vertex) in our samples of interest (described below). Averaging over the 1,000 simulations, we create measures of two types of network exposure:

$$\begin{aligned} \text{General network exposure: } & N^g_{i,t} \equiv |\mathbb{I}_{i,t}| \\ \text{Exposure to participants: } & N^p_{i,t} \equiv |\mathbb{I}_{i,t}| \cap \mathbb{P}, \\ & \quad \text{where } \mathbb{P} \text{ is the set of participants} \end{aligned}$$

In other words, the *general network exposure* of person i at time t ($N^g_{i,t}$) is the number of people informed about her behavior up to and including time t; her *exposure to participants* (an alternate measure of exposure discussed in Section 4.2.5) is the number of other *participants* (people who signed the petition or attended the protest) informed about i's behavior up to and including time t.

It is worth emphasizing that *general network exposure* is a pure measure of network structure that does not take into account any information about whether the people reached by the diffusion process are political participants themselves. As a result, our analysis is not impacted by one of the major confounds of studies of political participation and network exposure: the possibility that individuals surrounded by political participants are more likely to participate themselves not because of the influence of their peers, but because they have a latent preference for political participation that drove them to associate with

[15]Results are similar when we normalize by $degree(v)$ rather than $log(degree(v) + 1)$.

other participants (political homophily).

As we noted in the introduction, no simple statistic—such as average degree or average shortest path length—captures the speed of information diffusion through a network (Newman et al., 2006); moreover, the relationship between information diffusion and proxies for network structure in the literature—such as membership in civic organizations or self-reported number of friends—is generally unknowable. Appendix Table 8 reports the correlation between network exposure and *eigenvector centrality*, a common measure that calculates the centrality of a node as the (scaled) sum of the centrality of its neighbors (Jackson, 2008, p. 41); the two measures turn out to be highly correlated (≈ 0.7), but not identical. If replicated in other social network data sets, this has important implications for the interpretation of analyses of eigenvector centrality (as it implies centrality is also closely related to network exposure), as well as future studies of information diffusion, as eigenvector centrality is generally easier to calculate than the simulation-based measure used here. (Appendix Table 8 also reports the correlations between network centrality at one time step and network centrality at other time steps; we return to this below.)

To the best of our knowledge, therefore, our simulation-based estimate of network exposure is the only direct empirical measure of the quantity that matters in much of the theoretical literature (see Section 2). For interested scholars, the code for implementing our simulation-based is freely available.[16]

3.3 Measuring Political Participation

Measuring petition signing is straightforward. The (identified) list of 1.7 million signatories was circulated; we merge the list (on national I.D. number) with our measures of network exposure.

To measure protest activity, we classify as a *protester* any cell-phone subscriber with at least one call routed through a cell tower along the protest route during the time of the protest—except those who live or work nearby. Specifically, we exclude from the sample those who live or work in parishes adjacent to the protest route (parish is a sub-county administrative unit). We do this

[16] Our measure can be applied to any network data; the base code has been contributed to the open-source `LightGraphs.jl` library, and is freely available here. Our most recent implementation (with log-normalization of diffusion probabilities) will be integrated into the `LightGraphs.jl` library in the near future, and is available upon request if readers would like a copy prior to integration.

in order to reduce the incidence of false positives: the imprecision of cell-tower-based geolocation prevents us from distinguishing *presence at the protest* from *presence at a nearby home or business*. The data identify approximately 160,000 Partner Telecom subscribers who place a call in the vicinity of the protest during protest hours on September 1; when we drop those who live or work in parishes adjacent to the protest route, we are left with a sample of approximately 44,000 protesters.[17]

The use of two distinct *behaviorial* measures of political participation constitutes one of the major contributions of this analysis. First, because the measures are behavioral and not self-reported, we are able to avoid issues of desirability bias that are especially likely to appear in self-reports if political participation is indeed influenced by social pressure. And second, as discussed in more detail below, the ability to measure to categorically different forms of political participation allow us to ensure our results are not driven by other features of a specific form of political participation (like the possibility that people who attend protests are just more sociable an like hanging out in big groups, a confound not present in petition signing).

3.4 Matching participants to similar non-participants

The theoretical results summarized in Section 2 imply that, in equilibrium, people with greater network exposure should be more likely to participate than otherwise identical people who are less exposed.

The "otherwise identical people" part of these models poses a challenge for our empirical tests. In theoretical work, network position and other characteristics—like socioeconomic status—are assumed to be independent. In reality, they are correlated. We address this challenge by matching on observables.

Specifically, for each protester in a random sample of $5,000$ protesters, we identify a non-protester who is (a) registered to vote at the same polling place, (b) of the same political party, (c) likely to spend a similar amount of time in Caracas in a given month (excluding the protest day), (d) of the same gender, (e) on the same type of cell-phone plan (pre- or post-paid), (f) of a similar age, and (g) of a similar level of mobility (as measured by the spatial variance, across days, of her cell phone). Of these variables, matching on polling place appears to account for the largest share of variation due to the relatively

[17]That we observe only 160,000 of the approximately 700,000 (estimated) protesters at the *Toma de Caracas* makes sense given (a) the fact that we only have cell-tower-routing data for subscribers of our Partner Telecom, and (b) the fact that we only have cell-tower-routing data for phone calls, not for text messages.

small size of polling places and geographic socio-economic segregation. Results are similar when also matching on average user call frequency. For details of our matching strategy, see Appendix B. This strategy — made possible by exceptional data availability — constitutes a substantial improvement in efforts to create comparable comparison groups in large scale studies of social networks and political protest, like Larson et al. (2019).[18]

4 Results: How Network Exposure Drives Political Participation

4.1 Descriptive analysis

Descriptively, how does network exposure relate to socio-economic characteristics and to political participation? In this section, we study these correlations using a representative sample of individuals from six large Venezuelan states. For these six states, we obtained census-tract maps that allow us to link neighborhood characteristics (like education) to electoral precincts (and thus to the voters in our data).[19] We re-weight the sample so that it resembles the overall Venezuelan population.

Figures 1b and 1c reveal that, contrary to our expectations, network exposure is only weakly correlated with neighborhood socioeconomic characteristics. In particular, it is only weakly correlated with the proportion of households in the neighborhood that have a cement floor. This captures whether a neighborhood has formal or informal housing, or whether the neighborhood is a *barrio* (in Venezuelan terms); the distribution of the proportion is bimodal, with modes close to zero and close to 0.8.[20] It is also only weakly correlated with the proportion of adults (over 25) in the neighborhood who have a college degree. (Note that the scales of the y-axes in these figures span the first 90 percent of the distribution of network exposure).

In contrast, Figures 1e and 1f reveal that network exposure is strongly correlated with both protest participation and petition signing. To investigate

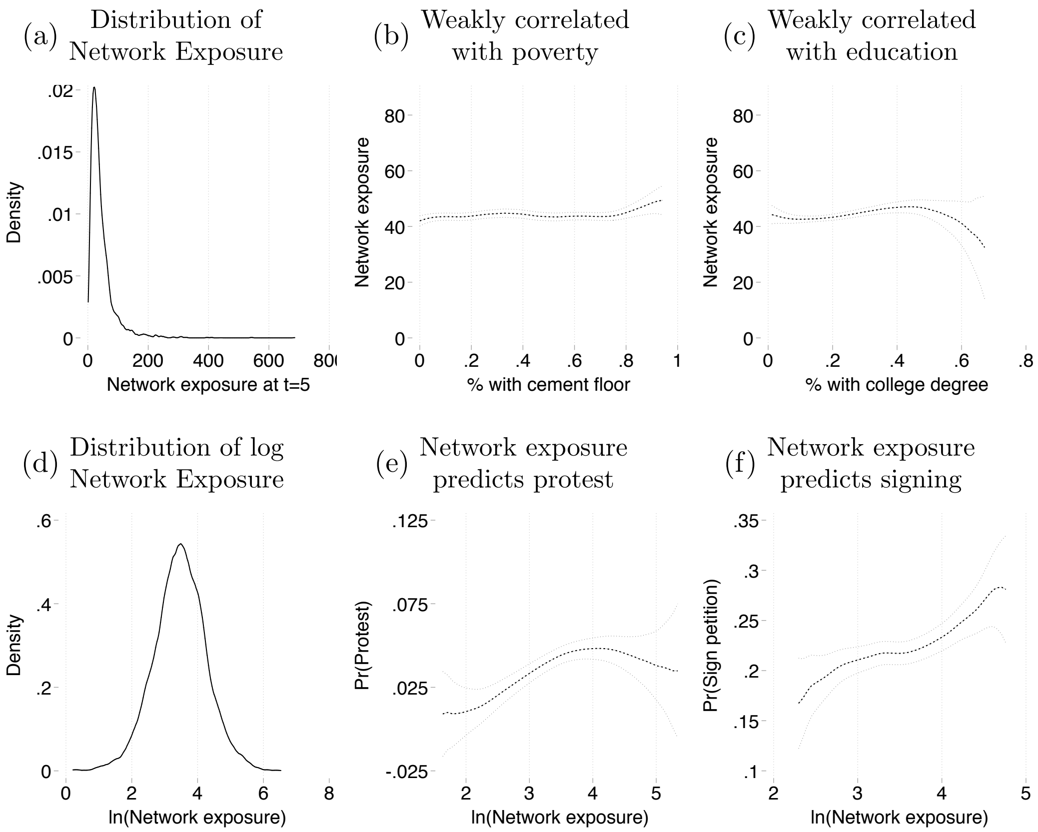
[18]Larson et al. (2019) compared Twitter users who issued a geo-coded tweet from the location of the Charlie Hebdo protests using a related hashtag to Twitter users who also issued a geo-coded tweet using a related hashtag from within Paris but from a location more than 5km from the protest site. This ensured the creation of subjects matched on interest and ability to attend, but not other demographic features.

[19]The six states are Aragua, Carabobo, the Federal District, Lara, Miranda, and Vargas.

[20]By "neighborhood," we mean the census tract in which a person's polling place is located.

Figure 1: Network exposure predicts participation in the population

These figures describe the relationships between network exposure and (a) neighborhood (census-tract) poverty, (b) neighborhood (census-tract) education, (c) protest participation, and (d) petition-signing, all using a representative sample of the population of six large Venezuelan states.



how this correlation changes when we account for other factors that affect participation, we estimate:

$$\text{Participate}_i = \gamma_0 + \gamma_1 N^a_{i,5} + \boldsymbol{\gamma_2 X_i} + \eta_i \tag{1}$$

where $\boldsymbol{X_i}$ includes (a) municipality fixed effects, (b) whether person i is registered with the United Socialist Party of Venezuela (PSUV, the party of the government), (c) the proportion of adults (over 25) in the neighborhood who have a college degree; and (d) the proportion of household in the neighborhood with a cement floor.

Table 1 compares the predictive power of network exposure to the predictive power of party registration and neighborhood socio-economic factors. All coefficients are scaled so as to capture the predicted change in participation associated with moving from the 5th to the 95th percentile of the independent

Table 1: Network exposure predicts participation in the general population

Predicted change in participation associated with moving from the 5th to the 95th percentile of each independent variable (or from zero to one, for indicators).

	Protest			Petition		
	(1)	(2)	(3)	(4)	(5)	(6)
Network exposure ($N^g_{i,5}$)	0.027	0.040	0.040	0.046	0.042	0.041
	(0.009)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)
Registered in PSUV (gov't party)		-0.020	-0.018		-0.113	-0.101
		(0.01)	(0.01)		(0.02)	(0.01)
Signed petition		0.017	0.016			
		(0.006)	(0.006)			
% Neighborhood w/ cement floor			-0.009			-0.042
			(0.009)			(0.05)
% Neighborhood w/ college degree			0.007			0.081
			(0.008)			(0.05)
Observations	8033	8033	8033	8033	8033	8033
Municipality FEs		✓	✓		✓	✓

variable. Network exposure predicts larger changes in protest participation than does party registration or neighborhood socioeconomic characteristics. And moving from the 5th to the 95th percentile of network exposure predicts a 4.5-percentage-point increase in the probability of signing the petition—about half of the difference associated with moving from the 5th to the 95th percentile neighborhood education. These comparisons indicate that, in descriptive terms, network exposure is a strong predictor of participation in the general population, even compared to the characteristics often emphasized in the literature on participation in Latin America.

4.2 Network Exposure and Political Participation: Matched Sample Analysis

This section compares the network exposure of participants and observationally similar non-participants. The goal of this comparison is to gauge how network exposure affects participation: if a person's social network suddenly shifted such that more people would hear about her behavior, would she be any more likely to participate? How much more likely? If an important link in one of our social networks unexpectedly disappeared, such that we expected fewer of our colleagues to know what we did last summer, would we have been any less likely to attend protests? How much less likely?

4.2.1 Participation increases with network exposure

The analysis in this section describes variation *within* matched pairs, asking whether protesters and/or petition signers have higher or lower network exposure than their observationally similar counterparts. We use $\overline{N_{pair}}$ to denote the within-pair mean of network exposure, such that $N_i - \overline{N_{pair}}$ denotes person i's deviation from the within-pair mean of network exposure. Except where otherwise noted, we use network exposure measured at time step $t = 5$; results for other time steps are similar.[21]

Figure 2 plots the distribution of deviations from the within-pair means of network exposure to participants, i.e., the distribution of $N_i^g - \overline{N_{pair}^g}$. A person with a value of 1.5, for example, is exposed to three more people than her matched counterpart.

To better gauge the magnitude of the differences in Figure 2, Figure 3a plots the probability of protest participation given deviations from the within-pair mean of exposure to participants, $N_{i,5}^g - \overline{N_{pair,5}^g}$. Within matched pairs, the probability of protest participation increases sharply with network exposure, consistent with the notion that network exposure facilitates social sanctioning and thereby encourages protest participation.

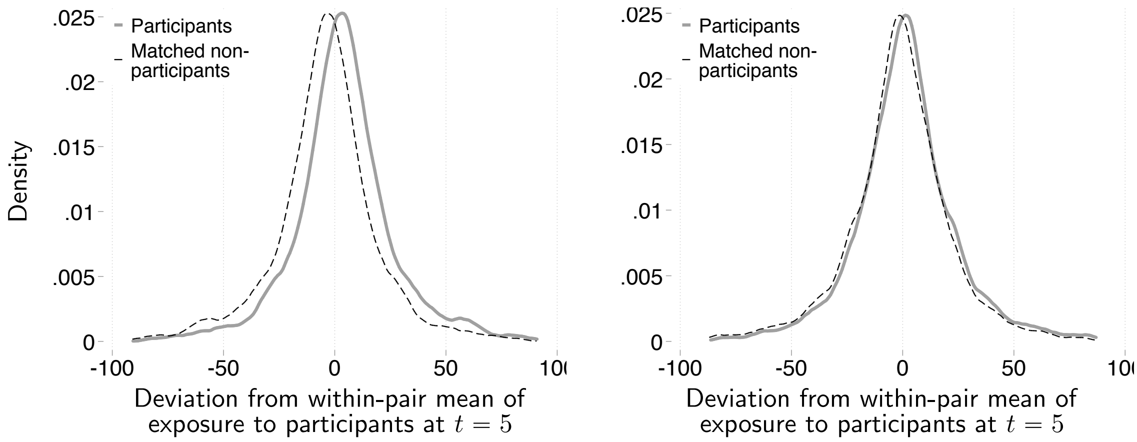
[21] In principle, it would be interesting to compare the effects of exposure at different time steps: does exposure at $t = 2$ matter more than exposure at $t = 5$? How much more? In practice, exposure is too highly correlated across time steps to allow us to make these distinctions, as Appendix Table 8 reveals.

Figure 2: Participants have higher network exposure than their observationally similar counterparts

Each figure plots the distribution of deviations from the within-pair mean of network exposure ($N^g_{i,5} - \overline{N^g_{pair,5}}$), separately for participants and matched non-participants.

(a) Protest (September 1) Overall mean of $N^g_{i,5} \approx 48.5$

(b) Petition signing Overall mean of $N^g_{i,5} \approx 48.2$



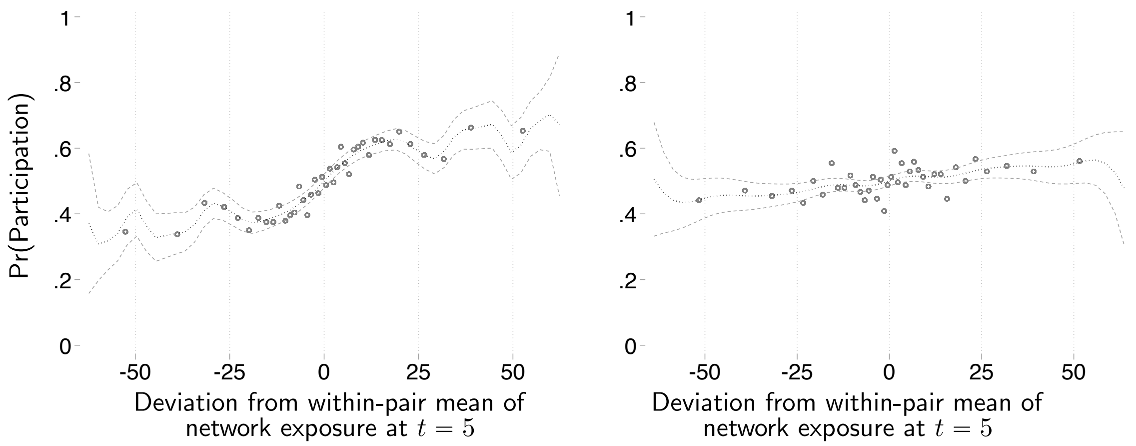
For visual clarity, both figures exclude observations in the top and bottom 1% of the distribution of within-pair differences in network exposure. We do not trim the sample when estimating Equation 2 below.

Figure 3: Exposure to participants predicts political participation

Figure (a) plots the probability that a person protests given her deviation from the within-matched-pair mean of network exposure ($N_{i,5} - \overline{N_{pair,5}}$). Figure (b) plots the probability that a person signs the recall referendum petition conditional on her deviation from the within-pair mean.

(a) Protest and network exposure

(b) Petition and network exposure



Lines shows fitted values (and 95% C.I.s) from local linear regression using a Gaussian kernel with the rule-of-thumb bandwidth from Fan and Gijbels (1996), p. 110–113. The regression is fit to the raw (unbinned) data. The points mark mean participation rates in bins of 2.4 percentiles (40 equally sized bins). For visual clarity, we exclude the top and bottom 2% (we do not trim the sample when estimating Equation 2).

4.2.2 Comparing protest to petition signing: what explains the difference?

One reason that we hesitate to interpret the pattern in Figure 3a as a causal relationship is that *sociability* is almost certainly correlated both with network exposure and with protest participation. A protest is an inherently social activity, usually attended in the company of friends, and thus protesters might have higher network exposure than similar non-protesters for reasons unrelated to social sanctioning or, for that matter, to any solution to the collective action problem. In other words, we might observe the same pattern if we were to compare (say) people at a nightclub to observationally similar homebodies, even though clubbing is not subject to the collective action problem.

Petition signing, in contrast, eludes the sociability confound. Like protests, signature drives suffer from the collective action problem: participation is costly, and everyone (participants and non-participants alike) benefits from success. The temptation to free ride on others' petition signatures might be especially strong in Venezuela, where signatories of an earlier petition were fired from government jobs and otherwise discriminated against (Hsieh et al., 2011). But unlike protesting, petition signing is not a social event.

Using the matched sample of petition signers and similar non-signers, Figure 3b plots the relationship between the probability of signing the petition and deviations from the within-pair mean of exposure to other petition signers $N^g_{i,5} - \overline{N^g_{pair,5}}$. Within matched pairs, petition signing rates increase more slowly with network exposure than do protest rates (more on this in Section 4.2.3 below); this suggests that, indeed, sociability may account for some of the observed relationship between network exposure and protest attendance. At the same time, the nonzero slope in Figure 3b suggests that sociability is not the whole story. Rather, the fact that network exposure predicts petition signing at all is consistent with the idea that network exposure facilitates social sanctioning and thereby discourages free riding.

Of course, sociability is not the only possible explanation for the difference in slopes in Figure 3—that is, for why protest participation increases faster with network exposure than petition signing increases with network exposure. Another possible explanation is that protesting might be more *visible*: perhaps people are more likely to know whether their peers protested than to know whether their peers signed a petition.

To evaluate this possibility, we included original questions on an in-person, nationally representative survey of Venezuelans conducted in September, 2018.[22] The results suggest that a person's decision to protest is no more or less visible than her decision to sign the petition: the proportion of people who report knowing whether a friend participated is similar for both activities.[23]

Nor is protesting more likely to incur social sanction (or social approbation). In a separate question, we asked whether friends or family members "told [the respondent] what they thought of [the respondent's] decision" to attend (or not attend) the protest, or whether "they kept their opinions to themselves." We asked the same question about petition signing. If people were more likely to opine about their friends' protest attendance (or non-attendance) than about their friends' petition signatures (or lack thereof), we might interpret the difference in slopes (Figure 3a vs. Figure 3b) as evidence simply that social sanctioning is a more common tool for protest mobilization than for petition drives. In fact, our survey data suggest the opposite: more respondents reported hearing others' opinions about their decision to sign (or not sign) the petition (44%) than about their decision to attend (or not attend) the protest (28%).

In addition to visibility, one may also worry that these results are driven by a link between extroverted personality types and a willingness to participate political. In other words, even if petition signing is not intrinsically social, if political participation is simply more likely among individuals with outgoing, social personalities, then it is possible that this relationship is driven by that correlation rather than social pressure. However, while political participation is often cited as being related to extroversion, recent work has shown this is only the case for social forms of political participation. In his study of personality types and political participation, for example, Mondak (2010, p. 159-160) finds that "extraversion operates as a strong determinant of the tendency to engage in those forms of political participation that involve social interaction, especially interaction in large groups, but the influence of extraversion on political participation dissipates when focus turns to more individualistic behaviors" (like putting up yard signs or contributing to political campaigns. See also Gerber et al. (2011)). As petition signing took place over multiple days at hundreds of locations across the country (reducing crowding), and involved minimal social interaction, we do not believe personality type is a likely confound in this case.

Taken together, Figure 3, the survey data, and past work on personality types

[22] The firm Datanálisis included our questions on their regular quarterly survey.

[23] We asked each respondent both about protesting and about petition-signing, randomizing the order of the questions. Results are nearly identical when we restrict the analysis to the question that respondents answered first. Appendix D presents the full survey instrument, in English and Spanish.

and political participation suggests two conclusions, both of which we explore in the rest of this paper. First, the probability that a person signs the petition increases with her network exposure, consistent with theories linking social network structure to political participation. Second, sociability is likely correlated both with network exposure and with protest attendance. This correlation affects the interpretation both of our protest results and of related results in the literature.[24]

4.2.3 Parametric estimates

Figure 3 reveals that the within-pair nonparametric relationship between network exposure and political participation is close to linear. With this in mind, we estimate:

$$\text{Participate}_i = \gamma_p + \beta N_{i,5} + \epsilon_i \tag{2}$$

where Participate_i is an indicator for whether person i participates (in the protest, in signing the petition, or in a placebo outcome, as indicated in each table); γ_p are fixed effects for each pair p, $N_{i,5}$ is the network exposure of person i at $t = 5$;[25] and ϵ_i is a person-specific shock term. This specification thus exploits *within-matched-pair* variation in network exposure (Mummolo and Peterson, 2018).

Columns (1) and (2) of Table 2 report the estimates of Equation 2. A one-standard deviation of within-pair general network exposure predicts an 7.7-percentage-point increase in protest participation, while increasing network exposure one within-pair standard deviation predicts a 2.8-percentage-point increase in the probability of signing the petition. Again, this comparison suggest that sociability may drive some—but not all—of the observed relationship between network exposure and protest participation.

[24]For example, Enikopolov et al. (2017) find that an online social network (similar to Facebook) increased protest participation in Russia, interpreting this result as evidence that the social network helped "solve the collective action problem" by "reducing the costs of coordination." But if the social network also increased attendance at (say) concerts or dinner parties, we might interpret their results instead as evidence that the social network simply facilitated social activities, whether or not those activities were subject to the collective action problem. These two interpretations have different political implications: if the social network helped solve the collective action problem, we might expect that it would also enable other types of political activities; if it merely facilitates social events, we would not expect that it would affect other, less-social political activities, like voting. And indeed, Enikopolov et al. (2017) find no evidence that the social network increased anti-government votes.

[25]As noted above, results are similar for other time steps.

Table 2: Network Exposure Predicts Political Participation
Predicted change in probability of participation associated with a one-standard-deviation increase in network exposure (that is, one s.d. of the within-pair distribution of network exposure), based on estimates of Equation 2.

	Protest (1)	Petition (2)
General network exposure (N_i^g)	0.077	0.028
	(0.01)	(0.010)
N	10K	10K

Standard errors, clustered by pair, in parentheses.

4.2.4 Unobserved dimensions of socio-economic status

A potential concern is that, despite our effort to match protesters to comparable non-protesters, the slope in Figure 3a might nevertheless capture the effect of some unobserved dimension of socioeconomic status that remains unbalanced across the two groups (and that is correlated with network exposure). For example, because we exclude anyone who lives in eastern Caracas from our sample of protestors (to avoid false-negatives in the identification of protestors) we know that everyone in our protester sample travelled to reach the protest. Consequently we might expect that our protester sample is richer than matched non-protesters (again, despite matching on observables correlated with income).

To evaluate this possibility, we repeat our core analysis with a placebo outcome that should be associated with socioeconomic status but not (necessarily) with general network exposure: presence in the area of Caracas where the protest took place, but on non-protest days ("placebo dates"). This entails (a) drawing a new sample of "placebo protestors" (individuals in the protest area on a set of non-protest days), (b) drawing a new sample of individuals who are observationally similar to these placebo protestors but who were not in the protest area on the placebo date, (c) measuring network exposure for these individuals, and (d) comparing network exposure for those two groups by re-estimating Equation 2 for these populations.

If our matching strategy failed to capture important dimensions of socioeconomic status, we would expect people who travel to the protest area in eastern Caracas on *any* date to have higher network exposure than their matched counterparts. Instead, as the estimates in Table 3 reveal, network exposure is much less predictive of the placebo outcome than it is of protest participa-

Table 3: It's not just an unobserved dimension of socio-economic status

Predicted change in probability of participation associated with a one-standard-deviation increase in network exposure (that is, one s.d. of the within-pair distribution of network exposure), based on estimates of Eqn. 2.

	8/4	8/11	8/18	8/25	9/1	9/8	9/15	9/22	9/29
General network exposure (N_i^g)	0.031	0.016	0.003	0.013	0.077	0.006	0.010	-0.000	0.013
	(0.010)	(0.01)	(0.010)	(0.010)	(0.01)	(0.010)	(0.010)	(0.010)	(0.010)
N	10K	10K	10K	10K	10K	10K	10K	10K	10K

Standard errors, clustered by matched pair, in parentheses.

tion: moving up standard deviation in the distribution of network exposure predicts a 5,750.0-percentage-point increase in the probability of presence in Caracas on placebo dates (compared to 7.7 percentage points for actual protest participation).[26]

4.2.5 Network exposure to other participants

One concern the analysis thus far is that we have focused on *general network exposure*, or the total number of peers who are likely to hear about a person's behavior, whereas (in theory) only *participants* have the authority to sanction non-participation. Who are we to criticize non-participation if we stayed home?

In this section, we show that the results are robust to focusing instead on network exposure *to participants*—that is, the number of protest attendees or petition signers likely to hear about a person's behavior.

Table 4 shows the relationship between participation and *exposure to participants*. Clearly, exposure to participants does predict participation—indeed, within matched pairs, it predicts an even larger change in participation rates. Columns (2) and (4) report the estimates: a (within-pair) one-standard-deviation increase in network exposure produces a 20.1-percentage-point increase in protest participation and a 4.9-percentage-point increase in the probability of signing the petition (compared to 7.7-percentage-points and 2.8-percentage-points, respectively, for one-standard-deviation changes in general network exposure).

[26] In principle, we could conduct an analogous placebo-day analysis using network exposure to participants, rather than general network exposure. In practice, though, we would not expect these placebo-day estimates to be zero: assuming some homophilic tendency among people who travel to downtown Caracas on a given day, network exposure to "participants" (scare quotes indicating placebo-day participants) *should* predict "participation."

Table 4: Network Exposure Predicts Political Participation

Predicted change in probability of participation associated with a one-standard-deviation increase in network exposure (that is, one s.d. of the within-pair distribution of network exposure), based on estimates of Eqn. 2.

	Protest		Petition	
	(1)	(2)	(3)	(4)
General network exposure (N_i^g)	0.077		0.028	
	(0.01)		(0.010)	
Exposure to participants (N_i^p)		0.201		0.049
		(0.01)		(0.010)
N	10K	10K	10K	10K

Standard errors, clustered by pair, in parentheses.

While these results are consistent with our claims, this is not our preferred specification. First, it is unclear whether the moral authority to sanction is restricted to participants; people unable to attend (for health or work reasons) might nevertheless pressure others to participate. Likewise, participants and non-participants alike can provide social approbation as a reward for participating. And moreover, any tendency for participants to preferentially befriend other participants (a form of homophily) would inflate the correlation between exposure to participants and participation rates.

4.3 Political consequences of social network structure

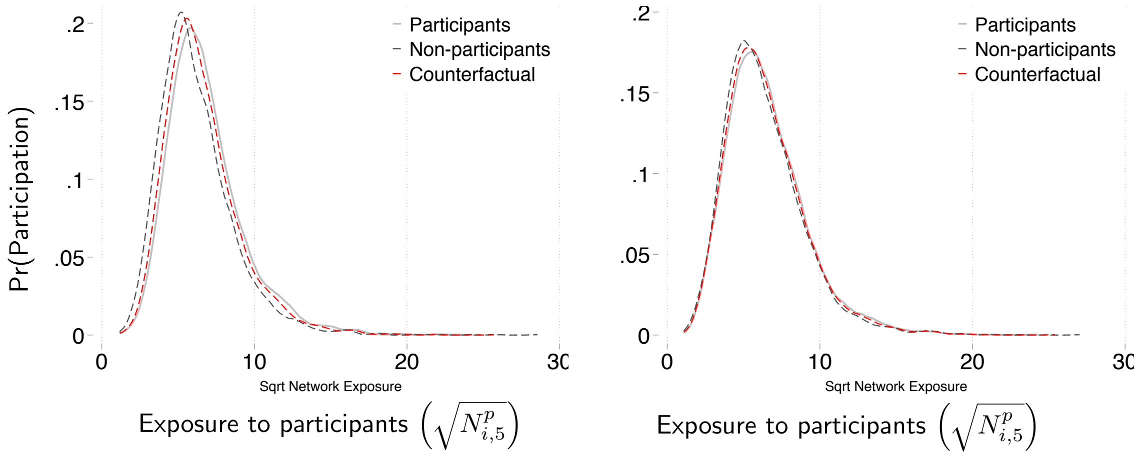
While these simple comparisons are informative, it is unclear how to gauge the *magnitude* of the differences. In other words, when we learn that a person with an extra standard deviation of network exposure at $t = 5$ is 7.7 percentage points more likely to participate than an otherwise comparable person with the least network exposure, what do we conclude (beyond the fact that the difference is positive and nonzero)?

Before considering the magnitude (rather than just the sign) of the effect of network exposure on participation, we note that the estimates in Table 2 are specific to the population of people with observable characteristics similar to those of protesters (or to be even more precise, protesters who do not reside in eastern Caracas). That is because unlike in our analysis in Section 4.1, the results in Table 2 is estimated using protestors and a matched sample of individuals who are observationally similar to protestors (without re-weighting). As this population is likely characteristic of the *mobilizable* population (or "population at risk of protesting"), we see that it is a population of political interest (not just of econometric convenience).[27]

Figure 4: What if non-participants had greater exposure to participants?

These figures plots the distribution (across individuals in the data) of network exposure (at $t = 5$), separately for protest participants and a set of non-participants who are similar on observables. The dotted lines mark possible counterfactual distributions in which politically isolated non-participants are integrated into the network.

(a) Protest and network exposure, $t = 5$ (b) Petition and network exposure, $t = 5$



For visual clarity we plot the distribution of $\sqrt{N_{i,5}}$ rather than $N_{i,5}$.

Because the estimates in Table 2 come from samples in which we matched participants and non-participants on all observables in our data, we cannot compare the magnitude of these differences to how protest probability changes with (say) proximity to Caracas or neighborhood education level. But we can get a sense of the political significance of the magnitude of these effects by considering a counterfactual: what if non-participants knew more participants?

Figure 4 plots the distribution (across individuals in the data) of network exposure, separately for participants and (matched) non-participants. The dotted lines plot a possible counterfactual: a world in which party leaders or participants bring politically isolated non-participants into the fold.[28] A counterfactual along these lines is especially relevant to the Venezuelan case, where, for most of the presidencies of Hugo Chávez and Nicolás Maduro, the political opposition was infamously divided: divided over whether to participate in or abstain from elections; divided over whether to take to the streets or make demands via other channels (Trinkunas, 2018).

[27]Because protestors are such a small share of the population, the simple strategy of drawing a random sample of Venezuelans and drawing a random matched sample is simply not viable. It is for that reason we use a re-weighting approach in Section 4.1.

[28]Specifically, the counterfactual marks a convex combination of the two observed distributions, in which the weight on the *participants* distribution declines with network exposure.

In this world, protest participation in the mobilizable population would be 0.7 percentage points higher, and petition signing in the mobilizable population would be 0.1 percentage points higher.

If we imagine that the mobilizable population is four times as large as the number of participants—for every participant, there are four similar potential participants—the protest would be 2.9% larger in the counterfactual world, and the petition would have 0.4% more signatures (percent, not percentage points). In our view, these differences are politically meaningful. Certainly, if we were to recklessly extrapolate these estimates to the collective political activity of voting, differences of this size would have avoided the election of Nicolás Maduro in 2013: a 1.9% increase in opposition turnout would have reversed the outcome.

5 Discussion

Until recently, it was impossible to observe social networks at scale. That rendered unknowable many facts about the role of social ties in solving collective action problems. In this paper, we take advantage of this newly available data to study the relationship between a person's network exposure—how quickly others hear about her—on political participation. We use cell-phone meta-data to map the full structure of social networks for the entire country of Venezuela. We then use this exceptional data to develop a *theoretically-derived* measure of network structure, and pair it with behavioral data on two categorically different forms of political participation in order to test the relationship between network exposure and political participation. To help narrow the space for unobserved heterogeneity, we then analyze this relationship by comparing a random sample of protesters and signatories with a sample of individuals who are observationally similar across a much richer range of demographic characteristics than has ever been possible before. We find non-participants who have similar observable characteristics. Within this matched sample, network exposure strongly predicts protest participation. We use placebo tests to address concerns about unobservables. Beyond establishing the sign of the effect of network exposure (positive), we consider the magnitude, learning that slight changes in the social network might improve political outcomes for the Venezuelan opposition. We use qualitative information and survey data to argue that social sanctioning drives the relationship between network exposure and political participation.

Some aspects of the Venezuelan case are common to all settings for collective action. Other aspects are unique, or at least unusual. For example, few protests are as large as the one studied in this paper, and we might imagine that the power of social sanctioning changes with protest size. And as previously noted, we suspect that in many contexts where protests are less widely advertised in the public media, or where risk of government disruption if protest plans are made publicly, networks will play a number of roles *in addition* to supporting social sanction regimes. In general, however, we have little reason to believe that the ability of networks to support peer pressure documented here is not a generalizable result.

Together, these results provide rigorous empirical evidence not only that "networks matter" for political participation, but that at least one of the ways in which they matter is for facilitating peer pressure regimes.

References

Albertus, M. (2015). The role of subnational politicains in distributive politics: political bias in venezuela's land reform under chávez. *Comparative Political Studies*.

Anoll, A. P. (2018). What Makes a Good Neighbor? Race, Place, and Norms of Political Participation. *American Political Science Review*, 112(3):494–508–508.

Banerjee, A., Chandrasekhar, A., Duflo, E., and Jackson, M. O. (2014). Gossip: Identifying Central Individuals in a Social Network. *SSRN Working Paper Series*, 20422.

Booth, J. A. and Seligson, M. A. (2006). Inequality and democracy in latin america: Individual and contextual effects of wealth on political participation. *Duke University Workshop on Poverty and Democracy*.

Centola, D. (2018). *How Behavior Spreads: The Science of Complex Contagions*. Princeton University Press, Princeton.

Centola, D. and Macy, M. (2007). Complex Contagions and the Weakness of Long Ties. *American Journal of Sociology*, 113(3):702–734.

Christensen, D. and Garfias, F. (2018). Can You Hear Me Now?:How Communication Technology Affects Protest and Repression. *Quarterly Journal of Political Science*.

Datanalisis (2016). Encuesta ómnibus. *Primer Trimestre*.

Dionne, K. Y. (2015). Social Networks, Ethnic Diversity, and Cooperative Behavior in Rural Malawi. *Journal of Theoretical Politics*, 27(4):522–543.

Enikopolov, R., Makarin, A., and Petrova, M. (2017). Social media and protest participation: Evidence from russia. *Working Paper*.

Eubank, N., Grossman, G., Platas, M., and Rodden, J. (2018). Social Networks and Turnout in Low Salience Elections: Evidence from Uganda. *Working Paper*.

Fafchamps, M. and Vicente, P. C. (2013). Political Violence and Social Networks: Experimental Evidence from a Nigerian Election. *Journal of Development Economics*, 101:27–48.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. CRC Press.

Fearon, J. D. and Laitin, D. (1996). Explaining Interethnic Cooperation. *American Political Science Review*, 90(4):715–735.

Ferrali, R., Grossman, G., Platas, M., and Rodden, J. (2018). Peer effects and externalities in technology adoption: Evidence from community reporting in uganda. *Working Paper*.

Gerber, A. S., Green, D. P., and Larimer, C. W. (2008). Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment. *American Political Science Review*, 102(1):33–48.

Gerber, A. S., HUBER, G. A., Doherty, D., and Dowling, C. M. (2011). The Big Five Personality Traits in the Political Arena. *Annual Review of Political Science*, 14(1):265–287.

Gerber, A. S., Huber, G. A., Doherty, D., and Dowling, C. M. (2016). Why People Vote: Estimating the Social Returns to Voting. *British Journal of Political Science*, 46(2):241–264.

Gonzalez, F. (2017). Collective action in networks: Evidence from the chilean student movement. *Working Paper*.

Granovetter, M. S. (1973). The Strength of Weak Ties. *American Journal of Sociology*, 78(6):1360–1380.

Haythornthwaite, C. (2005). Social Networks and Internet Connectivity Effects. *Information, Communication & Society*, 8(2):125–147.

Hsieh, C.-T., Miguel, E., Ortega, D., and Rodriguez, F. (2011). The price of political opposition: Evidence from venezuela's maisanta. *American Economic Journal: Applied Economics*.

Jackson, M. O. (2008). *Social and Economic Networks*. Princeton University Press.

Jackson, M. O. (2018). A typology of social capital and associated network measures. *Unpublished Manuscript*.

Jackson, M. O., Rodriguez-Barraquer, T., and Tan, X. (2012a). Social capital and social quilts: Network patterns of favor exchange. *American Economic Review*.

Jackson, M. O., Rodriguez-Barraquer, T., and Tan, X. (2012b). Social Capital and Social Quilts: Network Patterns of Favor Exchange. *The American Economic Review*, 102(5):1857–1897.

Jackson, M. O., Rogers, B. W., and Zenou, Y. (2017). The economic consequences of social-network structure. *Journal of Economic Literature*.

Kandori, M. (1992). Social Norms and Community Enforcement. *The Review of Economic Studies*, 59(1):63–80.

Kronick, D. (2015). Why would a president wreck the economy before an election? *The Washington Post*.

Kuran, T. (1991). Now out of never: The element of surprise in the East European revolution of 1989. *World Politics*, 44(01):7–48.

Larson, J. M. (2017). Why the West Became Wild. *World Politics*, 69(4):713–749.

Larson, J. M. and Lewis, J. I. (2017). Ethnic Networks. *American Journal of Political Science*, 61(2):350–364.

Larson, J. M., Nagler, J., Ronen, J., and Tucker, J. A. (2019). Social Networks and Protest Participation: Evidence from 130 Million Twitter Users . *Forthcoming, American Journal of Political Science*.

Little, A. T. (2015). Communication Technology and Protest. *Journal of Politics*, 78(1):152–166.

Miritello, G., Moro, E., Lara, R., Martínez-López, R., Belchamber, J., Roberts, S. G. B., and Dunbar, R. I. M. (2013). Time as a limited resource: Communication strategy in mobile phone networks. *Social Networks*, 35(1):89–95.

Mondak, J. J. (2010). *Personality and the Foundations of Political Behavior*. Cambridge University Press, Cambridge, MA.

Motahari, S., Appala, S., Mengshoel, O. J., Zoia, L., Reuther, P., and Shah, J. (2012). The Impact of Social Affinity on Phone Calling Patterns: Categorizing Social Ties from Call Data Records - Semantic Scholar. In *The Sixth SNA-KDD Workshop*.

Mummolo, J. and Peterson, E. (2018). Improving the interpretation of fixed effects regression results. *Political Science Research and Methods*.

Newman, M., Barabási, A.-L., and Watts, D. J. (2006). ***The Structure and Dynamics of Networks***. Princeton University Press.

Noticia al Día (2016). Así se desarrolla en maracaibo el segundo día de recolección de firmas para el revocatorio. *https://bit.ly/2M8bFwP*.

Olson, M. (1965). *The Logic of Collective Action*. Harvard University Press.

Onnela, J.-P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., Kertész, J., and Barabási, A.-L. (2007). Structure and Tie Strengths in Mobile Communication Networks. *Proceedings of the National Academy of Sciences*, 104(18):7332–7336.

Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press.

Panagopoulos, C. (2010). Affect, Social Pressure and Prosocial Motivation: Field Experimental Evidence of the Mobilizing Effects of Pride, Shame and Publicizing Voting Behavior. *Political Behavior*, 32(3):369–386.

Putnam, R. D., Leonardi, R., and Nanetti, R. (1993). *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton University Press, Princeton, New Jersey.

Rodriguez, F. (2016). Chronicle of a march foretold. *Torino Capital Venezuela this Week*.

Rodriguez, F. (2017). As u.s. sanctions maduro and hints at regime change, a debate on resolving the crisis in venezuela. *Democracy Now*.

Rojo, G. and Wibbels, E. (2014). Clientelism, Public Services and Elections in the Slums of Jaipur and Udaipur. *Presented at Stanford Comparative Politics Seminar*.

Rolfe, M. (2012). *Voter Turnout: A Social Theory of Political Participation*. Cambridge University Press, Cambridge.

Siegel, D. A. (2009). Social Networks and Collective Action. *American Journal of Political Science*, 53(1):122–138.

Sinclair, B. (2012). *The Social Citizen: Peer Networks and Political Behavior*. The University of Chicago Press, Chicago.

Steinert-Threlkeld, Z. C. (2018). Activism and Protest in Authoritarian Regimes. *Working Paper*.

Toro, F. (2014). Rash repression in venezuela. *The New York Times*.

Toro, F. (2016). Signed, sealed, and delivered. *Caracas Chronicles*.

Trinkunas, H. (2018). The tragedy of the venezuelan opposition. Foreign Affairs.

Valesco, A. (2015). *Barrio Rising: **Urban Popular Politics and the Making of Modern Venezuela***. University of California Press, Oakland, CA.

Wiese, J., Min, J.-K., Hong, J. I., and Zimmerman, J. (2015). *You Never Call, You Never Write: Call and SMS Logs Do Not Always Indicate Tie Strength*. Call and SMS Logs Do Not Always Indicate Tie Strength. ACM, New York, New York, USA.

Wolitzky, A. (2012). Cooperation with network monitoring. *Review of Economic Studies*.

A Network Specification

A.1 Motivation for Unweighted Network Specification

Two considerations motivated our decision not to use the frequency or duration of calls as weights, as other studies have done (Onnela et al., 2007; Miritello et al., 2013).

First, it is unclear whether frequency of communication is a good indicator of influence in a relationship. Within a given type of relationship (e.g. among co-workers), there is some evidence that frequency of communication in one electronic medium is a useful proxy for intensity of communication across all mediums (Haythornthwaite, 2005), but whether this holds across types of relationships is unknown.[29]

For example, people may make more calls to co-workers and business partners than to family members—even if the family relationships are more influential. For example, using data from California, Motahari et al. (2012) shows that calling patterns among family members are qualitatively different from calling patterns with others: calls to family members are more frequent but shorter. These findings illustrate how the mapping from call frequency or duration to significance-of-tie may vary across types of connections. Similarly, in a survey of 40 U.S. individuals who agreed to share phone records and fill out questionnaires about their connections, Wiese et al. (2015) finds that while call frequency and duration do predict self-reported tie strength, "many people in all tie strength levels had very little communication" (Wiese et al., 2015, p.5). Wiese concludes that this is driven by substitution to in-person communication, substitution to email or other non-phone communication, and the fact that "[f]amily is close regardless of communication" (p. 7).

Second, in the Venezuelan context we know that a non-trival share of communications take place via WhatsApp, and thus do not appear in our data. As these communications are especially likely among younger users, using text and phone frequency as a measure of importance of connections would necessarily privilege connections among older people, potentially biasing results. Indeed, fear of excluding ties among younger users is one reason that we use such a low threshold for connection inclusion.

[29]Indeed, (Haythornthwaite, 2005, p. 125) concludes only "that media use *within groups* conformed to a unidimensional scale." (Emphasis added.)

A.2 Motivation for Undirected Network Specification

We treat connections as undirected because, first, information exchange in phone communications is inherently bi-directional. Second, and perhaps more importantly, the direction of a call can be surprisingly difficult to establish in the Venezuelan context. In Venezuela, the cost is borne by the person placing the call. As a result, many users engage in the practice of giving more affluent contacts a "missed call" (they call, let the phone ring once, then hang up) as a signal that they would like the more affluent contact to call them back, allowing the more affluent party to be billed. These missed calls do not appear in the data (call detail records are collected primarily for billing purposes, and missed calls aren't billed), so many bi-directional relationships may appear uni-directional in the data.

B Matching Strategy

For each protester p, we locate an observationally similar match m using the following recursive matching algorithm:

1. First, we create a set of potential matches including all individuals who subscribe to our Partner Telecom[30] and are registered to vote as the same polling place as p.
2. The algorithm looks for individuals who are perfect (exact) matches for p in terms of:
   - political party (registered member of the PSUV, or not, `PSUV`
   - whether they spent any weekdays in Caracas in the month preceding or following the protest (excluding the week of the protest) (binary, `any_weekday_in_caracas`)
   - whether they spent any weekends in Caracas in the month preceding or following the protest (excluding the week of the protest) (binary, `any_weekend_in_caracas`)
   - exact number of weekdays in Caracas in the month preceding or following the protest (excluding the week of the protest, `in_caracas_weekday`)
   - exact number of weekend days in Caracas in the month preceding or following the protest (excluding the week of the protest,

[30] Recall that all identified protestors are also Partner Telecom subscribers, because real-time geo-location is only available for Partner Telecom subscribers.

`in_caracas_weekend`)

- gender (binary, from voter registration data, `registration_female`)
- whether the user has a pre-paid or post-paid cellular plan (related to socio-economic status) (`post`)

3. If there is at least one person who is a perfect match for p along these dimensions (which happens in $\sim$ 70% of cases), then a distance score is calculated with respect to normalized measures of age and spatial mobility,[31] and the individual with the lowest distance score becomes m.

4. If no individuals are identified who perfectly match p along the exact-match attributes, then the bottom-most attribute is moved from the list of attributes for which an exact-match is sought and is moved into the list of attributes for which a distance score is calculated. Steps 2 and 3 are then repeated until a single match is identified.

Success rates for various exact-match attributes can be found in Tables 5- 7.

Table 5: Successful Matches on Exact-Match Attributes, Sept 1st

	Num exact matched by var
post	3515
registration_female	3539
in_caracas_weekend	3877
in_caracas_weekday	4057
any_weekend_in_caracas	4832
any_weekday_in_caracas	4946
psuv	5000
total matches	5000

C Additional tables and figures

[31] The day-to-day variance in the location of the user.

Table 6: Successful Matches on Exact-Match Attributes, PSUV

	Num exact matched by var
post	4792
registration_female	4805
in_caracas_weekend	4863
in_caracas_weekday	4895
any_weekend_in_caracas	4980
any_weekday_in_caracas	4989
psuv	5000
total matches	5000

Table 7: Successful Matches on Exact-Match Attributes, MUD

	Num exact matched by var
post	4700
registration_female	4739
in_caracas_weekend	4803
in_caracas_weekday	4839
any_weekend_in_caracas	4961
any_weekday_in_caracas	4991
psuv	5000
total matches	5000

Table 8: Network exposure at different time steps and eigenvector centrality: correlations

	$N^a_{i,1}$	$N^a_{i,2}$	$N^a_{i,3}$	$N^a_{i,4}$	$N^a_{i,5}$	$N^a_{i,6}$	$N^a_{i,7}$	$N^a_{i,8}$	$N^a_{i,9}$	$N^a_{i,10}$	$N^a_{i,11}$	$N^a_{i,12}$	$N^a_{i,13}$	$N^a_{i,14}$	$E.C.$
$N^a_{i,1}$	1.000														
$N^a_{i,2}$	0.980	1.000													
$N^a_{i,3}$	0.940	0.989	1.000												
$N^a_{i,4}$	0.896	0.965	0.993	1.000											
$N^a_{i,5}$	0.854	0.936	0.978	0.996	1.000										
$N^a_{i,6}$	0.817	0.908	0.959	0.985	0.997	1.000									
$N^a_{i,7}$	0.784	0.882	0.940	0.973	0.990	0.998	1.000								
$N^a_{i,8}$	0.757	0.860	0.923	0.960	0.982	0.994	0.999	1.000							
$N^a_{i,9}$	0.736	0.842	0.908	0.949	0.974	0.988	0.996	0.999	1.000						
E.C.	-0.006	-0.009	-0.011	-0.012	-0.013	-0.014	-0.014	-0.015	-0.015	1.000					

D Survey instrument

The order of the two blocks was randomly assigned across respondents.

D.1 Block 1, English

1. In 2016, opposition parties collected signatures on a petition requesting a recall referendum. Hundreds of thousands of people signed the petition, but millions chose not to sign, and many others wanted to sign but just didn't have time or energy. Did you happen to sign the 2016 recall referendum petition?

2. Did any of your friends or family members tell you what they thought of your decision [to sign the petition] [not to sign the petition], or did they keep their opinions to themselves? [Read responses]
   - They told me what they thought
   - They kept their opinions to themselves
   - No response

3. We're interested in whether people talked to their friends about whether to sign the recall referendum petition in 2016, or whether they kept their decisions to themselves. Think for a minute of your three closest friends. I will not ask you whether or not they signed the recall referendum petition, but I'm curious whether you know one way or the other. [Read responses]
   - Thinking of the first of the three friends you have in mind, do you know whether he or she signed? [Read responses]
     - Yes, I know whether or not he or she signed the petition
     - No, I don't know
     - No response

- And now thinking of the second of the three friends you have in mind, do you know whether he or she signed?
  - Yes, I know whether or not he or she signed the petition
  - No, I don't know
  - No response
- And now thinking of the third of the three friends you have in mind, do you know whether he or she signed?
  - Yes, I know whether or not he or she signed the petition
  - No, I don't know
  - No response

D.2 Block 2, English

1. In 2016, opposition parties organized protests to pressure the government to hold a recall referendum. Hundreds of thousands of people attended, but millions chose not to attend, and many others wanted to attend but just didn't have time or energy. Did you happen to attend the recall referendum protests in 2016, such as the Toma de Caracas?

2. Did any of your friends or family members tell you what they thought of your decision [to attend the protest] [not to attend the protest], or did they keep their opinions to themselves? [Read responses]
   - They told me what they thought
   - They kept their opinions to themselves
   - No response

3. We're interested in whether people talked to their friends about whether to attend the recall referendum protests in 2016, or whether they kept their decisions to themselves. Think for a minute of your three closest friends. I will not ask you whether or not they attended the protests in 2016, but I'm curious whether you know one way or the other. [Read responses]
   - Thinking of the first of the three friends you have in mind, do you know whether he or she protested? [Read responses]
     - Yes, I know whether or not he or she protested
     - No, I don't know
     - No response
   - And now thinking of the second of the three friends you have in mind, do you know whether he or she protested?
     - Yes, I know whether or not he or she sprotested

    - No, I don't know
    - No response
- And now thinking of the third of the three friends you have in mind, do you know whether he or she protested?
    - Yes, I know whether or not he or she protested
    - No, I don't know
    - No response

D.3 Block 1, Spanish

1. En 2016 los partidos de oposición recogieron firmas para solicitar un referéndum revocatorio. ¿Firmó usted la petición del referéndum revocatorio de 2016?
    - Sí
    - No
    - No contesta
2. ¿Alguno de sus amigos o familiares le dijeron qué pensaban sobre su decisión [de firmar la peticón] [no firmar la petición], o fueron reservados con respecto a sus opiniones? (Enc. Leer opciones. Aceptar una sola respuesta)
    - Me dijeron lo que pensaban
    - Fueron reservados con respeto a sus opiniones
    - No contesta
3. Nos interesa saber si las personas hablaron con sus amigos sobre la firma de la petición para el referéndum revocatorio en 2016. Piense por un minuto en sus tres amigos más cercanos, recuerde que no queremos saber qué hicieron sus amigos, nos interesa saber si ellos compartieron con usted la decisón que tomaron.
    - Pensando en el primero de los tres amigos que tiene en mente, ¿sabe si firmó? (Enc. Leer opciones. Aceptar una sola respuesta)
        - Sí, sé si firmó o no la petición
        - No, no lo sé
        - No contesta
    - Y ahora, pensando en el segundo de los tres amigos que tiene en mente, ¿sabe si firmó? (Enc. Leer opciones. Aceptar una sola respuesta)
        - Sí, sé si firmó o no la petición
        - No, no lo sé

    - No contesta

- Y ahora, pensando en el tercero de los tres amigos que tiene en mente, ¿sabe si firmó? (Enc. Leer opciones. Aceptar una sola respuesta)
    - Sí, sé si firmó o no la petición
    - No, no lo sé
    - No contesta

D.4 Block 2, Spanish

1. En 2016 los partidos de oposición organizaron protestas para presionar al gobierno a celebrar un referéndum revocatorio. ¿Asistió usted a las protestas a favor del referéndum revocatorio en 2016, como la denominada Toma de Caracas?
    - Sí
    - No
    - No contesta
2. ¿Alguno de sus amigos o familiares le dijeron qué pensaban sobre su decisión [de asisir a la protesta] [no asistir a la protesta], o fueron reservados con respecto a sus opiniones? (Enc. Leer opciones. Aceptar una sola respuesta)
    - Me dijeron lo que pensaban
    - Fueron reservados con respeto a sus opiniones
    - No contesta
3. Nos interesa saber si las personas hablaron con sus amigos sobre si asistirían o no a las protestas referéndum revocatorio en 2016. Piense por un minuto en sus tres amigos más cercanos, recuerde que no queremos saber qué hicieron sus amigos, nos interesa saber si ellos compartieron con usted la decisón que tomaron.
    - Pensando en el primero de los tres amigos que tiene en mente, ¿sabe si protestó? (Enc. Leer opciones. Aceptar una sola respuesta)
        - Sí, sé si protestó o no
        - No, no lo sé
        - No contesta
    - Y ahora, pensando en el segundo de los tres amigos que tiene en mente, ¿sabe si protestó? (Enc. Leer opciones. Aceptar una sola respuesta)
        - Sí, sé si protestó o no
        - No, no lo sé

  - No contesta
- Y ahora, pensando en el tercero de los tres amigos que tiene en mente, ¿sabe si protestó? (Enc. Leer opciones. Aceptar una sola respuesta)
  - Sí, sé si protestó o no
  - No, no lo sé
  - No contesta