AJPS AMERICAN JOURNAL of POLITICAL SCIENCE

Social Networks and Protest Participation: Evidence from 130 Million Twitter Users

Jennifer M. Larson Vanderbilt University
Jonathan Nagler New York University
Jonathan Ronen Berlin Institute for Medical Systems Biology
Joshua A. Tucker New York University

Abstract: *Pinning down the role of social ties in the decision to protest has been notoriously elusive, largely due to data limitations. Social media and their global use by protesters offer an unprecedented opportunity to observe real-time social ties and online behavior, though often without an attendant measure of real-world behavior. We collect data on Twitter activity during the 2015* Charlie Hebdo *protest in Paris, which, unusually, record real-world protest attendance and network structure measured beyond egocentric networks. We devise a test of social theories of protest that hold that participation depends on exposure to others' intentions and network position determines exposure. Our findings are strongly consistent with these theories, showing that protesters are significantly more connected to one another via direct, indirect, triadic, and reciprocated ties than comparable nonprotesters. These results offer the first large-scale empirical support for the claim that social network structure has consequences for protest participation.*

Replication Materials: The data and materials required to verify the computational reproducibility of the results, procedures and analyses in this article are available on the *American Journal of Political Science* Dataverse within the Harvard Dataverse Network, at: https://doi.org/10.7910/DVN/RLLL1V.

On January 7, 2015, gunmen killed 12 people at the offices of the French satirical magazine *Charlie Hebdo*. Four days after the terrorist attack, millions took to the streets in protest. That a protest took place in response to such a tragedy is not puzzling. However, any individual's choice of behavior that day, the sum of which aggregated into a large-scale protest, *is* puzzling, and relates to a larger question with wide-reaching implications for the social sciences: Why do some decide to attend a protest while others do not, and how do they arrive at this decision?

Despite how foundational this question is to an understanding of protests and collective action more broadly, how consequential protests can be given their role in policy change and the overthrow of governments, and how salient the topic has become due to the worldwide wave of twenty-first-century protests, the answer remains elusive.

The problem concerns data. Conventional wisdom suggests that an individual's decision to protest depends on the decisions of others in her social network, a logic that underlies much of the existing theory on protest (Centola 2013; Chwe 2000; Granovetter 1978; Kim and Bearman 1997; Marwell, Oliver, and Prahl 1988; Siegel 2009). However, tests of this wisdom require data that are fundamentally difficult to collect. Traditional methods entail tracking down protest participants after the fact and rely on their recall of behavior and *ex ante* motivations. To precisely test theoretical claims about networks, researchers must measure a full set of ties—protesters' ties

Jennifer M. Larson is Associate Professor, Department of Political Science, Vanderbilt University, 230 Appleton Place, Nashville, TN 37203 (jennifer.larson@vanderbilt.edu). Jonathan Nagler is Professor, Department of Politics, New York University, 19 W. 4th Street, New York, NY 10012 (jonathan.nagler@nyu.edu). Jonathan Ronen is PhD Fellow, Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine, Robert-Rössle Straβe, Berlin, Germany 13125 (yablee@gmail.com). Joshua A. Tucker is Professor, Department of Politics, New York University, 19 W. 4th Street, New York, NY 10012 (joshua.tucker@nyu.edu).

The data were collected by the NYU Social Media and Political Participation (SMaPP) laboratory (https://wp.nyu.edu/smapp/), of which Nagler and Tucker are co-directors along with Richard Bonneau and John T. Jost. The SMaPP lab is supported by the INSPIRE program of the National Science Foundation (Award SES-1248077), the New York University Global Institute for Advanced Study, the Moore-Sloan Data Science Environment, and Dean Thomas Carew's Research Investment Fund at New York University.

American Journal of Political Science, Vol. 63, No. 3, July 2019, Pp. 690–705

 DOI: 10.1111/ajps.12436

to other protesters, ties to nonprotesters, ties of ties—as well as the protest behavior, usually relayed secondhand, of everyone. Collecting this kind of information is error-prone and expensive, often prohibitively so in the context of protests. Consequently, few empirical studies to date consider the role of social networks in protest participation, and those that do are highly limited.[1]

We overcome this limitation by collecting data on Twitter use during the *Charlie Hebdo* protest.[2] Social media use generates an observable record of social ties that is complete on the platform and unfiltered by the memory of a respondent. Our data include user geolocation, which indicates physical presence or absence at the protest. Although others have explored the relationship between networks among Twitter users and *online* protest behavior (see, e.g., Barberá et al. 2015; González-Bailón et al. 2011; González-Bailón and Wang 2016), our data provide a rare window into the relationship between fine-grained network features and *offline* protest participation.

Our data set contains 764 people whose geolocation reveals that they were a participant in the Paris protest. Our data set also includes a set of 764 people who were tweeting about *Charlie Hebdo* in the same way as those who participated, but who were elsewhere in Paris during the protest. The latter serves as a baseline for controlled comparisons of protesters and similar nonprotesters. Furthermore, we collect full social network information measured out to two degrees—every user each of these people follows on Twitter, and every user each of these users follows. In total, our data set of protesters and nonprotesters, all users they follow, and all users whom those users follow contains 129,665,566 Twitter users.

These data provide a snapshot of the networks of protesters and the networks of nonprotesters who were similarly interested in the protest issue and eligible to attend. We show that "social theories" of protest participation—those holding that people influence each other in their decisions to protest—have implications that are testable with snapshots of networks like these. If social theories explain protest decisions well, the precise way that protesters are interconnected would be different from the way that similar nonprotesters are interconnected.

We begin by distilling social theories of protest to a simple premise: A person is more likely to attend a protest if she is exposed to closer social contacts who are also likely to attend the protest. We then show that if in the run-up to a protest social contacts positively influence one another to protest in this way, then there are observable implications for the structure of networks in data such as ours. The networks interconnecting protesters should look different from the networks interconnecting similar people who did not protest. In networks measured out two degrees, the differences should manifest not just in the number of ties to one another, but also in the extent of indirect, triadic, and reciprocated ties that are present in each network. Comparing networks serves as a large-scale empirical test: Finding no such differences would cast doubt on social theories of protest.

In fact, we find that all expected differences between the protester and the comparison networks are present, significant, and large in magnitude. Our results are consistent with social theories of protest behavior in which highly motivated individuals influence those most exposed to them socially. Interpreted in the context of Twitter, a person is more likely to attend a protest if she follows people who are motivated to attend, follows people who follow people who are motivated, follows a member of a small clique who is motivated, and follows motivated people who mutually follow the user in return.

This article makes three broad contributions to the study of protest. First, it offers the first characterization of a network among protesters measured beyond egocentric networks, which allows us to capture structural features beyond degree. This network is significantly different from the network among a baseline set of people in the same area with similar interest in the protest issue. Second, it translates existing theory of the role of networks in protest into hypotheses testable with cross-sectional social media data. We are able to move beyond offline studies that show some protesters have friends who protest, and online studies that show online ties relate to online behavior, to show that the full set of Twitter ties among real-world protest participants is interconnected in ways that are consistent with an influence-by-exposure process. Third, our approach of collecting data from online social media, assembling a baseline set for comparison, and utilizing nonparametric statistical methods and tools from social network analysis is novel and easily replicable for other twenty-first-century protests. Our hope is that this article will pave the way for a reinvigorated empirical investigation of social networks' role in individual protest participation.

[1]Although some studies include coarse measures of a social network feature, such as respondents' estimates of the number of their friends who participated (Opp and Gern 1993) or the level of support received from certain tie types (McAdam and Paulsen 1993), the onerous data requirements have severely limited the study of social network *structure* in protest participation with offline data.

[2]Online social media have become standard tools of twenty-first-century protesters (Tufekci and Freelon 2013). In the case of *Charlie Hebdo*, the slogan "Je Suis Charlie" became the worldwide statement of solidarity; that slogan originated on Twitter. By 72 hours after the massacre, the hashtag `#JeSuisCharlie` had appeared in over 5 million tweets (Goldman and Pagliery 2015).

In the next section, we situate our contribution within the theoretical, offline empirical, and online empirical literatures concerning networks and protests. The third section gives an overview of our approach, highlighting steps we take to allay selection concerns. The fourth section presents a theory of protest used to derive hypotheses about network structure from standard assumptions about protest behavior. The subsequent section presents our data, results, and robustness checks, and the last section concludes.

The Role of Network Structure in Protest

Social networks have the potential to affect protest behavior because individuals may influence each other. Links in a social network serve as channels of information and peer pressure. Different network structures—arrangements of links—may be more conducive to this influence, and hence to protests.

Traditional approaches to the study of protest participation stipulate the way that people decide to protest based on others in the network, and then derive the network structures most favorable to protest. One set of theories draws heavily on epidemiological models to represent influence, holding that individuals can "infect" each other with the desire to protest in an analogue to disease spread (Eguiluz and Klemm 2002; Goldenberg, Libai, and Muller 2001; Marwell, Oliver, and Prahl 1988), or via a more complex process that depends on the extent of exposure (Centola 2013; Centola and Macy 2007; Siegel 2009).

Another set of theories accounts for variance in how susceptible people are to others' influence. Whatever its source, this variance can be represented as personal "thresholds" that must be met before people are willing to protest: A person is only willing if enough others in total (Granovetter 1978; Macy 1991) or in her network neighborhood (Chwe 2000; Watts 2002) do so as well.

These approaches share two assumptions about protest decision making. First, in general, social ties to others who are likely to protest increase a person's propensity to protest. Second, the quality of ties matters. Ties must be "wide" enough to transmit social reinforcement (Centola and Macy 2007) or "strong" enough in a Granovetter (1973) sense to convey credible information (Chwe 2000; Siegel 2009). The model we present in the fourth section adapts these two assumptions to the context of Twitter. Despite the rich theory for why network structure should affect protests, large hurdles to collecting the appropriate data have limited empirical tests in both the offline and online domains.

Offline Studies of the Role of Social Ties in Protests

Although precisely measuring networks is a difficult, error-prone, and expensive task in general, doing so in the context of protests is particularly challenging. The biggest hurdles are identifying participants after the fact; tracking them down to survey them; eliciting reliable network information after-the-fact; and learning about enough of their ties, and enough about each of their ties, to answer meaningful questions about the role of network structure in their decision to participate. These problems are compounded when the protest was sensitive or attained notoriety—memory can be colored by *ex post* social judgment (Opp and Gern 1993, 665).

Most empirical work studying networks and protest using offline data documents the existence of ties between a small number of participants and other participants (see Oliver 1984; Snow, Zurcher, and Ekland-Olson 1980). Even the most comprehensive offline network studies of protest have been limited to a small subset of ties without record of interconnections among them for a small number of participants. McAdam and Paulsen (1993) use enrollment lists from the 1964 Mississippi Freedom Sumer Project to target respondents, receiving responses from 340 participants and individuals who withdrew. Questionnaires inquired about the existence of five relationship types and respondents' perceived level of support from each.

Opp and Gern (1993) collected data on social ties among a subset of participants and nonparticipants in the 1989 Leipzig rebellion in the German Democratic Republic. Since there are no enrollment lists in a protest, to identify some of the 70,000 participants in the 1989 demonstration, the authors surveyed 1,300 of the 450,000 residents of Leipzig in late 1990. Respondents were asked whether they participated in the protest and about the existence of ties to colleagues and friends. Questions were of the form "how close were your ties to your colleagues" (on a 4-point scale), "how many of your colleagues criticized the situation in the GDR," and "how many of your colleagues attended peace prayers, demonstrations and similar activities?" (Opp and Gern 1993, 673).

Both studies establish that, among the subset sampled, the existence of a tie recalled after the fact to a protest participant is associated with participation, and that the stronger the tie, the greater the association. Although these are important findings, they leave much unanswered about the role of networks in protest participation. Do those who protest have ties to other protesters because they have more ties in general, or because these ties were used to encourage participation? Do friends of

friends have any effect? If someone belongs to a clique containing a participant, is she more likely to participate? Is the network structure among protesters different from the network structure among similar people who could have protested? Our unusually rich data allow us to tackle these questions, and others like them, for the first time.

The Promising Era of Online Social Media

Online social media offer unprecedented data opportunities for the study of protest behavior. Platforms like Twitter, Facebook, and Instagram are online spaces where individuals create and share content like messages and photos. These platforms are "social" in that users officially establish other users as their contacts; these connections give privileged access to content and are observable to researchers. Online social media are widely used by participants in protests to coordinate, organize, inform, and report (Steinert-Threlkeld et al. 2015; Tufekci and Freelon 2013), and in general these observable online ties can be used as valid measures of offline social contacts (Bisbee and Larson 2017).

Because our *Charlie Hebdo* protest data are from Twitter, they have advantages over self-reported data. First, our sample is more comprehensive. Because we did not use long, costly surveys, we have data on a large set of protesters and nonprotesters. Second, the record of behavior was generated in real time, freeing our data of potentially large recall and social desirability biases. Third, our measure of the social network is rich. The network is observable, behaviorally verified, and complete on Twitter: We know everyone to whom someone is connected, all of their connections, and any connections among them. Finally, and an advantage over many online studies as well, our data contain the precise geo-coordinates of the user at the time of the activity. This offers a measure of real-world behavior, confirming presence at the protest site, again freeing us from reliance on self-reported measures of behavior that can be contaminated by social desirability bias and other confounders.

This article joins a burgeoning research area using online social media to study protests, much of which focuses on the tactics, coordination, and timing of protests (for an overview, see Tufekci and Freelon 2013). Research exploring the role of network *structure* in protest behavior using online social media data is rarer.

The few exceptions focus on the role of network structure in *online* protest behavior. González-Bailón et al. (2011) consider the relationship between network position and Twitter message-sending activity during a monthlong window containing 2011 protests in Spain. Barberá et al. (2015) find that those in peripheral network positions played an important role in the wide reach of messages on Twitter about Istanbul's Taksim Gezi Park protest in 2013 and the United for Global Change demonstration in 2012. González-Bailón and Wang (2016) find that networks that spread messages on Twitter are fragmented, and certain network positions are essential to the wide reach of tweets. These findings, while important, are limited to explaining online behavior: creating and sharing messages on Twitter. Our interest extends beyond who talks about a protest on Twitter to who physically attends.

Steinert-Threlkeld (2017) also relates Twitter activity to offline behavior; the study uses a database of 13 million tweets to argue that peripheral members of a network are more influential in mobilizing a protest than central members are because they provide a more credible signal of the low costs of protest. This study, like ours, uses geocoded tweets to measure real-world participation. However, the network among participants is measured coarsely—for each user, only the number of followers on Twitter is recorded. Since no connections between ties, or ties beyond a single degree, are measured, no information on the structure of the network beyond degree can be explored.

Our unusual data map the networks of protesters and nonprotesters out to two degrees. These networks, combined with a measure of real-world protest participation, allow us to test hypotheses about the precise role of network structure in protest participation.

Overview of Our Data and Approach

Our data are drawn from the universe of Twitter users who sent at least one tweet about *Charlie Hebdo* containing at least one of the following hashtags: `#CharlieHebdo`, `#JeSuisCharlie`, `#Charlie Hebdo`, `#JeSuisAhmed`, `#JeNeSuisPasCharlie`, `#Beinfait`, and `#JeSuisKouachi`. Our analyses use two subsets of these users, which we label "Protesters" and "Comparison Set." Protesters are all users who sent at least one tweet containing at least one of the seven hashtags above during the time of the protest that was geotagged to be within the protest site, Paris's Place de la République. There are 764 such users. Our comparison set is a random sample of 764 of the users who were interested and eligible to participate: They sent at least one tweet containing at least one of the seven hashtags above during the protest (i.e., *Charlie Hebdo* was salient to them) that was geotagged to be in Paris (i.e., they were geographically near enough) but more than

five kilometers away from the protest site.[3] Because these sets only include users who geotag, our analyses rely on the assumption that the proportion of a user's neighbors who geotag is not a function of that user's protest attendance.[4]

We measure the full Twitter network for both the protestors and the comparison set measured out to two degrees. For each user, we collect the usernames of all other users whom she follows (her "ties") and the usernames of all whom these users follow (her "ties of ties").[5] We call this web of following relations constructed for the protesters the "protester network," and the same for the comparison set the "comparison network."

We collect information on protesters and the comparison set simultaneously because we will seek to detect traces of influence by exposure among protesters and verify that those traces do not exist when we look for them in a comparison set of similar users at the same point in time. In order to know what to look for, the next main section specifies a simple social theory of protest that identifies differences in network structure that we should observe between these two networks if people really do influence each other to attend in the way theory stipulates.

Assessing the Role of Networks with Observational Data

Our data on network ties and protest participation are observational. We will show that the way protesters are connected to one another on Twitter is significantly different from the way that eligible nonprotesters are connected to one another, and these differences are strongly consistent with a social theory of people influencing one another to protest. However, as with any observational study, we are limited to showing that the data are consistent with the theory; we cannot be sure that exposure to others in the network *caused* participation.

On the one hand, the fact that the networks under study were not exogenously varied—either by manipulating users' Twitter experiences or by assigning lab participants to new, artificial networks—limits our ability to attribute causality. On the other hand, the fact that the networks we do observe are the real networks used by real protest participants in their natural setting bolsters our study's external validity.

The biggest threat to causality is a selection story: It could be that people who choose to participate in protests also select into networks in which they are relatively well connected to others who choose to participate in the same protests. In such a case, we could not distinguish whether selection into the network or influence by exposure explained the differences we observe.

Although there are limits to addressing this issue with observational data, our strategy is to carefully select the set of nonprotesters in order to hold constant some key sources of selection. If protesters and the comparison set of nonprotesters *both* have attributes that lead people to befriend one another and to participate in the *Charlie Hebdo* protest, then differences in their networks cannot be explained by those common attributes.

For example, one potential source of selection is an interest in politics. If people interested in politics are more likely to befriend others interested in politics, and if those interested in politics are more likely to attend the *Charlie Hebdo* protest, then we might observe highly interconnected protesters simply due to their shared interest. To account for this possibility, we assemble the comparison set from users who used the same seven hashtags about *Charlie Hebdo*, which ensures that all protesters and comparison set users were interested in this political issue. Of course, this still leaves open the possibility that protesters had a *stronger* interest than those in the comparison set. However, we can demonstrate that protesters and those in the comparison set divided their attention across the seven hashtags similarly, and both sets were tweeting about *Charlie Hebdo* at similar rates in the days leading up to the protest (see SI Section 2.6). To the extent that these measures capture the character and intensity of interest in this political issue, the protesters and the comparison set do not differ.

Furthermore, we consider the possibility that, although the protesters and the comparison set appear to be equally interested in *this* political issue, protesters are more politically active in general. However, we show that the protesters and our comparison set follow verified accounts, which include news media and politicians, at

[3]We collect our comparison set from Paris to maximize comparability. Although these users sent a tweet from a location more than five kilometers from the protest during the protest and did not send a tweet from the protest, it is conceivable that some nonetheless traveled to the protest. Hence, some in the comparison set may have actually participated. However, this potential overlap makes the differences we observe conservative. In Supporting Information (SI) Section 2.2, we verify all results with an additional comparison set drawn from France but not Paris.

[4]Geotagging is an optional account setting, and most tweets are not geotagged. SI Section 2.4 reports that protesters geotag 11.6% of tweets, and users in the comparison set geotag 10%. Although the users who geotag could differ from those who do not, our comparisons will all be between users who geotag.

[5]"Following" is the basic social relationship on Twitter. If one user elects to follow another, the home page of the follower will regularly display the other's Twitter activity. These relationships can be asymmetric—user *a* can follow *b* without *b* following *a*.

similar rates (see SI Section 2.6). We also use a latent Dirichlet allocation decomposition to cluster users based on the accounts they follow. If politically active users follow certain political or news accounts more than nonpolitically active users, and if protesters are systematically more politically active than nonprotesters, then we should expect to see protesters cluster separately from nonprotesters. Reassuringly, SI Section 2.5 shows that this is not the case. By these measures, protesters do not appear more politically active than our set of nonprotesters. Consequently, if the politically interested and active select into networks with one another, we should observe a high extent of interconnectedness among both the protesters and the comparison set. Insofar as these measures capture political interest, the differences we observe cannot be due to selection driven by shared political interest.

Finally, we consider sociable personality traits as a source of selection. It could be that some people are more outgoing or gregarious and forge more ties in general. If this type of person is also more likely to protest, we might observe more connections among protesters because they form denser networks among themselves and not because they influenced one another to protest. As the section "*Charlie Hebdo* Protest" shows, the protesters do have more ties and more ties of ties than those in the comparison set. Consequently, in our analyses, we will compare groups based on the proportion of their ties and ties of ties *that they have to one another* rather than the raw number of ties within each group, which will control for this difference. Furthermore, we show that in the days between the massacre and the protest, protesters and comparison set users were tweeting (about anything) at similar rates, and about the protest at similar rates as well (and both were tweeting more frequently, and hence appear more gregarious, than a third set of users in France that was collected without restriction to any particular hashtag use). To the extent that gregariousness is captured by tweet volume and patterns, both groups similarly exhibit this personality trait.

In short, finding that protesters are substantially more interconnected than a purely random set of nonprotesters might be unconvincing because the protesters may be interconnected due to a shared interest in politics or outgoing personalities. But by sampling nonprotesters from those who were equally engaged with the topic of *Charlie Hebdo* in the same place at the same time, we hold constant potentially large sources of homophily and selection. We can be more confident (though still not sure) that differences we observe between the two similar groups' networks are at least in part the by-product of influence by exposure.

Deriving Predictions about Network Structure from Social Theories of Protest

In this section, we show that theories holding that people influence one another in the run-up to a protest have implications for the observed *structure* of networks among those who do protest. We specify a simple, social theory of protest participation in order to pin down expectations for network data collected from protesters and nonprotesters using Twitter.

Our theory distills social theories of protest to a simple premise: Individuals will join a protest if they value doing so highly enough, and their valuation depends in part on their exposure to others' valuations. The greater the extent to which people are exposed to others who value the protest highly, the more likely they are to value the protest highly. The extent of exposure to another's protest valuation depends on how socially proximate the two are—if they are friends, or merely friends of friends, or friends of friends of friends—and on the strength of their relationship—if they are close, or if their relationship is less intimate.

We remain agnostic about why exposure matters in order to capture a wide range of social theories. It could be that a high valuer actively cajoles her contacts to attend. It could be that she passively signals that attending would be met with her approval, or is likely to be worthwhile. It could be that her interest drives her to share information about the protest that convinces her friends. Our approach detects evidence that exposure in a network matters. We leave an examination of how and why for future research.[6]

Our data are drawn from Twitter, a vast network of ties that are indicative of personal relationships and that can function as sources of exposure.[7] On Twitter, users can inform others via messages ("tweets"), secondhand messages ("retweets"), and shared links to web pages. Users' activity also conveys something about the users and the content that they find relevant or endorse. In this way, ties are windows into the opinions, values, and intentions of other users. A tweet that says, "Protest in the Place de la République tomorrow, come if you can!"

[6]SI Section 2.9 presents the results of human coders classifying the content of a random subset of protesters' tweets. Protesters' tweets contain a variety of content, including simple logistical details and exhortations to attend, and contain both explicit and subtle indications of whether a user will attend.

[7]We use Twitter to measure the underlying social network in which exposure can take place. Whether the exposure happens on Twitter or happens in a real-world encounter between the two people who share a Twitter tie is irrelevant to our approach.

not only conveys information about the time and location of a protest, but also suggests to followers that the user endorses the protest and prefers that others attend.

The next section shows that an influence-by-exposure process leading up to a protest results in protesters who are highly connected to one another directly, indirectly, with reciprocated ties, and as parts of cliques. A sample of similar nonprotesters would not be as interconnected in these ways.

Modeling the Decision to Participate in a Protest

Consider a simple model of protest participation among a set of individuals N who are situated in a network g. These individuals base their decision to protest or not in part on the decisions of others. Suppose that attending a protest entails some cost so that a person will only participate if she finds joining the protest sufficiently valuable to offset this cost. Person i's net valuation, V_i, can depend on benefits to herself or to others, and sources of value can vary across individuals; whereas one may value attending the protest highly because one expects its outcome to positively impact the world, another may value attending highly because she expects to win the favor of her friends who care about the protest. So long as, for whatever reasons, i's net valuation is positive, $V_i > 0$, i will attend the protest.

Let exposure to others' valuations depend on two features of network position: distance and strength. Call $d_{i,j}$ the *network distance* between i and j: the length of the shortest path from i to j so that $d_{i,j} = 1$ if i has a tie to j, and, following convention, $d_{i,j} = \infty$ if there is no path from i to j. Call $s_{i,j}$ the *strength* of the path between i and j such that $0 \leq s_{i,j} \leq 1\ \forall i,\ j \in N$ where $s_{i,j} = 0$ if there is no path from i to j, $s_{i,j}$ is the strength of the tie between i and j if $d_{i,j} = 1$, and if $d_{i,j} > 1$, $s_{i,j}$ is the strength of the weakest tie in the path from i to j.[8] Now we can define one's exposure to another in a network:

Definition 1 (Exposure). *The extent to which individual i is exposed to individual j,*

$$E_{i,j}(d_{i,j}, s_{i,j}), \tag{1}$$

is decreasing in network distance $d_{i,j}$ and increasing in path strength $s_{i,j}$.

That is, an individual i is more exposed to j if i and j are closer to each other in the network, and if i and j are connected by a stronger path. Individual i's exposure to j is greatest when i follows j directly ($d_{i,j} = 1$) and when i and j's tie is as strong as possible ($s_{i,j} = 1$). Suppose that $E_{i,j}(d_{i,j} = \infty, s_{i,j}) = 0$ and that $E_{i,j}(d_{i,j}, s_{i,j} = 0) = 0$ so that i has no exposure to j if there is no path from i to j or if the strength of the path from i to j is 0. Further suppose that $E_{i,j}(d_{i,j} = 2, s_{i,j} > 0) > 0$ so that indirect ties generate positive exposure as long as neither tie in the path from i to j has zero strength.

Now we can specify the way a person's net valuation of the protest depends on her exposure to other individuals:

Definition 2 (Protest Valuation). *An individual i's valuation of a protest, V_i, is a function of i's exposure to others and their valuations. Let $H = \{1 \ldots n\}$ be the subset of individuals in N such that $V_h > 0\ \forall\ h \in H$. These are the individuals who value the protest highly. Individual i's valuation is a function of his exposure to these individuals in the network,*

$$V_i(E_{i,j_1} V_{j_1}, E_{i,j_2} V_{j_2}, \ldots, E_{i,j_n} V_{j_n}, Z_i), \tag{2}$$

where $j_1, \ldots, j_n \in H$, and Z_i captures i's private reasons for valuing the protest, independent of the valuations of others.

Recall that when $V_i > 0$, i prefers to participate in the protest. Our key modeling assumption is that V_i is increasing in $E_{i,j} V_j$ for any $j \in H$.[9]

This assumption means that the more that individual i is exposed to an individual j who values protesting highly enough to participate, the higher will be i's own valuation of the protest. In addition, the more an individual j to whom i is exposed values the protest, the higher will i value the protest. Since exposure is increasing in network proximity and tie strength, individuals who have strong, direct ties to others who value protesting highly will value the protest especially highly.[10]

[8]We present the model in its most general form; in principle, long paths can affect protest decisions. However, our ability to test arguments about paths here is limited since our data contain paths of length at most two. Path strength defined as the strength of the path's weakest link means that strong ties can transmit information better than weak ones, and that the consequences of weak-tie transmission are permanent.

[9]Our theory leaves much unspecified about the dynamic process by which individuals form their initial valuation and update over the time period leading up to the protest. Our interest is simply in positing a minimal set of assumptions that capture broad intuitions of existing theory.

[10]It is conceivable that for particularly controversial or high-risk protests, there could also be some trying to influence others *not* to attend. In such a case, so long as there are also some with moderate viewpoints, the comparisons between protesters and others that we derive hold. It is also conceivable that, for more partisan protests, the effect of exposure may vary with the party affiliation of the source. Since *Charlie Hebdo* was not a particularly partisan issue in France, we leave an exploration of this possibility for future research.

Network Hypotheses

If a process with the features laid out above was at play when people were deciding whether or not to protest, what should we expect to observe when we measure a network among a set of individuals who actually participated in a protest?

Our hypotheses make use of some network notation. We use the convention $ij \in g$ to mean that a directed tie from i to j is present in the network g. To refer to everyone to whom i is tied in g, i's "neighborhood," we will write $N_i(g)$. That is,

$$N_i(g) = \{j | ij \in g\}. \tag{3}$$

The neighborhood of i is the set of all other individuals to whom i is tied in g; these are called i's "neighbors." On Twitter, this is the set of everyone whom i follows. Likewise, we can define the set of i's ties of ties accordingly:

$$N_i^2(g) = \{j | d(i, j) = 2\}. \tag{4}$$

$N_i^2(g)$ is the set of all individuals to whom i is connected in a path through the network of length two. On Twitter, this is the set of individuals that those whom i follows follow.

Now we can consider the first of the two inputs to exposure, network distance.

Network Distance Hypotheses. The hypotheses are stated in terms of a set of individuals who are observed to participate in a protest, P, and a set of individuals who are observed to not participate in a protest, C. The network that includes all ties and ties of ties of everyone in P will be called g^P; the network that includes all ties and ties of ties of everyone in C will be called g^C. We present the intuition underlying each hypothesis here; the supporting information contains the formal derivation of each (SI Sections 1.1–1.6).

In our theory, the key assumption is that high exposure to individuals who value the protest highly increases one's valuation of protest, and high enough valuation results in a person actually participating in the protest. When we observe a set of individuals who did participate in a protest, we know that at the conclusion of this process, they must have valued the protest highly enough. Although we cannot know which of the protesters influenced which other of the protesters to attend during the process that led to the protest, if influence by exposure was at play, then protesters should occupy network positions that are highly exposed to one another. Individuals in P should be more exposed to each other in g^P compared to the extent to which individuals in C are exposed to each other in g^C. In other words, protesters should cluster in the network.[11]

Since exposure $E_{i,j}$ is decreasing in $d_{i,j}$, the first hypothesis holds that protesters should cluster with respect to direct ties. We formulate a strong version of this hypothesis that compares the proportion, rather than the number, of ties. On average, a protester's neighborhood should contain a greater share of ties to other protesters when compared to the share of an eligible nonprotester's ties to other sampled nonprotesters. Specifically, we expect

$$\frac{1}{\#P} \sum_{i \in P} \frac{\#\{j | j \in P,\ j \in N_i(g^P)\}}{\#\{j | j \in N_i(g^P)\}} > \frac{1}{\#C} \sum_{i \in C} \frac{\#\{j | j \in C,\ j \in N_i(g^C)\}}{\#\{j | j \in N_i(g^C)\}}, \tag{5}$$

which is our first hypothesis:

H1: On average, the proportion of each protester's ties that are to other protesters in g^P is greater than the proportion of each eligible nonprotester's ties that are to other eligible nonprotesters in g^C (Inequality 5).

Similarly, since $E_{i,j}(d_{i,j} = 2, s_{i,j} > 0) > 0$, individuals are also exposed to others' valuations via indirect connections, their ties of ties. We should also observe a relatively high degree of interconnectedness via ties of ties among the protesters. Specifically,

$$\frac{1}{\#P} \sum_{i \in P} \frac{\#\{j | j \in P,\ j \in N_i^2(g^P)\}}{\#\{j | j \in N_i^2(g^P)\}} > \frac{1}{\#C} \sum_{i \in C} \frac{\#\{j | j \in C,\ j \in N_i^2(g^C)\}}{\#\{j | j \in N_i^2(g^C)\}}. \tag{6}$$

This can be stated as the following hypothesis:

H2: On average, the proportion of each protester's ties of ties to other protesters in g^P is greater than the proportion of each eligible nonprotesters' ties of ties that are to other eligible nonprotesters in g^C.

[11]The derivation of our hypotheses accounts for the way our data were collected. Consider the universe of people eligible to attend the protest. The protesters were selected from this universe *because they participated.* The comparison set was selected at random from the remainder of this universe. Influence by exposure would result in correlated valuations of the protest and hence attendance in the network. Selecting on participation would result in a set of highly interconnected users. Selecting from a large universe of eligible users at random would not. If instead social ties do not influence one another, selecting on participation should look roughy like selecting eligible participants at random from the network.

These hypotheses capture the intuition that if exposure via ties and ties of ties mattered to protest decisions, then individuals who did in fact protest should hold positions in the underlying Twitter network in which they were highly exposed to one another via ties and ties of ties.

Tie Strength Hypotheses. Since exposure via strong ties is more impactful, by the same reasoning as above, we expect protesters to have more strong ties to each other compared to eligible nonprotesters.

Tie strength cannot be directly measured on Twitter. We use two proxies for the attributes such as intimacy and trust thought to underly strength (see Gilbert and Karahalios 2009). Both reduce a potentially continuous measure $s_{i,j}$ to a binary variable taking two values, "strong" and "weak."[12]

In the first operationalization, we consider the arrangement of ties among triads. In particular, given that i follows both j and k, we assume that i's ties to both j and k are stronger if j and k themselves share a tie. The triangle formed by i, j, and k is one measure of a strong relationship between each pair, perhaps indicative of shared attributes or membership in a close-knit social clique (Granovetter 1973). On Twitter, this means that a user's tie to someone she follows is stronger if there is a second person she follows and at least one of the two follows the other.

To simplify notation, let $ijk \in g$ mean $ij \in g$, $ik \in g$, and either $jk \in g$ or $kj \in g$. Then we expect

$$\frac{1}{\#P}\sum_{i\in P}\frac{\#\{ijk \in g^P \mid j \in P \text{ or } k \in P\}}{\#\{ijk \in g^P\}} > \frac{1}{\#C}\sum_{i\in C}\frac{\#\{ijk \in g^C \mid j \in C \text{ or } k \in C\}}{\#\{ijk \in g^C\}}. \quad (7)$$

In words, our next hypothesis is the following:

H3: The average proportion of protesters' triangles that contain at least one other protester is larger than the average proportion of eligible nonprotesters' triangles that contain at least one other of these eligible nonprotesters (Inequality 7).

If triangles comprise a person's strong ties, then the proportion of a person's triangles that entail a protester is a rough measure of the proportion of a person's strong ties that are connections to protesters. We also capture this value a second way, operationalized with reciprocity. A tie is considered strong if it is reciprocated and weak if it is not (see Friedkin 1980). On Twitter, this means that a tie to someone a person follows is strong if that person follows her in return and weak otherwise.

Influence by exposure implies that protesters have more reciprocated ties among one another than eligible nonprotesters have among one another:

$$\frac{1}{\#P}\sum_{i\in P}\#\{j \mid ij \in g^P, ji \in g^P, j \in P\} > \frac{1}{\#C}\sum_{i\in C}\#\{j \mid ij \in g^C, ji \in g^C, j \in C\}. \quad (8)$$

This becomes our final hypothesis:

H4: The number of ties in g^P that are between individuals in P and reciprocated is larger than the number of ties in g^C that are between individuals in C and reciprocated (Inequality 8).

Table 1 Network Attributes for Protesters and Paris Comparison Set

	Protesters	Comparison	*t*-stat
Mean # ties	833 (2,491)	418 (590)	4.5
Mean # ties of ties	134,623 (52,780)	85,830 (49,864)	18.6
Mean # reciprocated	471 (1,989)	113 (416)	4.9
Mean transitivity	0.098 (0.053)	0.108 (0.071)	−3.2

Note: Standard deviations are in parentheses. The *t*-statistic tests the null hypothesis that attributes for protesters and the comparison set are the same. The distributions of the number of ties and transitivity have long right tails; the null can be rejected with high statistical confidence for the log-transformed distributions as well (*t*-statistic 11.6 and 14.9, respectively).

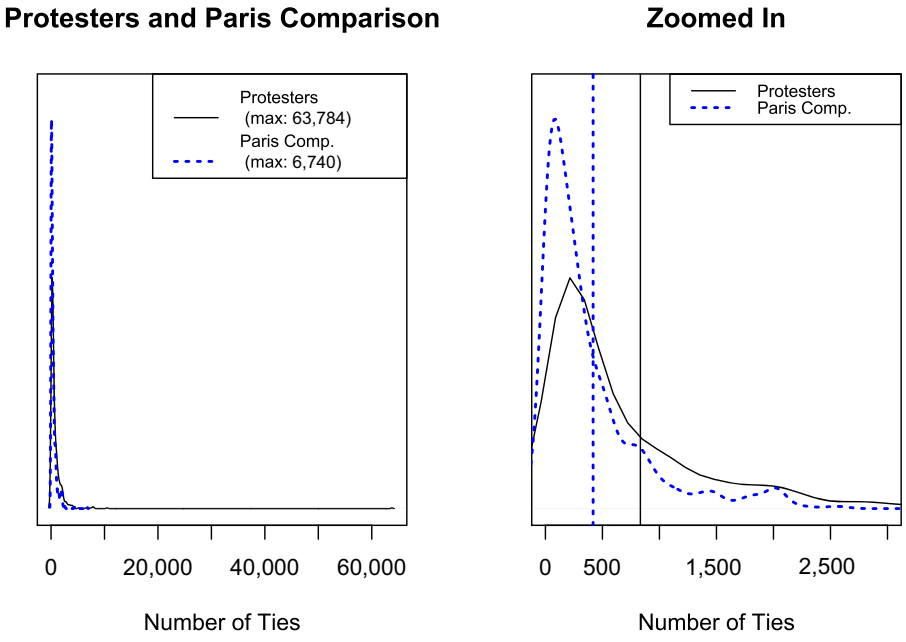
Charlie Hebdo Protest

The *Charlie Hebdo* protest took place on January 11, 2015, in Paris's Place de la République. The protester network we analyze, g^P, centered around the 764 individuals geotagged to be present at the protest and measured out two degrees, contains 93,009,971 distinct nodes. On average, each protester has 833 ties and 134,623 ties of ties. The comparison network, centered around 764 nonparticipants and measured out two degrees, contains a total of 106,116,658 nodes, with 418 ties and 85,830 ties of ties on average. Table 1 compares summary statistics for the two networks.[13]

[12] A binary measure magnifies contrast. We encourage future work, especially using data with paths longer than length two, to consider more categories of strength in order to unpack the role of relationship quality.

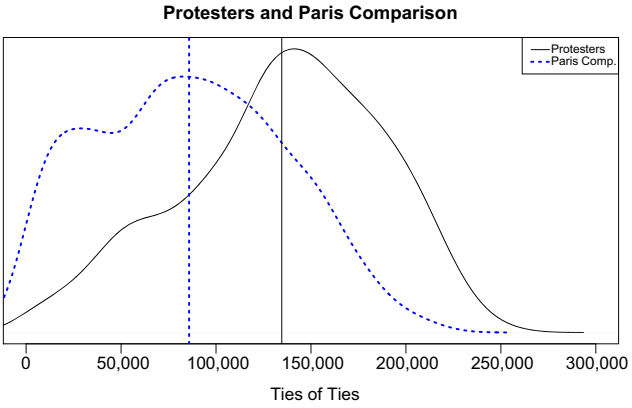
[13] Although there exist other network-level statistics, they require precise measurement of the *absence* of a link between any two nodes

FIGURE 1 Distribution of the Number of Ties per User



Note: The left panel shows the distribution of the number of ties per user in the set of protesters and the Paris comparison set. The right panel shows the same, zoomed in on the mode. Vertical lines indicate the distributions' means.

FIGURE 2 Distribution of the Number of Ties of Ties per User in the Set of Protesters and the Paris Comparison Set



Note: Vertical lines indicate the distributions' means.

Protesters have more ties and ties of ties than individuals in the comparison set. These distributions are displayed in Figures 1 and 2. Note that this will make support for Hypotheses 1 and 2, which pertain to proportions of these quantities, conservative. Protesters also have a larger number of reciprocated ties, and lower transitivity.[14] These characteristics imply that protesters occupy positions of high reach within the Twitter network. Although these individuals would have the greatest exposure to initial high valuers who, according to González-Bailón et al. (2011), tend to be scattered roughly randomly throughout a network, we leave future exploration of the timing and dynamics of influence for future investigation. Now we turn to our hypothesis tests.

Assessing Support for Hypotheses

We present the main results of our hypothesis tests here, with additional analyses and robustness checks in the supporting information.

Support for Network Distance Hypotheses. First we consider the hypotheses that pertain to ties and ties of ties. Hypotheses 1 and 2 hold that since exposure is greatest when ties are direct, and still positive for ties of ties, if protesters influenced one another via these channels,

in the network. An egocentric network measured out two degrees cannot detect links between individuals included as ties of ties of the egos. Hence, statistics like diameter and centrality measures other than degree are not meaningful here.

[14]Network transitivity measures the frequency of triangles—three-person cliques—in the network. The reported value is the average node's ratio of number of three-person cliques to total possible three-person cliques.

TABLE 2 Assessing the Network Distance Hypotheses

	Protesters	Comparison	*t*-statistic	Support
H1: Prop. ties within	0.0040 (0.0063)	0.0004 (0.0017)	15.0 [log: 11.0]	Y
H2: Prop. ties of ties within	0.0332 (0.0332)	0.0057 (0.0080)	22.2 [log: 35.0]	Y

Note: Standard deviations are in parentheses, and the *t*-statistic tests the null that values for protesters and the comparison set are the same. The *t*-statistic on log-transformed data appears in square brackets. Both hypotheses are supported with high statistical confidence.

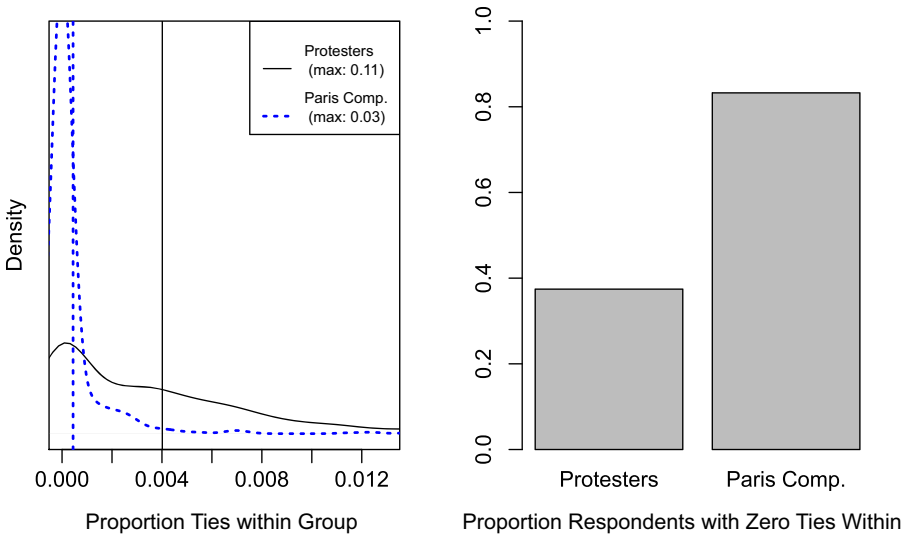
they should be highly interconnected in terms of these features. We expect to observe many ties and ties of ties *connecting protesters to one another*, and this value should be greater for g^P than for g^C.

Table 2 summarizes the results of comparisons between protesters and those in the Paris comparison set. The network among protesters exhibits features consistent with both hypotheses: g^P and g^C differ substantially in the expected direction.

A larger proportion of protesters' ties are to other protesters than the proportion of eligible nonprotesters' ties are to one another in the comparison set (row 1). Protesters also have many more ties of ties to one another than the eligible nonprotesters of the comparison set have to one another (row 2). Both comparisons are highly statistically significant and robust to a correction for the long right tail. The SI (Section 2) shows that the comparisons are also robust to excluding verified Twitter accounts, excluding outliers, and using a different comparison set drawn from a larger geographic area.

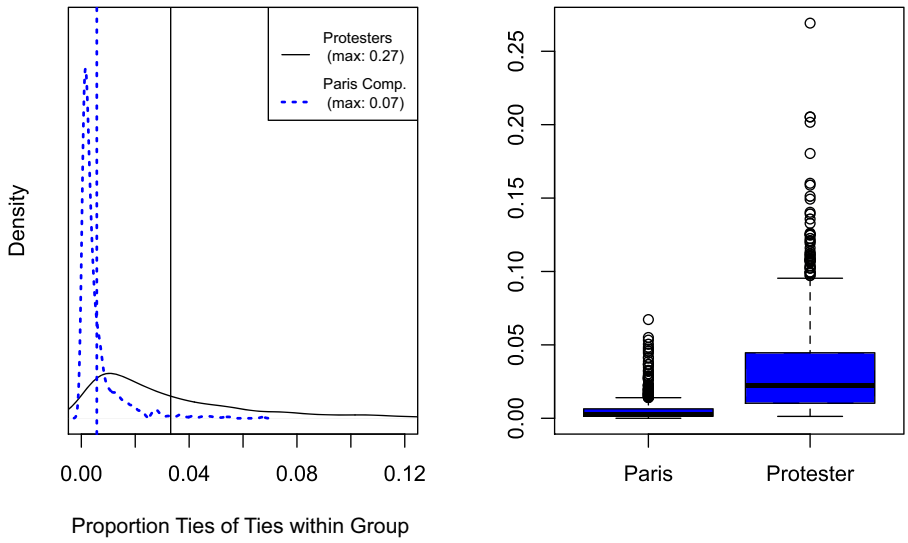
Figure 3 shows the distribution of the proportion of ties within each group, zoomed in to better show the mass of the distribution. The plot on the right highlights just how many more individuals in the comparison set have no ties to others in the comparison set. Over 80% of the eligible nonparticipants in the Paris comparison set have no ties to others in that set, compared to less than 40% of participants who have no ties to other participants. Consistent with Hypothesis 1, those who protest are much more likely to be directly connected to others who protest, even as a share of their

FIGURE 3 Distribution of the Proportion of Ties within Each Group, Zoomed In



Note: The left panel shows the distribution of the proportion of protesters' ties to other protesters and the distribution of the proportion of users' ties in the Paris comparison set to others in the Paris comparison set. Vertical lines show the distributions' means. The right panel shows the proportion of each sample with zero ties to others within the sample. Strong support is indicated for Hypothesis 1.

FIGURE 4 Distribution of the Proportion of Ties of Ties among Each Group, Zoomed In



Note: The left panel shows the distribution of the proportion of protesters' ties of ties to other protesters and the distribution of the proportion of users' ties of ties in the Paris comparison set to others in the Paris comparison set. Vertical lines show the distributions' means. The right panel shows the boxplot for both distributions. Strong support is indicated for Hypothesis 2.

total links (which, recall from the last section, is larger for protesters).

Figure 4 shows the distribution of the proportion of ties of ties among each group, again zoomed in. The plot on the right displays the distribution as a box plot. Even though the protesters have more ties of ties on Twitter overall, a larger proportion of their ties of ties is to other protesters compared to the proportion of eligible nonprotesters' ties of ties that are to one another.

These results are strongly consistent with an influence-by-exposure process on Twitter that functioned as follows: Individuals who followed people who valued the protest highly, and who followed people who followed people who valued the protest highly were influenced to value the protest highly and ultimately attend. Consequently, those who attended the *Charlie Hebdo* protest have a larger share of their ties and ties of ties to other protesters compared to the share of ties and ties of ties that a sample of eligible nonprotesters have to others in that sample.

Support for Tie Strength Hypotheses. Since exposure is theorized to be increasing in tie strength, we expect to observe more strong ties among protesters than among a comparison set of eligible nonprotesters.

Hypothesis 3 regards ties to be stronger when they are "triadic," or part of a triangle—when two of a person's ties themselves share a tie. Hypothesis 4 regards ties to be stronger when they are reciprocated. Table 3 shows that, by both measures, protesters have more strong ties to other protesters than individuals in the comparison set have to others in the comparison set.

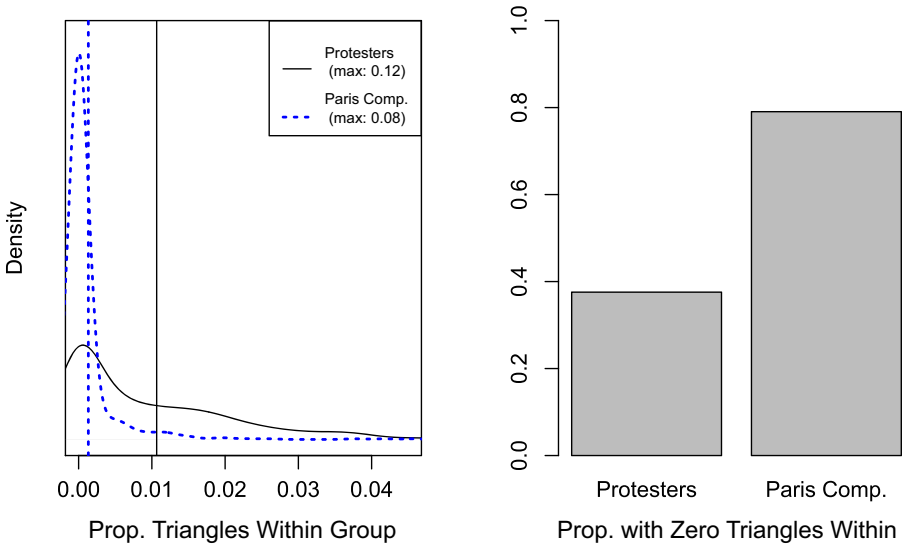
Specifically, the first row of Table 3 shows that a much larger proportion of protesters' triangles in g^P entail another protester compared to the proportion of eligible nonprotesters' triangles in g^C that entail another in this

TABLE 3 Assessing the Tie Strength Hypotheses

	Protesters	Comparison	*t*-statistic	Support
H3: Prop. triangles within	0.0107 (0.0150)	0.0013 (0.0055)	16.1 [log: 8.5]	Y
H4: # Recip. within	3.08 (5.77)	0.18 (0.67)	13.8 [log: 14.0]	Y

Note: Standard deviations are in parentheses, and the *t*-statistic tests the null hypothesis that attributes for protesters and the comparison set are the same. The *t*-statistic on log-transformed data is shown in square brackets. Both hypotheses are supported with high statistical confidence.

FIGURE 5 Distributions of Individuals' Triangles That Entail Another in the Relevant Set



Note: The left panel shows the distribution of the proportion of protesters' triangles that entail at least one other protester, and the distribution of the proportion of individuals in the Paris comparison set's triads that entail at least one other in that set. Vertical lines show the distributions' means. The right panel shows the proportion of respondents with no triangles that entail anyone else in the relevant set. Strong support is indicated for Hypothesis 3.

set. Figure 5 shows the distributions of individuals' triangles that entail another in the relevant set, the set of protesters or the set of eligible nonprotesters. The bar graph on the right shows that over 70% of nodes have no triangles that include another in their reference group for the comparison set, whereas fewer than 40% have none in the protester set.

The second row of Table 3 shows that, on average, protesters have about three reciprocated ties to other protesters, whereas for members of the comparison set, only about 1 out of 5 people have a single reciprocated tie to someone else in the comparison set. Figure 6 shows the distributions of reciprocated ties present in both networks for reference on the left, and the incidence of reciprocated ties among the set of protesters and the Paris comparison set of eligible nonprotesters on the right.

Once again, both hypotheses pertaining to tie strength are strongly supported by our data. These results are consistent with a process by which individuals decide whether to protest based in part on their exposure to others: If a person follows someone who values the protest highly and follows him in return, or follows someone who values the protest highly within a tight-knit clique, that person is more likely to protest. This process would result in many more strong ties among a set of individuals who turned up to protest than among a comparison set of eligible nonprotesters, which is precisely what we see for the *Charlie Hebdo* protest.

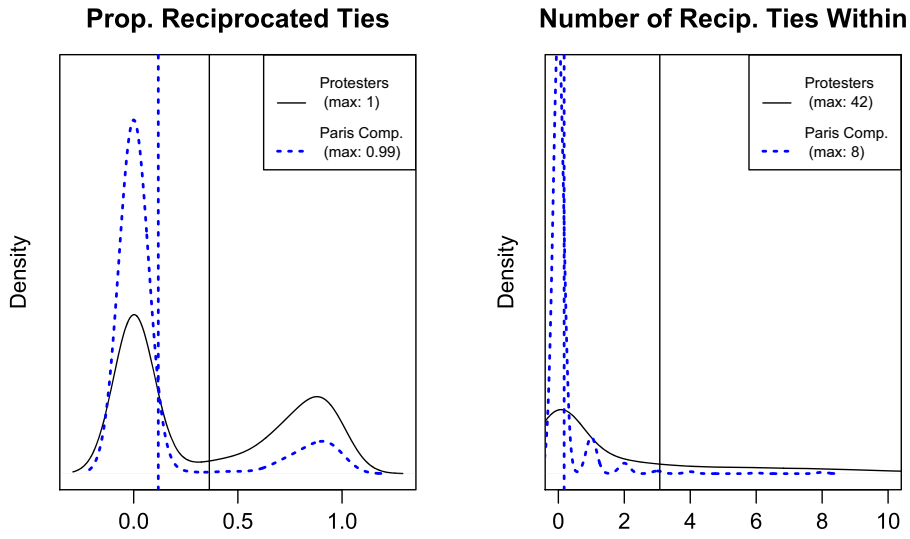
Robustness

In this section, we address potential concerns with our interpretation of the results. Here, we describe the main analyses we perform to alleviate concerns; the details and additional analyses can be found in the supporting information.

The results presented above compare protesters with nonprotesters. We select the nonprotesters from Paris. Twitter networks may differ when users are drawn from vastly different levels of geographic dispersion. Our hope is that Paris is roughly the correct level of resolution for both sets.

We could be wrong in one of two directions. Paris could be too small a pool from which to draw our comparison set, or it could be too large. If drawing from a pool of the incorrect size mechanically introduces differences between our protesters and the comparison set, then the differences we attribute to network influence by exposure may in fact be artifacts of constructing the comparison set. To ensure that our results are not driven by a comparison set drawn from too small a pool, we construct a second comparison set, which we call the "France

FIGURE 6 Reciprocated Ties



Note: The left panel shows the distribution of the proportion of respondents' ties that are reciprocated. The right panel shows the distribution of the number of respondents' reciprocated ties that are to another member in the relevant set—protesters or the Paris comparison set. Vertical lines show the distributions' means. Strong support is indicated for Hypothesis 4.

Comparison Set." We construct this set from users who were geolocated to be in France in the week following the protest irrespective of whether they used one of the seven *Charlie Hebdo* hashtags. We repeat all analyses comparing the protesters to this comparison set, and comparing the two comparison sets to each other (see SI Section 2.2).

The results are reassuring in two ways. First, all of the differences between the protesters and a comparison set hold when the France comparison set is substituted in. Second, in all cases, the Paris and the France comparison sets are much more similar to each other—often statistically indistinguishable—than either is to the set of protesters. This further corroborates our claim that the set of protesters is meaningfully different because of the process that drove individuals to protest, and not because of the geographic size from which it was drawn.

Of course, it is also possible that Paris was too *large* a pool from which to draw our comparison set if, by sampling individuals who tweeted from the Place de la République during the protest, we picked up a set of individuals who were from a very small geographic area, one smaller than the geographic area spanned by individuals drawn from all of Paris. To address this, we construct a measure of the mobility of any Twitter user in our protester set and calculate the same for users in the Paris comparison set. It turns out that protesters are at least as mobile as those in the Paris comparison set: In the lifetime of their Twitter accounts, they have sent tweets from places as least as geographically dispersed as the those in the comparison set.

Furthermore, although over 80 million tweets were geocoded to be in France between January 14, 2015, and September 14, 2015, a window of time after and excluding the protest date, exactly zero were located at the protest site. Hence, those who tweeted from this site during the protest were visiting a small geographic area for the purpose of protesting, but they are unlikely to reside in it.

An additional concern is that some ties on Twitter are to other users who may not be the sources of personal influence to which our theory pertains. These users —celebrities, news sites, and so on—are categorized as "Verified Accounts" on Twitter. To be sure that our results are not artifacts of ties to these accounts, we repeat all of the above analyses with the subset of our data that excludes verified accounts. All results continue to hold (see SI Section 2.3).

We could also worry that something about the geolocation setting is driving the difference between networks among protesters and networks among those in the comparison sets. However, first note that all users in the protest and both comparison sets are included in the sample because they had geolocation activated. Therefore, differences between users who geolocate and users who do not geolocate cannot drive the differences we observe (though knowing more about these differences would help establish to what extent we can generalize from our

findings). For geolocation behavior to be a problem for our results, it would have to be the case that attending not only affects a user's propensity to geolocate, but also that attending *with* a Twitter follower makes *both* users more likely to decide to geolocate compared to their propensity to geolocate when not attending a protest. Although we are not convinced that this behavior is plausible, we further rule it out by measuring the quantity of geolocated tweets by each user in our samples. Overall, the quantity of geolocated tweets is similar across all three of our samples (see SI Section 2.4).

Finally, we could worry that individuals are not actually exposed to the protest valuations or intentions of their social ties. While exposure may take place on Twitter or in real-world interactions, we only have a record of Twitter interactions. As a simple verification check, we randomly selected 200 protesters from our data and collected all tweets they sent between the massacre and the protest. We created a replica of their accounts that their followers would have had access to (SI Figure 27 shows an example) and asked French speakers to classify the content. Approximately 44% of users tweeted specifically about the upcoming protest (as opposed to the issue of *Charlie Hebdo*, which 100% of users in our sample necessarily mentioned with at least a hashtag). Additionally, without being told that all users eventually attended, coders suspected that 64% were at least as likely as even chance to attend. Given that people could have been exposed offline as well, this simple validation reassures us that people had ample opportunity to be exposed to their social contacts' attitudes toward the upcoming *Charlie Hebdo* protest.

Conclusion

Using a novel data set that records both real-world participation in a protest and fine-grained social network information, we offer the first large-scale empirical support for a claim that theory has long hinted should be true: Individuals are influenced by one another in social networks when deciding whether to participate in protests.

Our approach distills existing theory into a simple model of protest participation from which we derive hypotheses about network structure. If exposure to people who value the protest increases a person's propensity to protest, and if exposure is a function of network proximity and tie strength, then the observed social networks among protest participants should look different from the observed social networks among a sample of people who were interested and eligible to participate in the protest but did not.

Indeed, the networks among those who participated in the 2015 *Charlie Hebdo* protest in Paris differ significantly from the networks among individuals in Paris for whom *Charlie Hebdo* was equally salient but who did not participate. The large differences are strongly consistent with soccial theories of influence by exposure: Individuals who participated in the *Charlie Hebdo* protest are more likely than nonparticipants to be connected to one another via direct ties. Moreover, protesters are also more likely to be connected to one another via indirect, reciprocated, and triadic ties, offering support for an even stronger claim that the way ties are arranged—the network *structure*—matters for protest behavior.

Taken together, these results suggest that someone connected by strong, direct ties to people highly motivated to participate in a protest is more likely to participate herself than someone occupying a network position farther away from, or connected with weaker ties to, others intent on protesting. Because our evidence includes verified, real-world participation in a protest and social ties that were observed rather than self-reported, we believe this is the most powerful support to date for the claim that networks play a meaningful role in individuals' protest participation.

Of course, much work remains to understand the process by which an individual comes to protest. With these large network differences established, future work can delve deeper into the presence and possible mechanisms of influence, and the strength of our findings highlights the promise of using Twitter data as one of the tools to do so. Replicating this study for protests that are riskier or more partisan will help determine the scope of our findings, establishing when participation by social contacts is a strategic complement and when it might be a strategic substitute (Cantoni et al. 2017). By measuring Twitter networks over time, tracking the addition of new followers, classifying the content of tweets, and exploiting lifetime Twitter activity and geolocation, researchers can build an even richer picture of the role of networks in protest.

References

Barberá, Pablo, Ning Wang, Richard Bonneau, John T. Jost, Jonathan Nagler, Joshua Tucker, and Sandra González-Bailón. 2015. "The Critical Periphery in the Growth of Social Protests." *PLOS ONE* 10(11): e0143611.

Bisbee, James, and Jennifer M. Larson. 2017. "Testing Social Science Network Theories with Online Network Data: An Evaluation of External Validity." *American Political Science Review* 111(3): 502–21.

Cantoni, Davide, David Y. Yang, Noam Yuchtman, and Y. Jane Zhang. 2017. "Protests as Strategic Games: Experimental Evidence from Hong Kong's Democracy Movement." https://goo.gl/aUSqvo.

Centola, Damon M. 2013. "Homophily, Networks, and Critical Mass: Solving the Start-Up Problem in Large Group Collective Action." *Rationality and Society* 25(1): 3–40.

Centola, Damon, and Michael Macy. 2007. "Complex Contagions and the Weakness of Long Ties." *American Journal of Sociology* 113(3): 702–34.

Chwe, Michael Suk-Young,. 2000. "Communication and Coordination in Social Networks." *Review of Economic Studies* 67(1): 1–16.

Eguiluz, Victor M., and Konstantin Klemm. 2002. "Epidemic Threshold in Structured Scale-Free Networks." *Physical Review Letters* 89(10): 108701.

Friedkin, Noah. 1980. "A Test of Structural Features of Granovetter's Strength of Weak Ties Theory." *Social Networks* 2(4): 411–22.

Gilbert, Eric, and Karrie Karahalios. 2009. "Predicting Tie Strength with Social Media." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ed. Dan R. Olsen Jr., Richard B. Arthur, Ken Kinckley, Meredith R. Morris, Scott Hudson and Saul Greenberg. Boston, MA: ACM, 211–20.

Goldenberg, Jacob, Barak Libai, and Eitan Muller. 2001. "Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-Mouth." *Marketing Letters* 12(3): 211–23.

Goldman, David, and Jose Pagliery. 2015. "#JeSuisCharlie Becomes One of Most Popular Hashtags in Twitter's History." *CNN Money*, January 9. http://money.cnn.com/2015/01/09/technology/social/jesuischarlie-hashtag-twitter/index.html.

González-Bailón, Sandra, Javier Borge-Holthoefer, Alejandro Rivero, and Yamir Moreno. 2011. "The Dynamics of Protest Recruitment Through an Online Network." *Scientific Reports* 1(197): 1–7.

González-Bailón, Sandra, and Ning Wang. 2016. "Networked Discontent: The Anatomy of Protest Campaigns in Social Media." *Social Networks* 44(1): 95–104.

Granovetter, Mark S. 1973. "The Strength of Weak Ties." *American Journal of Sociology* 78(1): 1360–80.

Granovetter, Mark S. 1978. "Threshold Models of Collective Behavior." *American Journal of Sociology* 83(1): 1420–43.

Kim, Hyojoung, and Peter S. Bearman. 1997. "The Structure and Dynamics of Movement Participation." *American Sociological Review* 62(1): 70–93.

Macy, Michael W. 1991. "Chains of Cooperation: Threshold Effects in Collective Action." *American Sociological Review* 56(1): 730–47.

Marwell, Gerald, Pamela E. Oliver, and Ralph Prahl. 1988. "Social Networks and Collective Action: A Theory of the Critical Mass. III." *American Journal of Sociology* 94(3): 502–34.

McAdam, Doug, and Ronnelle Paulsen. 1993. "Specifying the Relationship between Social Ties and Activism." *American Journal of Sociology* 99(3): 640–67.

Oliver, Pamela. 1984. "'If You Don't Do It, Nobody Else Will': Active and Token Contributors to Local Collective Action." *American Sociological Review* 49(5): 601–10.

Opp, Karl-Dieter, and Christiane Gern. 1993. "Dissident Groups, Personal Networks, and Spontaneous Cooperation: The East German Revolution of 1989." *American Sociological Review* 58(5): 659–80.

Siegel, David A. 2009. "Social Networks and Collective Action." *American Journal of Political Science* 53(1): 122–38.

Snow, David A., Louis A. Zurcher Jr., and Sheldon Ekland-Olson. 1980. "Social Networks and Social Movements: A Microstructural Approach to Differential Recruitment." *American Sociological Review* 45(5): 787–801.

Steinert-Threlkeld, Zachary C. 2017. "Spontaneous Collective Action: Peripheral Mobilization during the Arab Spring." *American Political Science Review* 111(2): 379–403.

Steinert-Threlkeld, Zachary C., Delia Mocanu, Alessandro Vespignani, and James Fowler. 2015. "Online Social Networks and Offline Protest." *EPJ Data Science* 4(1): 1–9.

Tufekci, Zeynep, and Deen Freelon. 2013. "Introduction to the Special Issue on New Media and Social Unrest." *American Behavioral Scientist* 57(7): 843–47.

Watts, Duncan J. 2002. "A Simple Model of Global Cascades on Random Networks." *Proceedings of the National Academy of Sciences* 99(9): 5766–71.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Appendix 1: Derivation of Hypotheses
Appendix 1.1: Setup and Working Assumptions
Appendix 1.2: Intro to Proofs
Appendix 1.3: Proof of Density Hypothesis
Appendix 1.4: Proof of Hypothesis 1
Appendix 1.5: Proof of Hypothesis 2
Appendix 1.6: Proof of Hypotheses 3 and 4
Appendix 2: Additional Analyses and Robustness Checks
Appendix 2.1: Sensitivity of Descriptive Statistics
Appendix 2.2: Robustness to France Comparison Group
Appendix 2.3: Robustness to Omitting Verified Accounts
Appendix 2.4: Mobility and Geolocation
Appendix 2.5: Clustering by Accounts Followed
Appendix 2.6: Potential Baseline Differences
Appendix 2.7: Robustness to Following Protesters in Both Sets
Appendix 2.8: Did the Protest Cause the Network Ties?
Appendix 2.9: Direct Evidence of Exposure on Twitter