Communities of Protest*

Jennifer M. Larson†, Jonathan Nagler‡, Jonathan Ronen§, and Joshua A. Tucker¶∥

May 18, 2016

Abstract

Pinning down the role of social ties in the decision to protest has been notoriously elusive, largely due to data limitations. The era of social media and its global use by protesters offers 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 protests in Paris which, unusually, record both real-world protest attendance and high-resolution network structure. We specify a theory of participation in which an individual’s decision depends on her exposure to others’ intentions, and network position determines exposure. Our findings are strong and consistent with this theory, showing that, relative to comparable Twitter users, protesters are significantly more connected to one another via direct, indirect, triadic, and reciprocated ties. These results offer the first large-scale empirical support for the claim that social network structure influences protest decisions.

†NYU, Department of Politics
‡NYU, Department of Politics
§Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine
¶NYU, Department of Politics
∥Larson, Nagler, and Tucker developed the research question and overall design of the study. Larson wrote the original draft of the paper. Ronen wrote the code and conducted the data analysis. All four authors contributed to the revision and editing of the text, as well as the choice of the particular empirical tests conducted. The data was 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.
1 Introduction

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 as a sign of unity and 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 of fundamental importance to the social sciences: why do some decide to join 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; how consequential protests can be given their role in policy change, the overthrow of governments, and violent revolution; and how salient the topic has become due to the world-wide wave of 21st 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 (Marwell, Oliver and Prahl 1988, Kim and Bearman 1997, Chwe 2000, Siegel 2009). However, attempts to gather data relevant to testing this wisdom have run up against two serious barriers. First, traditional methods require finding protest participants after-the-fact and rely on their recall of behavior and ex ante motivations. Second, rich network data requires identifying a full set of ties—protesters’ ties to other protesters, ties to non-protesters, the ties of ties— and the protest behavior, perhaps relayed second-hand, of everyone. Due to these barriers, few empirical studies to date consider the role of social networks in protest decisions and those that do are highly limited.

The era of online social media offers an unprecedented opportunity to obtain large amounts of information about social ties, with the unique feature that for any platform, these can be measured in full, in real time, and unfiltered by the memory of a respondent. We take

---

1 For instance, although some studies include coarse measures of an aspect of social networks, such as respondents’ estimates of the number of their friends who participated in the protest (Opp and Gern 1993) or the level of support they received from certain types of ties (McAdam and Paulsen 1993), the onerous data requirements have precluded the study of social network structure in protest decisions to date.

2 This opportunity is particularly present in the case of protests since online social media has become a standard tool of 21st century protesters (Tufekci and Freelon 2013). The Charlie Hebdo protest studied here is no exception.
advantage of this opportunity and record Twitter use during the Charlie Hebdo protest in Paris, including user geolocation that indicates physical presence or absence at the protest. Our dataset thus contains all 764 individuals who can be geolocated at the Paris protest and a set of 1,000 individuals who were elsewhere in Paris during the protest who serve as our control. Furthermore, we collect full social network information measured out to two degrees—every user each person follows on Twitter, and every user each of these users follow. Our network of protesters thus contains 93,009,981 nodes and our control network contains 106,116,658 nodes.

We specify a theory of protest participation that accounts for the social networks in which prospective participations are situated. In the theory, an individual will attend a protest if she values it highly enough. Her valuation depends in part on how others in her social network value the protest and her exposure to them. If she is heavily exposed to individuals who value the protest highly, she is more likely to value the protest highly and hence to ultimately participate. Her exposure to any other individual depends on their relative positions in the social network. Exposure to an individual is greater when that individual is closer in the network and is connected by stronger ties. If individuals are influenced by others in their network in this way, then the network observed among those who ultimately attend a protest should look different than the network observed among a control set. The differences should manifest in terms of the extent of direct, indirect, triadic, and reciprocated ties that are present among the protesters and among the control set.

In fact, using statistical methods from social network analysis and community detection, we find that all expected differences between the protester network and the control network are present, significant, and large in magnitude. Our data are consistent with an influence-by-exposure process of protest behavior: highly motivated individuals influence those most exposed to them on Twitter.

Using these unusual data, we contribute a new case to the study of protest behavior that

---

The slogan “Je Suis Charlie” became the worldwide statement of solidarity with the victims and support of non-violent expressions of free speech. That slogan originated on Twitter. The day after the massacre, the hashtag expressing it, “#JeSuisCharlie,” appeared in 6,500 Tweets per minute; by 72 hours after the massacre, the hashtag had appeared in over 5 million Tweets (Goldman and Pagliery 9 January 2015).
is the first to include high-resolution social network information. This article makes three broad contributions to the study of protests. First, we offer empirical support for the claim that social networks influence protest outcomes, a claim that has reached “consensus” status without complete testing (Kim and Bearman [1997] p. 70). Second, we advance evidence for an even stronger claim, that the structure of the full network in which participants and non-participants are embedded matters for decisions to protest. Although decades of research on networks suggests this should be true in the case of protests, the data to confirm this have not previously been available (Siegel [2009], p. 123). Third, our approach which collects data from online social media and utilizes methods derived from social network analysis and community detection is easily replicable for other 21st century protests. Our hope is that this article will pave the way for a reinvigorated empirical investigation of individual protest decisions, one that includes full networks among both protesters and non-protesters.

2 Measuring the Role of Social Ties in Protests

2.1 The Role of Network Structure

The social networks in which people are embedded could impact their decision to protest for many different reasons. In the process we envision (and make precise in Section 4), an individual’s decision to protest depends in part on her exposure to others and their feelings toward the protest. Exposure to someone who values the protest highly makes a person more likely to value the protest highly, all else equal. The network matters because it determines the extent of exposure. An individual is more exposed to another if both are closer together in the network and if they are connected by a tie that is stronger in the Granovetter [1973] sense—based on greater intimacy or trust.

In a Twitter network, individuals can Tweet about an upcoming protest, or share the Tweets of others by reTweeting, all the while filtering content and editorializing. In this way, users can both learn information about the protest, and can learn how others feel about the protest. Exposure to this information may influence one’s own valuation of the protest
(and ultimately one’s willingness to attend the protest) for a number of reasons. Exposure may reveal information that allows a person to form her private valuation of the protest and attend. Exposure may entail learning that a close friend is eager to protest, which may convince someone that if protesting is right for this close friend, then protesting is right for herself. Perhaps learning that a number of social contacts value the protest highly establishes that someone could be subject to social shame if she refrains.

If a process like the one described here was at play in individuals’ decisions to protest, then the networks observed among protesters and among a control set will have significantly different characteristics. However, and importantly for our data collection strategy, these differences can only be detected with fine-grained network data. If we would collect only ties to other protesters, or only a subset of each users’ ties, or only the number but not the identity of their ties, we would mask the differences that speak to support for the theory. For this reason, our approach which records all ties and ties-of-ties on Twitter has considerable advantages over existing approaches which, in the offline world, require measuring small subsets of ties at best, and in the online world to date, record coarser measures of networks to reduce storage size.

2.2 The Limitations of Offline Data

Networks research suggests that behavior at the individual level, and the aggregate of that behavior into a group outcome like a protest, likely depends on the full network structure. However, measuring full networks is a difficult task in general, and particularly difficult in the context of protests.

Most empirical work studying networks and protest measures a subset of social ties of a subset of participants. This approach has a long history in Sociology. Early work focused on personalized recruitment to movements, and showed that those recruited tend to have relationships with existing participants in movements [Bibby and Brinkerhoff 1974, Harrison 1974, Oliver 1984]. Taking stock of the empirical research to that point, Snow, Zurcher Jr and Ekland-Olson (1980) point out that much is impressionistic, though the data that do exist (with small samples ranging from 31 to 310) are consistent with the interpretation that social
networks are an important factor in recruitment.

McAdam and Paulsen (1993) offer a similar characterization of empirical work on the role of social ties, noting that “the result is a growing body of studies that appear to attest to the causal importance of organizational ties [...] or prior contact with a movement participant as strong predictors of individual activism. But while they remain important, these studies are nonetheless plagued by a troubling theoretical and empirical imprecision that raises important questions about their ultimate utility” (McAdam and Paulsen 1993 p. 640). In a step toward greater precision, the authors used the enrollment lists of the 1964 Mississippi Freedom Summer Project to send questionnaires after the program to participants and enrollees who withdrew, receiving responses from 340 individuals. The questionnaires gathered more detailed social network information than those of previous studies, asking about not just one but five types of relationships—parents, friends, civil rights organizations, other volunteers, and religions groups. Respondents were asked to report the level of support offered from each. This design allowed the authors to conclude that having a prior strong tie linking the applicant to another volunteer predicted participation well, and that support from other types of ties, like parents, also had an effect on participation.

Relatedly, Opp and Gern (1993) collect data on social ties among participants and non-participants in the 1989 Leipzig rebellion in the German Democratic Republic. The authors argue that preexisting social networks helped the spontaneous protests emerge in a repressive regime that did not allow free association: "Networks of friends, colleagues, or neighbors constitute micro-contexts for mobilizing citizens. Members of these networks can communicate relatively easily and argue and exchange rewards that promote participation in political action" (Opp and Gern 1993 p. 662). To identify some of the 70,000 participants in the demonstration in the fall of 1989, the authors surveyed 1,300 of the 450,000 residents of Leipzig in November and December of 1990. Respondents were asked if they participated in the protest, and asked questions about ties to colleagues and friends. Questions were of the form “how close were your ties to your colleagues” with a 4-point scale of very weak to very strong, “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 p.
Regression analysis shows that having friends critical of the GDR is positively related to mobilization.

The empirical literature offers suggestive support for the conventional wisdom that “social networks are of central importance in explaining social movements and political protest” (Opp and Gern, 1993, p. 659). However, despite the innovative designs, the data limitations—many explicitly acknowledged in this literature—limit the scope of support. The number of respondents tends to be small, in part because surveys are costly, and in part because identifying participants can be difficult. Surveys are conducted after, sometimes long after, decisions to participate are made. This can be particularly problematic when the protest was sensitive or attained notoriety—memory can be colored by after-the-fact social judgement (Opp and Gern, 1993, p. 665).

The most severe data constraint is that social network information is measured by survey questions asking respondents to assess the influence that a certain type of tie, or at best a few types of ties, played in their decisions. This method provides some network information, but tends to miss ties of other types and ties to non-participants, and most significantly, precludes a measure of the full network structure—how those ties are related to each other, ties of those ties, and so on.

These limitations are inherent to a methodology that collects information via surveys after-the-fact. An alternate approach in the literature, both in sociology as well as in economics and political science, eschews data: the approach begins with the premise that social ties matter for recruitment, and designs game-theoretic models or simulations to explore ways that networks matter. Marwell, Oliver and Prahl (1988, p. 502) begins “It is widely agreed that participants in social movement organizations are usually recruited through preexisting social ties...But exactly how and why social ties are important is less well established.” The author uses simulations to show that given this agreement, density of ties among actors in a group improve its chances for collective action. Relatedly, Chwe (2000) explores the minimally connected network configurations among actors playing a strategic threshold game that result in full-scale protests among the actors, and Siegel (2009) presents a model of agents with varied motivations to participate who adjust desires to participate over time in response to their social
ties in local networks to explore the consequences of network classes and the addition of certain kinds of ties.

While these theoretical approaches are indisputably valuable, our aim is to use a new data opportunity to revive the empirical exploration of motivations to protest. Our hope is that by contributing to knowledge about the empirical role of social networks in protest decisions, we enrich the future theoretical investigation of protest as well.

2.3 The Promising Era of Online Social Media

Online social media offer unprecedented data opportunities for the study of protest behavior. Platforms like Tumblr, Facebook, Twitter and Instagram have become increasingly popular online spaces where individuals create and share content like photos and messages. These platforms have a social networking component whereby users officially establish other users as their social contacts; these connections give privileged access to content. On Twitter, the main social connection is the “following” relationship where one user elects to follow another. By following her, a user sees the Tweets posted by the other user on her homepage.

Not only are social media platforms increasingly popular, each with millions of users worldwide and growing, but these platforms are also increasingly used by participants in protest (Tufecki and Wilson 2012; Tufecki and Freelon 2013). Protesters organize, coordinate, persuade, inform, and report on social media (Pollock 2011). This use generates a trail of data unlike previous sources of data on protest in a number of ways.

First, data are available for all participants and non-participants who use social media. Researchers are not limited to a small subsample or constrained by the costs of long surveys—in principle they have access to all on social media.

Second, and crucially for the study of potentially sensitive or socially significant events, the data are generated in real time. Messages sent, connections between participants, online behavior, is all recorded as it happens. Researchers can access this information without surveying participants after the fact, avoiding potentially large recall or social desirability biases.

Third, and in large departure from previous methods, social media data contain rich social network information. For the first time, researchers have access to a measure of social con-
nections that is complete in the social media domain: the researcher can access everyone to whom a participant is connected, everyone to whom they are connected, connections among them, and even has access to some behavioral measures of the character and strength of these ties. Because the network information is observable, these measures avoid the difficulties of eliciting networks from surveys.

Finally, many social media platforms offer the option for precise geolocation. If users elect to use the extra feature, all activity on social media is accompanied by the precise geocoordinates of the user at the time of the activity. Since an increasing amount of social media activity is generated by users with smartphones on the go, this can be particularly helpful for identifying protest participants, as we show here (and others have made use of as well (see Steinert-Threlkeld et al., 2015)).

Studying protest with social media data is an active research area with many promising insights uncovered about the tactics, methods, informational role and timing of protests (Segerberg and Bennett, 2011; Starbird and Palen, 2012; Tufekci and Freelon, 2013; Valenzuela, 2013; Earl et al., 2013; Rød and Weidmann, 2015; Steinert-Threlkeld et al., 2015; Munger et al., 2015). However, to date, little work makes use of the social network information that can speak to earlier, foundational lines of research in the social sciences.

An exception is Steinert-Threlkeld (2015), which uses a database of over thirteen 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. The network is measured by the number of followers of each user in the dataset. This coarse measure of the network allows a measure of in-degree for each user and no other statistic. Since no connections between ties, or ties beyond a single degree are measured, no further information on the structure of the network can be considered.

This coarsening is in keeping with others studying networks among social media users to date, and follows from the size requirements of network information. Although online social media allows access to full network information, storing this information for a large number of users can create datasets that exceed researchers’ physical storage capacity. Collecting full network information out n degrees from a set of users generates a set of nodes that grows...
exponentially in n. The network measured out 2 degrees from 1,762 users contains 129,665,566 nodes.

Nonetheless, this big data holds the key to answering questions that have previously eluded study. The dataset of Tweets that we construct has four unusual features that allow us to speak to the role of social networks in protest decisions. First, thanks to the geolocation service activated for some Twitter users, we have a measure of real world behavior to combine with the online data.

Second, we have comparable groups of protest participants and a randomly selected set of other Twitter users who are similar to the protest participants. This allows us to make use of methods from community detection and social network analysis to explore network features among participants at the group level with reference to a control group.

Third, our dataset includes unusually rich social network information, allowing us to map the network of who the protesters and non-protesters follow – their social ties– and who these individuals follow– their ties of ties. The data storage demands of a network this large have precluded others from recording this information, but without it, the network structure could not be explored.

Fourth, our dataset is large and records information from real time as opposed to retrospective surveys. While this feature is common among datasets culled from online social media, it is unusual when accompanied by a measure of offline behavior and rich social network information.

3 Measuring the Role of Networks in Protest

3.1 Data

Our data are a subset of Twitter users who tweeted about Charlie Hebdo. We collected all Tweets that contained hashtags #CharlieHebdo, #JeSuisCharlie, #Charlie Hebo, #JeSuisAhmed, #JeNeSuisPasCharlie, #Beinfait, and #JeSuisKouachi. We assemble two subsets of the users who sent these Tweets: Protesters, and a Control Set. All users who sent at least one
Tweet during the time of the protest geotagged to be within the Place de la République are included in our Protesters data: there are 764 such users. A random sample of 1,000 drawn from all users who sent at least one of these Tweets during the protest geotagged to be in Paris but not within the area of the protest are included in our Control Set.

For the Protesters and the Control Set, we measure their full Twitter network out to two degrees. That is, for each user, we collect the usernames of all other users whom she follows (her “ties”) as well as the user names of all users whom these users follow (her “ties-of-ties”). This web of following relations constructed for the protesters is called the “protester network,” and the same for the control set is the “control network.”

Our approach specifies a simple theory of protest behavior in which individuals base their decision of whether or not to protest in part on how others value participating in the protest. We use this theory to derive observable implications for networks measured during a protest. We compare the network connecting known participants in the Charlie Hebdo protest to a network connecting individuals who were in Paris and for whom the Charlie Hebdo incident was salient, but who were geotagged to be away from the protest.

Our theory generates testable hypotheses at the network-level for expected differences between the protester network and the control network. Since our approach measures differences in observed networks that result from different underlying behavior, it is related to commu-

---

3These users were chosen conditional on being at least five kilometers away from the Place de la République. Users have the option to activate geotagging in their settings. Although the users who geotag could differ from those who do not, our comparisons will all be between users who geotag. The supporting information contains additional analyses of user location after the protest, as well as a third set of Twitter users that we call the “France Control.” This third set are geolocated to be within France the week following the protest, but never Tweeted using any of the above hashtags.

4The basic social relationship on Twitter is the following relationship in which one user elects to “follow” another, which results in the home page of the follower regularly displaying the Twitter activity of the user who is followed. These relationships can be symmetric, as when two users each follow each other, or they can be asymmetric as when one user follows the other but the other does not follow the first in return. Users are part of a vast directed network of following relationships on Twitter.

5We collect our control set of users from Paris to maximize similarity between our protest participants and a randomly selected subset of the network which serves as our control. It is conceivable that, although these Twitter users in Paris did not send any Tweets from the protest, they sent Tweets while geographically away from the protest site and then attended the protest without Tweeting. For this reason, calling the control group “non-participants” is not necessarily accurate. However, this potential overlap between participants and individuals in our control should attenuate differences between the two sets; any differences we do observe are then even more conservative estimates of differences between the set of attenders and a randomly selected set of Twitter users. Furthermore, we collected a second control set from France but not Paris. We repeat all analyses presented below using the France control as the comparison set and all results hold. These can be found in the online appendix.
nity detection methods for complex networks (Fortunato 2010; Lancichinetti and Fortunato 2009). These methods also strive to detect differences in sets of nodes theorized to result from differences in underlying processes, and have been used in the sciences to identify clusters of shared functionality, like functional modules in metabolic networks (Palla et al. 2005) and common specific within-cell functions in protein-protein interaction networks (Chen and Yuan 2006), and in the social sciences to identify small communities among large groups of people, like social cliques among high schoolers in a friendship network (Moody and White 2003) and intellectual communities within a collaboration network (Girvan and Newman 2002). In the standard approach, these methods use algorithms to search an observed network for the best guess of boundaries around likely subcommunities that comprise the network. While the specific techniques vary, the approaches extract from an observed network evidence of communities within it.

The analogous approach here would be to observe the full Twitter network and ask whether algorithms detect our protesters as a relatively cohesive community within it. However, since there are over 300 million active Twitter users, describing the social network requires accounting for the over $9 \times 10^{16}$ possible relationships, an unwieldy dataset. Instead, our data include a subset of the Twitter network— the 764 users whose geolocation places them at the Charlie Hebdo protest. The simplest formulation of our question asks whether these 764 form a relatively cohesive set of users in the Twitter network, or whether they are just a scattered, disconnected set of users. To answer this question without measuring the whole vast Twitter network, we collect another subset of similar size from the Twitter network. This control subset is what a scattered, disconnected set of users in the Twitter network look like. Our analog to community detection asks to what extent the attenders look like a cohesive community and relate in patterned ways implied by our theory of protest relative to a randomly selected set of users from roughly the same geographic concentration at the same time who activate the same account feature.

Naive storage of the presence or absence of one directed relationship per byte would require over 83 million terabytes to store the Twitter user network at any given point in time.
4 A Network Theory of Protest

In this section, we present a theory of protest participation that accounts for the role of Twitter. First we lay out a set of assumptions for the way individuals decide whether to protest and how they may influence each other’s decision. Next we derive expectations about individual-level protest behavior. Finally, we use these expectations to formulate hypotheses about the networks among Twitter users that should be observed at the time of protest.

Our theory holds that individuals will join a protest if they value do so highly enough, and their valuation depends in part on their exposure to the valuation of others. The more individuals are exposed to people who value the protest highly, the more likely they are to value the protest highly. The extent to which an individual is exposed to another’s protest valuation depends on how socially proximate the two are– are they friends, or merely friends of friends, or friends of friends of friends– and on the strength of their relationship– are they close, or is their relationship less intimate.

We conceptualize Twitter as a vast network of ties– “following” relationships between users– that serve as sources of exposure. Users who encounter content of another user can learn information based on messages the user creates, links to web pages the user recommends, and messages created by other users that the user retweets. Importantly, the messages, links, and retweets shared by a user also convey something about the type of content that the user finds relevant or endorses.

Ties in the Twitter network therefore not only represent channels of information, but also windows into the opinions, values, and even intentions of other users. A Tweet that says “Protest in the Place de la République tomorrow, come if you can!” not only conveys information about the time and location of a protest, it suggests to the follower that the sender endorses the protest and prefers that others attend.

The extent of a Twitter user’s exposure to another’s information depends on how that user is situated relative to that other user in the Twitter network. We hold that user $i$ is more exposed to user $j$ if $i$ follows $j$ than if $i$ follows someone who follows $j$. Furthermore, we hold that user $i$ is more exposed to user $j$ than user $k$ if $i$ and $j$ have a stronger, more meaningful
relationship than $i$ and $k$ do. If $i$ and $j$ have been close friends for years, trust each other, and spend a lot of time together, then we assume that $i$ will be more receptive to $j$ than to a near stranger $k$. This could be because the information transmitted is more credible when sent from a stronger source, or because consequences of normative judgement are greater when ties are deeper (McAdam and Paulsen [1993]).

4.1 Deciding Whether to Participate in a Protest

Consider a simple model of protest participation among a set of individuals $N = \{1 \ldots n\}$ who are situated in a network $g$. These individuals base their decision of whether or not to protest in part on the decisions of others. Suppose that attending a protest incurs some cost $c$ so that a person will only participate if she finds joining the protest sufficiently valuable to offset this cost. Person $i$’s valuation $V_i$ can depend on benefits to herself or to others, and sources of value can vary across individuals; while one may value attending the protest highly because she expects its outcome to positively impact the world, another may value attending it highly because she expects to win the favor of her friends who care about the protest. So long as, for whatever reasons, $V_i > c$, $i$ attends the protest.

Suppose that an individual $i$ can be exposed to the valuation of others.

**Definition 1 (Exposure):** Let $d_{i,j}$ be the length of the shortest path from $i$ to $j$ so that $d_{i,j} = 1$ if $i$ and $j$ share a tie, $d_{i,j} = \infty$ if there is no path from $i$ to $j$. Let $s_{i,j}$ be the strength of the tie between $i$ and $j$ where $s_{i,j} = 0$ if $i$ and $j$ share no tie, and $0 \leq s_{i,j} \leq 1 \forall i \neq j$. The extent to which individual $i$ is exposed to individual $j$,

$$ E_{i,j}(d_{i,j}, s_{i,j}), $$

(1)

is decreasing in network distance $d_{i,j}$ and increasing in tie 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 $j$ share a stronger tie. $i$’s exposure to $j$ is greatest when $i$ directly follows $j$ and when $i$ and $j$’s relationship is based on trust and respect. Suppose that $\lim_{d_{i,j} \to \infty} E_{i,j} = 0$ and that $E_{i,j}(d_{i,j} = 2) > 0$ so that indirect ties generate positive exposure.
An individual’s valuation of the protest is then based in part on her exposure to other individuals.

**Definition 2 (Protest Valuation):** An individual $i$’s valuation of a protest is differentiable function

$$V_i(E_{i,j_1}, E_{i,j_2}, \ldots, E_{i,j_{n-1}}, Z_i)$$

for $j_1, \ldots, j_{n-1} \in N \setminus i$, where $Z_i$ is the aggregate of $i$’s private reasons for valuing the protest independent of the valuations of others.

Recall that when $V_i > c$, $i$ prefers to participate in protest. Our key modeling assumption is that:

For $j \ni V_j > c$, 

$$\frac{\partial V_i}{\partial E_{i,j}} > 0.$$ 

That is, 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 short, individuals who are more exposed to others who value protesting highly will value the protest more highly. Those who value protesting highly enough will participate. 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.

Consider an example of this process at play. Take two individuals, Mary and Sue, who have both heard there will be a protest the next day. Suppose neither is initially very interested. Mary logs onto Twitter and sees that her closest friends is trying to encourage others to attend, her most esteemed work colleague is sharing logistical details, and family members are retweeting other family members’ calls to action. Sue logs onto Twitter and finds that one acquaintance has unenthusiastically retweeted a stranger’s call to action. We claim that Mary is more likely to participate in the protest than Sue. Perhaps Mary wants her family and
colleague to think well of her, perhaps she trusts her closest friend’s assessment of activities they would both find worthwhile, or perhaps she is simply convinced by the information about the cause that all are sharing. Regardless of exactly why, Mary, due to greater exposure to high valuations of the protest on Twitter, is more likely to eventually attend than Sue.

This intuition generates a number of testable hypotheses about what to expect when we measure a network among those who did participate in a protest and compare it to a comparable network among a control set. We turn to these in the next section.

4.2 Hypotheses

Our interest is in deriving the observable consequences for the networks among those who ultimately protest. Given that some 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? How should this network differ from a network measured among a control set of individuals?

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 participated in a protest, then, we know that they must have ultimately valued the protest highly enough. Although we cannot know which protesters were influenced by which other protesters and how exactly the process leading up to the protest unfolded over time, the theory implies that those who protested should overall be connected to each other in the Twitter network in distinct, observable ways.

For each of our hypotheses, we first state the comparative static that the model implies in terms of an individual’s position relative to some other individual who values the protest highly. Then we derive the observable implication for the network among those who participate in a protest. All rely on the logic that high exposure to another person who values the protest highly increases one’s valuation of the protest, and that exposure depends on relative network positions.

Our hypotheses make use of some network notation. First, we will use the convention
Let \( ij \in g \) mean that a tie from \( i \) to \( j \) is present in the network \( g \). Second, to refer to everyone to whom \( i \) is tied in \( g \), we will use the convention \( N_i(g) \), or sometimes simply \( N_i \) when the network is obvious, to represent \( i \)’s network neighbors. That is,

\[
N_i(g) = \{ j | ij \in g \}.
\]

(2)

The neighborhood of \( i \) is the set of all other individuals to whom \( i \) is tied in \( g \). In the case of Twitter, this is the set of everyone whom \( i \) follows. Likewise, we can define \( i \)’s ties-of-ties accordingly:

\[
N^2_i(g) = \{ j | d(i, j) = 2 \}.
\]

(3)

\( N^2_i(g) \), or sometimes simply \( N^2_i \), 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.

### 4.2.1 Network Distance

Direct ties to individuals who value protesting highly increase one’s valuation of the protest (since \( E_{i,j} \) is decreasing in \( d_{i,j} \)). This implies two consequences for the network observed among those who attend a protest compared to the network observed among a control set.

First, we should expect to observe greater “density” in the network among \( P \) than in the network among \( C \). Specifically, we expect

\[
\frac{1}{\# P} \sum_{i \in P} \frac{\# \{ j | j \in N_i, j \in P \}}{\# P} > \frac{1}{\# C} \sum_{i \in C} \frac{\# \{ j | j \in N_i, j \in C \}}{\# C}.
\]

This expectation stipulates that the average proportion of each individual’s ties to others in the relevant set relative to the total possible number of ties to others in that set should be higher in the Protest set than in the Control set. This can be stated as our first hypothesis:
Hypothesis 1: The density of the network among protesters is greater than the density of the network among randomly chosen controls.

Second, we should expect to observe higher proportions of each users’ ties to others in the network among $P$ than in the network among $C$. Specifically, we expect

$$\frac{1}{\#P} \sum_{i \in P} \frac{\#\{j| j \in N_i, j \in P\}}{\#\{j| j \in N_i\}} > \frac{1}{\#C} \sum_{i \in C} \frac{\#\{j| j \in N_i, j \in C\}}{\#\{j \in N_i\}}.$$  

This leads to our next hypothesis:

Hypothesis 2: The average proportion of links to other protesters among protesters is greater than the average proportion of links to other controls for controls.

Moreover, individuals are also exposed to others’ valuations via indirect connections, their ties-of-ties. These indirect ties may influence a decision to join a protest by serving as the ultimate sources of direct ties’ influence—perhaps the direct ties value the protest so highly because their other ties did. These indirect ties may also impact a person’s perception of how widely a protest is supported. If a user gathers from her ties that there are others, perhaps strangers to the user, who also support the protest, then the user may feel more compelled to do what so many others find to be the right thing to do.

Similarly to the case of direct ties, in this case, we expect to observe a larger proportion of ties-of-ties among protesters than among controls:

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

This can be stated as the following hypothesis:

Hypothesis 3: The average proportion of ties-of-ties to other protesters among protesters is higher than the average proportion of ties-of-ties to other controls for controls.
4.2.2 Tie Strength

Operationalizing the network distance measures in the last section was straightforward: shorter paths between users admit greater exposure. Tie strength measures are less straightforward to operationalize. Ideally we could measure the strength of relationship—its intimacy, its trust, its duration. In our data, we use a proxy for these attributes that captures strength in two different ways.

In the first, we consider the arrangement of larger sets of ties. In particular, if \(i\) 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. In such a case, \(i, j\) and \(k\) form a “triad,” thought to be indicative of especially strong relationships between each pair \((\text{Granovetter} 1973)\). The tie connecting \(j\) and \(k\) could be indicative of \(i\) spending lots of time with each which brought them together, or of the three sharing much in common, or of a close-knit group of friends.

Taking ties entailed in triads to be stronger than ties not entailed in triads, we expect more ties among protesters to be entailed in triads than ties among the control. Specifically, in the triads present in the network, a greater proportion of these should contain two protesters than compared to the proportion of triads in the control network that should contain two members of the control set. Define the set of triads present in a network \(g^A\) for group \(A\) to be \(\text{TRIAD}^{g^A} = \{\{i, j, k]\mid i \in A, j \in A, k \in A, j \in N_i, k \in N_j,\text{ and either } i \in N_k \text{ or } k \in N_i\}\). Then we expect

\[
\frac{\#\{l \in \text{TRIAD}^{g^P}\} \text{ for } i, j, k \in l, \{i, j\} \subset P \text{ or } \{j, k\} \subset P \text{ or } \{i, k\} \subset P}{\#\text{TRIAD}^{g^P}} > \frac{\#\{l \in \text{TRIAD}^{g^C}\} \text{ for } i, j, k \in l, \{i, j\} \subset C \text{ or } \{j, k\} \subset C \text{ or } \{i, k\} \subset C}{\#\text{TRIAD}^{g^C}}.
\]

Our next hypothesis is then:

**Hypothesis 4:** The average proportion of protesters’ triads that contain at least one other protester is larger than the average proportion of controls’ triads that contain at least one other
Finally, we consider a second operationalization of tie strength. While strong ties have a range of substantive interpretations from greater intimacy to relationships of greater duration (see Gilbert and Karahalios, 2009), a standard method of operationalizing tie strength is via reciprocated directed ties (see Friedkin, 1980). On Twitter, a social tie between two individuals can be regarded as strong if both individuals follow each other. If one follows the other without reciprocation, the tie can be regarded as weak.

Call a tie strong when it is reciprocated: i and j share a strong tie when \( i \in N_j \) and \( j \in N_i \). If i follows j on Twitter and j follows i, that tie is considered strong. If i follows j but j does not follow i, call that relationship weak. Then, in networks among protesters and controls, we should expect to observe a greater number of reciprocated ties among protesters than among controls:

\[
\frac{1}{\#P} \sum_{i \in P} \# \{ j \in P | j \in N_i, i \in N_j \} > \frac{1}{\#C} \sum_{i \in C} \# \{ j \in C | j \in N_i, i \in N_j \}.
\]

This generates our final hypothesis:

**Hypothesis 5:** The number of ties among protesters that are reciprocated is larger than the number of ties among controls that are reciprocated.

## 5 Charlie Hebdo Protests

### 5.1 Descriptive Network Statistics

Before assessing support for the hypotheses derived in the last section, we present descriptive statistics of the protester network and the control network. Our data contain the network for the 764 protesters collected out to two degrees. For each protester, we record all individuals that she follows, and all individuals that those followed follow. We refer to a following relation
This network contains not only links among protesters, but also all other links that protesters have to non-protesters, links among these, and their other links.

The resulting protester network contains a total of 93,009,971 nodes. On average, each protester has 833 ties and 134,622 ties of ties. We report the same quantities for the control set of randomly-selected other Twitter users for comparison. Table 1 contains the summary statistics for both the protesters and the control set.

<table>
<thead>
<tr>
<th></th>
<th>Protesters</th>
<th>Control</th>
<th>T-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Degree</td>
<td>833 (2491)</td>
<td>456 (1067)</td>
<td>4.30</td>
</tr>
<tr>
<td>Mean 2-degree</td>
<td>134,622 (52779)</td>
<td>86,349 (49710)</td>
<td>19.67</td>
</tr>
<tr>
<td>Mean Transitivity</td>
<td>0.098 (0.053)</td>
<td>0.109 (0.072)</td>
<td>-3.43</td>
</tr>
</tbody>
</table>

Table 1: Node-level network attributes for protesters and control set with standard deviation reported in parentheses. The degree and transitivity distributions have a long right tail; the differences in the log transformed distributions are statistically significant as well (t-statistic 11.71 and 15.37, respectively).

Users on Twitter follow a large number of other users, who follow a large number of other users. Figure 1 shows the distribution of the number of ties for the protesters and the Paris control set. Both contain long right tails; the figure on the right zooms in on the left mass of the distribution. Figure 2 shows the distribution of ties of ties for protesters and the control set.

Figures 1 and 2 make clear that Twitter users who attended the Charlie Hebdo protest have significantly more ties and more ties of ties on Twitter than a randomly selected control group. In the appendix, we confirm that the difference between the two distributions is not driven by outliers; the median number of ties and the mean excluding the top twenty values is also larger.

On the one hand, these comparisons suggest that protesters occupy Twitter network positions with greater reach— they follow more people who themselves follow more people. This

---

7We prefer “tie” to “friend” because following relations on Twitter are a superset of many different social relationships. Some follow others on Twitter because of friendship, some because they are colleagues, some because they are relatives, and others for many different reasons. Previous studies were limited to one or a handful of categories, asking respondents specifically about “friends” or “coworkers.” We want to be explicit about the distinction: we capture a myriad of social relationships—all those salient enough to warrant an interaction on Twitter—by measuring links with Twitter following, and so prefer the general catch-all term “tie.”

8Likewise, the comparisons are statistically significantly different even on the logged distributions. The difference is also not due to a difference in sample size or peculiarities of Paris control set, discussed further below.
Figure 1: Distribution of the number of ties per user in the set of protesters and the Paris control set. The left plot shows the raw density for both; the right zooms in on the mass of the distribution. Vertical lines plot the distributions’ means.

Figure 2: Distribution of the number of ties of ties per user in the set of protesters and the Paris control set. Vertical lines plot the distributions’ means.

may give them improved access to content on average. On the other hand, these comparisons indicate that our hypothesis tests in the next section will be conservative. Our hypotheses pertain to the extent to which protesters are connected to other protesters and the controls are connected to other controls. We report differences in terms of the proportion of users’ ties.
Any differences we report are despite protesters having a larger diameter on average.

5.2 Assessing Support for Hypotheses

5.2.1 Support for Network Distance Hypotheses

First we consider the hypotheses that pertain to network distance; those about ties and ties-of-ties. Since the theory stipulates that exposure to others’ ideas about a protest is greatest when ties are direct, and still positive for ties-of-ties, we expect to observe many ties and ties-of-ties connecting protesters, especially compared to a control set.

Table 2 reports the comparisons between protesters and those in the control set that assess support for each of these hypotheses. In all three cases, the network among protesters is consistent with the hypotheses and significantly different than the network among the control set.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Protesters</th>
<th>Control</th>
<th>T-stat</th>
<th>Support</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1 Density</td>
<td>0.00437 (0.00749)</td>
<td>0.00033 (0.00101)</td>
<td>16.84 [log: 13.15]</td>
<td>Y</td>
</tr>
<tr>
<td>H2 Proportion within</td>
<td>0.00400 (0.00633)</td>
<td>0.00057 (0.00210)</td>
<td>15.91 [log: 11.95]</td>
<td>Y</td>
</tr>
<tr>
<td>H3 Proportion 2-ties</td>
<td>0.03260 (0.03275)</td>
<td>0.00706 (0.01020)</td>
<td>23.12 [log: 32.84]</td>
<td>Y</td>
</tr>
</tbody>
</table>

Table 2: Assessing the network distance hypotheses. Standard deviations in parentheses, t-statistic on log-transformed data in square brackets. All three hypotheses are supported with high statistical confidence.

Protesters have many more ties to other protesters than the control set has to other members of the control, measured both as the proportion of total possible ties to others in that set (row 1) and the proportion of all realized ties (row 2). Protesters also have many more ties-of-ties to protesters than members of the control set have to other members of that set (row 3). All three comparisons are highly significantly significant, and robust to a correction for the long right tail.

Figure 3 shows the distribution of the proportion of ties within the relevant group, zoomed in to better show the mass of the distribution. The figure on the right highlights just how many more individuals in the control set have no ties to others in the control set compared with protesters who have no ties to other protesters. As expected, those who protest are much
more likely to be directly connected to others who protest, even as a share of their total links (which, recall from the last section, is greater for protesters).

Figure 3: The plot on the left shows the distribution of the proportion of protesters’ ties to other protesters and the distribution of the proportion of users’ ties in the Paris control set to others in the Paris control set. Vertical lines show the distributions’ means. The plot on the right shows the proportion of each sample with zero ties to others within the sample. Protesters have a much higher proportion of ties to other protesters than individuals in the control set do to others in the Paris control set, and many more protesters have any ties to other protesters than individuals in the control set do to others in the control set.

Figure 4 shows similar plots for the proportion of ties-of-ties among protesters compared to those among the control set. Once again, the plot on the left zooms into the mass of the sampling distributions and shows that on average, protesters have many more ties-of-ties to other protesters than members of the control set have to others in the control set. The right shows a box plot to further highlight the differences. Even though the protesters have more ties-of-ties on Twitter in general, a larger proportion of their ties-of-ties are to other protesters than the proportion of controls’ ties to others in the control.

5.2.2 Support for Tie Strength Hypotheses

Next we consider the hypotheses that pertain to tie strength. Since exposure is theorized to be increasing in the strength of ties, we expect to observe more strong ties among protesters than among a control set.
Figure 4: The plot on the left shows the distribution of the proportion of protesters’ ties’ ties to other protesters and the distribution of the proportion of users’ ties’ ties in the Paris control set to others in the Paris control set. Vertical lines show the distributions’ means. The plot on the right shows a boxplot for both distributions. Protesters have a much higher proportion of ties’ ties to other protesters than individuals in the control set do to others in the Paris control set.

Hypotheses 4 and 5 rely on different operationalizations of tie strength. Hypothesis 4 regards ties to be stronger when they are entailed in a triad. Hypothesis 5 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 control set have to others in the control set.

<table>
<thead>
<tr>
<th></th>
<th>PROTESTERS</th>
<th>CONTROL</th>
<th>T-STAT</th>
<th>SUPPORT</th>
</tr>
</thead>
<tbody>
<tr>
<td>H4 Triads within</td>
<td>0.01062 (0.01496)</td>
<td>0.00167 (0.00659)</td>
<td>16.53 [log: 10.55]</td>
<td>Y</td>
</tr>
<tr>
<td>H5 Num Recip within</td>
<td>3.06 (5.77)</td>
<td>0.25 (0.97)</td>
<td>15.09 [log: 10.11]</td>
<td>Y</td>
</tr>
</tbody>
</table>

Table 3: Assessing the tie strength hypotheses. Standard deviations in parentheses, t-statistic on log-transformed data in square brackets. Both hypotheses are supported with high statistical confidence.

Specifically, the first row of Table 3 shows that a much larger proportion of triads present in the protesters network entail at least two protesters than the proportion of triads in the control network that entail at least two control individuals. Figure 5 shows the distributions of triads entailing at least two protesters or controls for the protesters and the Paris control group, respectively, as well as the incidence of no triads that entail anyone in the relevant set.
The control set contains a much larger proportion of triads that entail no one else from the control set.

Figure 5: The plot on the left shows the distribution of the proportion of protesters’ triads that entail at least one other protester, and the distribution of the proportion of Paris control’s triads that entail at least one other in the Paris control set. Vertical lines show the distributions’ means. The plot on the right shows the proportion of respondents with no triads that entail anyone else in the relevant set.

The second row of Table 3 shows that on average, protesters have about three reciprocated ties to other protesters while members of the control have only about a quarter of a reciprocated tie to others in the control set (or, more sensibly, about one out of four members of the control set have a single reciprocated tie to someone in the control set). Figure 6 shows the distributions of reciprocated ties, and the incidence of reciprocated ties among the set of protesters and the Paris control set.

Once again, both hypotheses pertaining to tie strength are strongly supported by our data. These analyses are consistent with a process by which individuals decide whether to protest based in part on their exposure to others who decide to protest. This process would result in many more strong ties among a set of individuals who turned up to protest than among a control set, which is precisely what we see.
Figure 6: The plot on the left shows the distribution of the proportion of respondents’ ties that are strong (reciprocated). The plot on the right shows the distribution of the number of respondents’ strong ties that are to another member in the relevant set—protesters or the Paris control set. Vertical lines show the distributions’ means in both.

5.3 Robustness

In this section, we address potential concerns that could confound our interpretation of the results. Here we describe the analyses we perform to alleviate concerns, and the details are presented in the appendix.

The results presented above make use of the geocoded Charlie Hebdo protesters’ full network data collected out to two degrees, and compare this information with a control set randomly selected from those geolocated in Paris but not at the protest. This control set is chosen to maximize similarity between the controls and those who attended the protest in order to establish the hard case for rejecting the null hypotheses above. Because they are all in the same city on the same day at the same time that the protesters are in that city, we are able to compare the subnetwork among protesters with the subnetwork among a random set drawn from the same geographic concentration.

One potential concern is that the differences in network features detected between the protesters and the Paris control set are artifacts of the protest site’s smaller geographic size than Paris. The concern would be that the network among those in this smaller geographic
unit is naturally characterized by more ties within the set. To address this concern, we take two steps.

First, we construct a measure of the mobility of any Twitter user in our protester set and calculate the same for users in the Paris control. It turns out that protesters are at least as mobile as those in the Paris control, which is to say that in the lifetime of their Twitter accounts, they have sent Tweets from places as least as geographically dispersed as the control. This suggests that the users who sent Tweets from the protest site are not residents of the protest site, but perhaps visited from a region at least as large as Paris.

Second, although over 80 million tweets were sent that were geolocated to be in France from January 14th, 2015 to September 14th, 2015, exactly zero were located in the protest site. Hence, those who tweeted from this site during the protest were visiting this small geographic space, but not from this small geographic space.

A second concern could speculate about the opposite: that the Paris control set is too small, and since the protesters may have traveled from a much farther distance, perhaps something about the small control set is driving misleading results. To alleviate this concern, we repeat all analyses using a larger control set, this time drawn at random from France. The appendix shows all the analyses repeated with the France control. These analyses offer two useful insights. First, all of the comparisons between the protesters and the Paris control set hold when instead the France control set is substituted in. Second, in all cases, the Paris and the France control sets are much more similar to each other—often statistically indistinguishable—than either is to the protest set. 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 they were drawn.

A third potential problem could arise if individuals in one of the two sets had a disproportionately large set of ties to “Verified Accounts,” users like news sites and celebrities, which made one of the sets more eligible to have some network features than the other. To address this concern, we repeat all of the above analyses with when Verified Accounts are excluded from the data. Once again, all of the results continue to hold, and these can be viewed in the online appendix.
Finally, we could worry that something about the geolocation service is driving the difference between networks among protesters and networks among the control sets. First, note that all users in the protest and both control 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 external validity). For geolocation behavior to be a problem for our results, it would have to be that attending a protest which other ties on Twitter attend makes sets of individuals activate geolocation together, while those who attend but are not ties on Twitter do not activate geolocation and while those who do not attend the protest activate geolocation at random. While we are not convinced that this behavior is plausible, we further rule it out by measuring the extent of geolocated tweets that each user in ours samples sends. By and large, geolocation is not a feature that is toggled with high frequency (those who geolocated in our sample tend to geolocate in general), and the extent of geolocated tweets is similar across all of our samples.

Hence, the network differences we observe between the protesters and the control set are robust.

6 Conclusion

Ties in social networks have the potential to transmit information about protests and about individuals’ feelings toward a protest. If decisions to join a protest are in part interdependent so that exposure to others who are strongly in favor of protesting influences a person to be more likely to protest, then the social networks constructed from the ties and ties-of-ties of a set of participants should look different than the social networks of a similar control set.

Indeed, the social networks among those who participated in the Charlie Hebdo protests look significantly different than the social networks of similar individuals from Paris, and the large differences are in the expected direction. Specifically, and consistent with an exposure theory of protest participation, the network constructed from the protesters exhibits substantially more direct, indirect, triadic, and reciprocated ties among the protesters than the
network constructed from the controls.

These results are consistent with a theory of protest participation in which individuals influence each other’s protest decisions such that the influence is greatest when social ties are strong and individuals are close to each other in the network. This is the first large-scale empirical support for this process at play. Our hope is that it will be the first of many.

## 7 Appendix

### 7.1 Verifying Descriptive Statistics

Protesters have more ties on average on Twitter than the Paris and the France control sets. Figure 7 shows the distribution of the number of friends for all three sets of Twitter users, and the same when the twenty largest values are excluded. Although the means compress slightly, protesters still have substantially more ties than either of the control groups.

![Figure 7: Distribution of the number of ties per user in the set of protesters, the Paris control set and the France control set. The left plot shows the raw density for all three; the right shows the densities for each, excluding the twenty largest values. Vertical lines plot the distributions’ means.](image-url)

Protesters and Paris Control

Exclude Top 20

Protesters (max: 63784 )
Paris Control (max: 28187 )
France Control (max: 69087 )

Protesters (max: 3222 )
Paris Control (max: 2054 )
France Control (max: 2382 )
Table 4 reports different summary statistics of the number of ties for users among protesters and the two control sets. The maximum number of ties in the protester set is substantially larger than the maximum in the Paris set, but not in the France set. However, these large values are more of an anomaly to the France control set than to the set of protesters: only 4.6% of the France control set have over 2,000 ties, compared to 8% of the protesters. The median value of ties, and the mean excluding the top twenty values, are greater for protesters.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Max</th>
<th>Mean - top 20</th>
<th>Prop &gt; 2k</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protesters</td>
<td>833</td>
<td>410</td>
<td>63784</td>
<td>638</td>
<td>0.082</td>
</tr>
<tr>
<td>Paris</td>
<td>456</td>
<td>228</td>
<td>28187</td>
<td>376</td>
<td>0.034</td>
</tr>
<tr>
<td>France</td>
<td>613</td>
<td>301</td>
<td>69087</td>
<td>464</td>
<td>0.046</td>
</tr>
</tbody>
</table>

Table 4: Comparisons of the set of protesters, the Paris control set and the France control set. Mean - top 20 is the maximum value reported, excluding the 20 largest. Prop > 2k is the proportion of the sample with at least 2,000 ties.

Finally, we worry that since the size of the control sets are larger (each with about 1,000 users), errors in measuring the underlying population may be correlated with size. We use the control samples to generate a sampling distribution of the average number of ties for samples of size 764 (the number in the protester set) to assess the likelihood of observing a mean as large as the protester set mean when samples are smaller. Figure 8 shows the results of 10,000 simulated samples of size 764 from both the Paris and France controls and the location of the observed protester mean relative to them. In both sets, the probability of observing a value as high as our observed value is highly statistically unlikely. The difference is unlikely to be driven by a difference in sample size.

Figure 9 shows the same distribution for the number of ties of ties for all three groups. Once again, excluding the twenty largest values compresses the means slightly (displayed as vertical lines), but protesters have substantially more ties of ties on average (though the maximum value in the France control set is larger).

Table 5 repeats the comparisons of Table 4 for ties of ties. Protesters in general have more ties of ties, both on average, and at the median. The comparison holds when the twenty largest values are excluded. As with the number of ties, the France control contains a higher
maximum value of ties of ties, and here the proportion in the right tail is larger for the France control. in other words, the extreme values of ties of ties are not greater among protesters, but they typical values– the mean and median– are greater among protesters.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>Max</th>
<th>No Top 20</th>
<th>Prop &gt; 200k</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protesters</td>
<td>134622</td>
<td>139944</td>
<td>255674</td>
<td>132006</td>
<td>0.102</td>
</tr>
<tr>
<td>Paris</td>
<td>86349</td>
<td>86430</td>
<td>218745</td>
<td>84075</td>
<td>0.008</td>
</tr>
<tr>
<td>France</td>
<td>116289</td>
<td>108558</td>
<td>323461</td>
<td>112802</td>
<td>0.135</td>
</tr>
</tbody>
</table>

Table 5: Comparisons of the set of protesters, the Paris control set and the France control set in terms of ties of ties. Mean - top 20 is the maximum value reported, excluding the 20 largest. Prop > 200k is the proportion of the sample with at least 200,000 ties of ties.

Figure 8 shows the distribution of transitivity for protesters compared to the same for the Paris and France control sets. Transitivity is computed as the ratio of weakly closed triangles to which a node is incident to the total number of possible triangles to which a node could be incident.
7.2 Robustness to France Control Group

Table 6 replicates Table 2 using a control group collected from Twitter users geo-located to be in France during the Charlie Hebdo protests and not at the protests in Paris. All results hold when using the control group from France instead of the control group from Paris.

<table>
<thead>
<tr>
<th></th>
<th>Protesters</th>
<th>FRANCE Control</th>
<th>T-stat</th>
<th>Reject</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH1 Density</td>
<td>0.00437 (0.00749)</td>
<td>0.00040 (0.00103)</td>
<td>16.55 (log: 14.65)</td>
<td>Y</td>
</tr>
<tr>
<td>NH2 Ratio within</td>
<td>0.00405 (0.00673)</td>
<td>0.00085 (0.00211)</td>
<td>14.07 (log: 8.86)</td>
<td>Y</td>
</tr>
<tr>
<td>NH3 Proportion within</td>
<td>0.00400 (0.00633)</td>
<td>0.00084 (0.00209)</td>
<td>14.61 (log: 8.86)</td>
<td>Y</td>
</tr>
<tr>
<td>NH4 Proportion 2-ties</td>
<td>0.03260 (0.03275)</td>
<td>0.00493 (0.01048)</td>
<td>24.90 (log: 38.76)</td>
<td>Y</td>
</tr>
<tr>
<td>NH5 Triads within</td>
<td>0.01062 (0.01496)</td>
<td>0.00207 (0.00632)</td>
<td>16.12 (log: 10.34)</td>
<td>Y</td>
</tr>
</tbody>
</table>

Table 6: Rejecting the null hypotheses that the network was irrelevant to the decision to protest.

Figure 11 shows the distribution of the proportion of ties within each of the groups for
Figure 10: Distribution of transitivity for protesters, Paris control set and France control set. Vertical lines are means.

Protesters are a much more cohesive group: a much larger proportion of each protester’s ties are to other protesters compared to the proportion of each member of the control set’s ties that are to other members of each control set.

Figure 11: Proportion of each user’s ties that are to other members of the relevant set—protesters to protesters, Paris control set to Paris control set, France control set to France control set.
Protesters have many more ties within the group than either of the two control groups do, as Figure 14 shows.

![Graph showing proportions of respondents with 0 ties within each group.](image)

Figure 12: Proportion of respondents in each set—Protesters, Paris Control, France Control—that have no ties to any other user in that set.

The same comparisons hold for the proportion of ties of ties within each set, regardless of the control set. Figure 13 shows the distribution of the proportion of ties’ ties within each group both in full (left) and zoomed in (right).

![Graph showing distribution of proportion of ties’ ties within each group.](image)

Figure 13: Proportion of each user’s ties’ ties that are to other members of the relevant set—protesters to protesters, Paris control set to Paris control set, France control set to France control set.

Finally, the proportion of triads in the protest set containing other protesters is larger than
Figure 14: Proportion of respondents in each set—Protesters, Paris Control, France control—that have no ties of ties to any other user in that set. The value is 0 for Protesters and the Paris Control set, and 1.3% for the France Control set.

The proportion of triads in either of the control sets containing other individuals in the control sets. Figure 15 shows the distributions.

Figure 15: Proportion of respondents’ triads in each set—Protesters, Paris Control, France control—that entail at least one other user in that set.

Table 7 replicates table 3 with the France control group instead of the Paris control group. Once again, all comparisons hold.

While many users in all sets have no strong ties, a much higher proportion of users in the control sets have no strong ties, shown in Figure 16.
<table>
<thead>
<tr>
<th></th>
<th>Protesters</th>
<th>FRANCE Control</th>
<th>T-stat</th>
<th>Reject</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Number Strong Ties</td>
<td>471 (1989)</td>
<td>209 (2232)</td>
<td>2.56 (log: 4.71)</td>
<td>Y</td>
</tr>
<tr>
<td>Mean Proportion Strong</td>
<td>0.36 (0.41)</td>
<td>0.15 (0.33)</td>
<td>11.89 (log: 2.63)</td>
<td>Y</td>
</tr>
<tr>
<td>Mean Number Strong w/i Group</td>
<td>3.06 (5.77)</td>
<td>0.32 (0.98)</td>
<td>14.70 (log: 12.30)</td>
<td>Y</td>
</tr>
</tbody>
</table>

Table 7: Reciprocated ties in general and among protesters and France controls.

Figure 16: Proportion of respondents’ strong ties in each set—Protesters, Paris Control, France control.

### 7.3 Robustness to Omitting Verified Accounts

Tables 8 and 9 replicates Tables 2 and 3, omitting all verified Twitter accounts. All comparisons are robust to excluding verified accounts.

<table>
<thead>
<tr>
<th></th>
<th>Protesters</th>
<th>Control</th>
<th>T-stat</th>
<th>Reject</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH1 Density</td>
<td>0.00397 (0.00714)</td>
<td>0.00033 (0.00101)</td>
<td>15.88</td>
<td>Y</td>
</tr>
<tr>
<td>NH2 Ratio within</td>
<td>0.00445 (0.00763)</td>
<td>0.00069 (0.00250)</td>
<td>14.47</td>
<td>Y</td>
</tr>
<tr>
<td>NH3 Proportion within</td>
<td>0.00437 (0.00722)</td>
<td>0.00068 (0.00245)</td>
<td>14.95</td>
<td>Y</td>
</tr>
<tr>
<td>NH4 Proportion 2-ties</td>
<td>0.02626 (0.02761)</td>
<td>0.00607 (0.01084)</td>
<td>20.94</td>
<td>Y</td>
</tr>
<tr>
<td>NH5 Triads within</td>
<td>0.01259 (0.02014)</td>
<td>0.00204 (0.00758)</td>
<td>14.81</td>
<td>Y</td>
</tr>
</tbody>
</table>

Table 8: Rejecting the null hypotheses that the network was irrelevant to the decision to protest with verified accounts omitted.

Figures 17, 18, 19 and 20 replicate the results of the main article with verified accounts removed.
Figure 17: Proportion of ties among protesters and controls with verified accounts omitted.

Figure 18: Proportion of ties’ ties among protesters and controls with verified accounts omitted.

<table>
<thead>
<tr>
<th></th>
<th>Protesters</th>
<th>Control</th>
<th>T-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Number Strong Ties</td>
<td>322 (650.72)</td>
<td>96 (362.91)</td>
<td>9.23</td>
</tr>
<tr>
<td>Mean Proportion Strong</td>
<td>0.35 (0.40)</td>
<td>0.11 (0.29)</td>
<td>14.35</td>
</tr>
<tr>
<td>Mean Number Strong w/i Group</td>
<td>2.80 (5.38)</td>
<td>0.25 (0.97)</td>
<td>14.62</td>
</tr>
</tbody>
</table>

Table 9: Reciprocated ties in general and among protesters and controls with verified accounts omitted.
Figure 19: Proportion of triads among protesters and controls that entail others in the group, with verified accounts omitted.

Figure 20: Proportion of strong ties, and incidence of strong ties among protesters and controls, with verified accounts omitted.
References


Steinert-Threlkeld, Zachary C. 2015. “Spontaneous Collective Action: Peripheral Mobilization During the Arab Spring.” *working paper*.


