Collective Action in Networks: Evidence from the Chilean Student Movement*

Felipe González†

Hundreds of thousands of students skipped school during the 2011 student movement in Chile to protest and reform educational institutions. Using administrative data on millions of students' daily school attendance on protest days, this paper presents robust evidence of school absenteeism following a threshold model of collective behavior. Students skipped school on a protest day only when more than 40% of the members of their networks also skipped school. Importantly, even though I show skipping school imposed significant educational costs on students, I also show it helped to shift votes towards non-traditional opposition parties in the 2012 local elections, candidates who were relatively more aligned with students' demands.

Keywords: collective action, networks, protests

*This version: July 2018. First version: October 2016. I would like to thank Fred Finan, Solomon Hsiang, Edward Miguel, and Suresh Naidu for their support and guidance. I also thank Ernesto Dal Bó, Supreet Kaur, Gerard Roland, Joachim Voth, Noam Yuchtman, and seminar participants at the Center for Global Development, FGV/EPGE, IESE Business School, PUC-Chile, RIDGE/LACEA-PEG Workshop on Political Economy, UC Berkeley, Universidad de Chile, Universidad de Santiago, Universidad Diego Portales, University of Gothenburg, and University of Warwick for comments and suggestions.

†Pontificia Universidad Católica de Chile, Instituto de Economía. Email: `fagonza4@uc.cl`.

1 Introduction

Individual participation in collective action has long puzzled social scientists due to the combination of common benefits and private costs. This "collective action problem" has given rise to a large amount of theoretical literature emphasizing that the actions of others are crucial in understanding individual participation.[1] Despite its importance for theory, empirical investigations estimating how individuals respond to the participation of others are surprisingly scarce. The reason for the lack of evidence is the enormous data requirements, particularly important when studying protest behavior.

This paper studies the 2011 student movement in Chile, one of the largest mobilizations in the country's history. During days of national protests, hundreds of thousands of high-school students across the country skipped school to protest with the goal of reforming the educational system. After constructing a large, partially overlapping network with billions of links across students, schools, and cities, I provide evidence for the role of networks in protest behavior using administrative data of millions of students' daily school attendance on protest and non-protest days. The main finding is that skipping school on a protest day followed a pattern consistent with Granovetter's (1978) "threshold model of collective behavior." Students were influenced by their networks to skip school on national protest days only when more than 40% of the members in their networks also skipped school. Importantly, even though skipping school imposed significant educational costs on students, it also helped to shift votes towards non-traditional opposition parties in the 2012 local elections, candidates who were relatively more aligned with students' demands. Taken together, these findings indicate that networks amplify the effect of protests in non-linear ways with potentially significant consequences for institutional change.

Two features of Chile allow me to empirically study protest behavior. First, the central government assembles an exceptionally rich dataset of daily school attendance. Thus, I can measure protest behavior for more than 800,000 high-school students using school absenteeism on national protest days. Second, students' peer social ties are mainly with classmates (Araos et al., 2014), and information about their *lifetime* history of classmates is available. The latter data allows me to construct a countrywide network with more than 600 billion interactions across students, schools, and cities.

[1] See Olson (1965), Tullock (1971), Granovetter (1978), Tilly (1978), Kuran (1989, 1991), Lohmann (1993), Marwell and Oliver (1993), Chwe (2000), Bueno de Mesquita (2010), Edmond (2013), Little (2016), and Barbera and Jackson (2017) among many others.

The empirical analysis is divided in two parts. The first part focuses on days of national protests to estimate network effects in protest behavior. I use variants of the "partially overlapping networks" approach proposed by Bramoullé et al. (2009) and De Giorgi et al. (2010), in a two-stage non-linear framework, to estimate how school absenteeism in the network affects the individual decision to skip school. Using the non-parametric estimation proposed by Newey et al. (1999), results support Granovetter's (1978) "threshold model of collective behavior." To improve our understanding of the context in which decisions are taking place, the second part presents suggestive evidence of the effect that skipping school on protest days had on students' academic performance and electoral outcomes in the 2012 local elections.

Given the absence of random allocation of students across classrooms, I employ two variants of a "partially overlapping networks" approach. The first strategy focuses on the first *massive* protest on June 16, when school absenteeism in networks varies from zero to one-hundred percent, and uses the networks' exposure to the inaugural protest of May 12 as a plausibly exogenous variation. College students, outside of the high-school network, organized this event. Exposure is measured as school absenteeism among students who belong to the network of networks. This strategy is a variant of the partially overlapping networks approach because it uses variation *across* protest days and focuses on students in *different* schools. The second strategy focuses on all days of national protest in May and June, uses fixed effects by student and protest day, and exploits plausibly exogenous variation from the decisions in the network of networks attending different schools. This strategy is also a variant of the partially overlapping networks approach because it uses student and day fixed effects, and decisions of students in different schools.

Both empirical strategies rely on a strong first-stage and deliver robust findings. Network exposure to the inaugural protest is highly predictive of school absenteeism in June 16, even after controlling for a large set of observable variables for students, networks, schools, and city fixed effects. In addition, all results are robust to the inclusion of fixed effects at the neighborhood, school, or school-grade level. Moreover, placebo checks using non-protest days confirm the importance of the inaugural protest. Panel data estimates also deliver similar results and are robust to the inclusion of student, network, and school characteristics interacted with protest day indicators. These results are again similar when including two-way fixed effects by city-day, school-day, and neighborhood-day.

The estimated coefficients reveal that a threshold model is a better representation of individual decisions than a linear model. If the share of students in the network that skips school is lower than 40 percent, I observe that the individual decision to skip school is not affected by the network. In contrast, after the 40 percent threshold, individual absenteeism increases rapidly with network absenteeism. This result suggests that a "critical mass" of individuals is needed to facilitate protest behavior.[2] In comparison, linear-in-means estimates suggest that a 10 percent increase in network absenteeism increases the probability of skipping school by 7 percent.

The critical mass of 40 percent should be interpreted as an average threshold. Students in larger schools, smaller networks, and smaller cities have lower thresholds, results broadly consistent with Olson's (1965) theory of groups. In addition, when augmenting the estimation to allow for differential non-linear effects within networks, the estimates suggest that students are more influenced by others who are similar to them. Taken together, the results reject a linear-in-means model and suggest a critical mass of others who are similar is needed to foster protest behavior. As a consequence, school absenteeism across network groups in the country follows a bimodal distribution with low and high levels of absenteeism.

To improve our understanding of the context in which school attendance decisions are taking place, I study the costs of skipping school on protest days and also their effect on electoral outcomes. A differences-in-differences analysis among primary and high-school students in the period 2002–2015 reveals that grade repetition increased by 60 percent, from a base of 6 percent, among high-school students in 2011. Skipping school on the June 16 national protest – which led to higher absenteeism in the following months – decreased GPA by 0.1 standard deviations and increased grade repetition by 33 percent. The cost of skipping school also follows non-linear network patterns. In addition, I provide suggestive evidence that the student movement was able to shift votes towards non-traditional opposition parties, which were relatively more aligned with the movement's demands. A cross-sectional regression using county-level electoral data suggests that a one standard deviation increase in the intensity of the movement in local schools increased vote shares for non-traditional parties by 5 percentage points, an increase coming mostly from right-wing candidates.

This paper contributes to the empirical understanding of participation in collective action. Only a few papers have studied *protest* behavior, and the role of social effects has been overlooked.[3] Notable exceptions include Cantoni et al. (2017), who show that beliefs about others'

[2]This tipping behavior is predicted by models of social interactions (e.g. Brock and Durlauf 2001). However, empirical evidence is limited. A notable exception is Card et al. (2008), who use Census tract data to provide evidence of tipping in the context of Schelling's (1971) dynamic model of segregation.

[3]There are studies of participation in other types of collective action. For example, McAdam (1986) shows that friends' participation in the 1964 Freedom Summer project predicts individual participation, and Yanagizawa-Drott (2014) shows that radios facilitated participation in the Rwandan genocide. There is also a vast literature studying the role of social interactions more generally, see Durlauf and Ioannides (2010) for a review.

turnout affect individual participation in the context of Hong Kong's democracy movement; Enikolopov et al. (2017) who show social image was an important motivation to participate in the 2011-2012 protests in Russia; and Larson et al. (2017) who use Twitter data to show that individuals' network position influenced attendance at the 2015 Charlie Hebdo protests in Paris. In contrast, my paper focuses on individual-specific networks and uses administrative data to test for threshold models in protest behavior.

Another branch of the literature focuses on the role of the media. Using an annual panel dataset of geographic cells in Africa, Manacorda and Tesei (2016) show that mobile phone coverage facilitated protests when countries experienced economic downturns. Enikolopov et al. (2018) show that the penetration of an online social network in Russia increased the probability of protest and the number of protesters across cities. There is also a growing literature studying the role of networks and online media in spreading political information (Halberstam and Knight, 2016; Qin et al., 2017; Yang, 2018), with some papers showing how citizens' discontentment online can predict real-world protest participation (Acemoglu et al., 2018). However, to the best of my knowledge this is the first paper to estimate how individual-specific networks affect individual protest behavior.

This paper also speaks to a literature estimating the consequences of protests.[4] Madestam et al. (2013) uses rainfall shocks as exogenous variation affecting the number of protesters in the Tea Party movement across U.S. counties to show how the movement affected electoral outcomes and the policies being implemented. Aidt and Franck (2015) show that the Swing riots in early 19th century Britain – credible signals of the threat of a revolution – facilitated democratic reforms. This paper contributes to this literature by providing novel evidence on the individual costs associated with protest behavior and suggestive evidence on the effect of that protest behavior on electoral outcomes.

The next section provides details about the 2011 student movement. Section 3 presents the data and the econometric strategy. Section 4 presents results. Section 5 estimates the costs of skipping school and its political effects, while Section 6 concludes.

[4]There is, of course, a large theoretical literature studying social unrest and political transformation. See, for example, Acemoglu and Robinson (2000).

2 The Chilean Student Movement

From the Tunisian demonstrations sparking the Arab Spring to Occupy Wall Street triggering a movement against inequality, 2011 was a year full of protests across the world. The global wave of citizens demanding a "new democracy" also took place in Chile, where high-school and university students revolted to reform the educational system installed by the Pinochet dictatorship, nowadays one of the most expensive and segregated in the world (Hsieh and Urquiola, 2006; OECD, 2013). Organized groups of students triggered one of the largest demonstrations in the country's history, which were recognized worldwide as one of the most important social movements of that year.

The student movement began in May 2011, two months within the academic year, and 14 months after a right-wing government took office democratically for the first time in 50 years.[5] Initial demonstrations were triggered by delays in the assignment of students' scholarships and bus passes. The first student-led national protest took place on May 12 and thousands of high-school and university students participated.[6]

The first protests were organized by the Confederation of Chilean Students, a national student organization, and had the objective of exerting pressure on the annual presidential speech on May 21, in which the government outlines the next year's policies. Students wrote a document proposing policies to decrease segregation in the educational system and increase government spending. After the presidential speech, the Confederation sent a letter to the Ministry of Education expressing their discontent with the presidential announcement (Confech, 2011). Students called for another national protest day in June 1, the last rally before the movement expanded in an unprecedented way.

After the national protest on June 1, and a failure to reach an agreement with the Ministry of Education in meetings held on May 30 and June 8, students intensified their protest activities. The movement was gradually supported by deans, teachers, prominent labor unions, and public figures. Over the weeks that followed, students occupied schools and universities, and protest activities spread across the country. In an attempt to prevent occupations, the Ministry of Education asked students "to stop protesting" and the president stated that "countries do not progress by occupying schools." The government's approval rating was low and continued to plummet

[5] Chronicles written by leaders of the student movement include Figueroa (2012), Vallejo (2012), and Jackson (2013). A brief history of the high-school movement can be found in Simonsen (2012).

[6] For additional context, Figure A.1 plots the daily number of protests in Chile in the period 1979-2013, and Figure A.2 plots economic indicators around the beginning of the student movement of 2011.

after the rise of the movement (Figure A.3). Students called for another national protest day on June 16, at the time the largest mobilization in the country's history. The government responded in June 25 with an offer, which students rejected, calling for yet another national protest day on June 30.

Education was the main topic of conversation during July and August. The leaders of the movement were regularly invited onto television and radio shows, and diverse protest activities filled the country. The president replaced the Ministry of Education on July 18 and the government responded to students' demands with offers on July 5, August 8, and August 17. Students rejected these offers and demonstrations continued after the July winter break, with the largest national protests taking place on August 24 and 25. These two days marked the peak of the student movement, and protest activities declined in the following months.

Various reasons explain the decay of the student movement, including the beginning of formal negotiations, the focus of popular media on violent protesters, and students' concerns about grade retention.[7] After months of protests, what were the consequences? Contemporary surveys show that 80 percent of citizens supported the movement (Adimark, 2011) and that education became a national priority (Figure A.4). Candidates in the 2012 local elections and 2013 Congress and presidential elections were constantly questioned about their ideological positions regarding education. Some of the older leaders of the movement founded political parties and four of them won seats at the congress.

3 Empirical framework

This section describes the data and two econometric strategies used to test for complementarities in school absenteeism decisions between students and their networks on protest days. Both strategies employ variations of the "partially overlapping networks" approach first proposed by Bramoullé et al. (2009) and De Giorgi et al. (2010). In the first strategy, the focus is on June 16, one of the largest protests and the first massive event, and exploits the exposure of students' networks to the *inaugural* protest in May 12 in *other* schools. The second strategy focuses on all national protest days in May and June and exploits within-student variation in the actions of

[7]"The constant emphasis on violence affected the strength of the movement" (Jackson, 2013, p. 22). The government threatened students with being held back, promoting the "Let's save the academic year" plan. In addition, public figures died in an airplane crash in September 2 – shifting public interest away from the movement – the movement's leaders had to face annual elections to renew their leaderships, and summer holidays caused the movement to slow down until the next academic year.

partially overlapping networks. I discuss estimating equations and econometric concerns in this section and provide details about the estimation procedure when presenting results in Section 4.

3.1 Administrative data

The analysis uses five administrative datasets. The first dataset measures *daily* school attendance in 2011. The academic year in Chile starts in March and ends in November, with a winter break in July. The second reveals students' enrollment information (school, grade, and classroom) for 2011 and previous years. There were approximately 975,000 high school students enrolled in 2,700 high-schools in 2011. The third dataset contains information about students' annual academic performance, i.e. GPA. The fourth dataset contains self-reported home addresses and is available for 35 percent of high-school students in 2011. I geocoded 50,000 home addresses in Santiago, the capital of Chile, to test for the role of neighborhoods, defined as small geographic units within cities. The last dataset describes schools. Approximately 40 percent of students were enrolled in public schools in 2011 and 60 percent in private schools. School addresses are available and I use these to construct geographic clusters that I refer to as "cities." There are 290 cities, with 8 high schools and 2,600 high-school students in the average city.[8] Table 1 presents descriptive statistics for the main variables.

Networks. Because students mainly interact with other students in their classrooms, I define student i's network $j(i)$ as her lifetime history of classmates. As of 2011, each high-school student had a unique set of classmates that I identify from historical enrollment information. This definition gives rise to a large network of students linked within *and* across classrooms, schools, and cities. Links across schools originate in the predetermined switching of students across schools before 2011, and I control for the switching decision in several ways in the econometric model. Overall, this network contains more than 600 billion potential interactions among students across the entire country, and more than 60 million existing links. The average student has 80 other students in her network, 60 percent attending the same school and 40 percent attending a different school in 2011.[9] Unfortunately the network is too large to calculate network statistics and solutions to this problem rely on approximations that are currently being evaluated (e.g. Brandes and Pich 2007 and Alghamdi et al. 2017).

[8]In practice, cities are isolated components in the spatial network of schools, where two schools are linked if these are closer than 5 kilometers from each other. Figure A.5 presents a map of cities.

[9]For computational reasons I only consider classmates in years 2007-2010. The calculation of student-specific network variables takes substantially more time when including more years. However, I estimated some regressions using more years and results are robust.

Protest days. To measure protest behavior related to student strikes I use school absenteeism among high-school students on national protests days. The government collects daily school attendance to track performance and allocate public programs. Several patterns in the data suggest school absenteeism is a useful way to measure protest behavior. First, there are significant spikes in school absenteeism on protest days. The upper panel in Figure 1 plots absenteeism throughout the 2011 school year. The first two national protest days (May 12 and June 1) are easy to observe. The sharp increase in school absenteeism between June 1 and June 16 corresponds to the real-time escalation of protest activities. Second, some schools were temporarily taken over by students, and these closures are observed in the data with the same dates reported in local newspapers. As examples, the lower panels in Figure 1 present three school-level time series. Schools with 100 percent absenteeism for prolonged periods correspond to schools occupied by students.

3.2 Econometric strategy using exposure to first protest

Consider the following regression relating a student's decision to skip school on a protest day as a function of school absenteeism in her network:

$$A_{isc} = f(A_{j(i)}) + g_1(x_i) + g_2(x_{j(i)}) + \delta x_s + \zeta_c + \epsilon_{isc} \tag{1}$$

where $A_{isc} \in \{0, 1\}$ takes the value of one if student i in school s, located in city c, decides to skip school on June 16. In addition, $f(A_{j(i)})$ is a function of a vector of absenteeism decisions in i's network $j(i)$, and $g_1(x_i)$ and $g_2(x_{j(i)})$ are flexible functions of observables that account for benefits and costs that may affect a student's decision. Finally, x_s is a vector of control variables at the school level, ζ_c is a city fixed effect, and ϵ_{isc} is an error term clustered by city.

The vector of control variables x_i include average school attendance in 2010, GPA in 2010, an indicator for grade retention in 2010, an indicator for gender, an indicator for students who switched school in 2010, and age. Averages of the same variables are included in $x_{j(i)}$, although results are robust to the use of more flexible functions such as fully saturated bins for all controls. In addition, student controls also include school absenteeism on previous (and smaller) protest days, i.e. May 12 and June 1. School-level controls include an indicator for public schools, reported quality signals (i.e. test score averages), the percentage of students who have repeated a grade in the past, and average household income.

The first part of the analysis uses a linear-in-means function f, i.e. the average absenteeism in networks $f(A_{j(i)}) = \frac{\sum_{k \in j(i)} A_k}{N_{j(i)}} \equiv \overline{A}_{j(i)}$. Then, I allow network absenteeism to flexibly influence individual decisions by using the following functional form for network decisions:

$$f(A_{j(i)}) = \beta_1 \cdot 1\left[\overline{A}_{j(i)} \in [0.1, 0.2)\right] + \cdots + \beta_9 \cdot 1\left[\overline{A}_{j(i)} \in [0.9, 1)\right] + \beta_{10} \cdot 1\left[\overline{A}_{j(i)} = 1\right] \tag{2}$$

where $\beta_1, \ldots, \beta_{10}$ are the parameters of interest and $1[\cdot]$ is an indicator function that takes the value of one when the statement in square brackets is true. I use eleven indicators, although results are robust to using more; the first takes the value of one if absenteeism in networks is between 0 and 10 percent, the second takes the value of one for 10-20 percent absenteeism in networks, and so on progressively until I reach 100 percent absenteeism in networks. The omitted category is network school absenteeism lower than 10 percent.

There are three concerns with an estimation of equation (1) using OLS. First is the classical reflection problem emphasized by Manski (1993): students affect their networks and networks affect students. Second, given the absence of random allocation of students across classrooms, there may be unobservable variables causing students and their networks to make similar decisions. Both concerns imply that an OLS estimation will overestimate the effect of networks. A third problem is known as "exclusion bias" and causes OLS estimates to be biased *downwards* (Guryan et al., 2009; Angrist, 2014; Stevenson, 2015; Caeyers and Fafchamps, 2016). To solve the former two issues, I use three sources of variation in an instrumental variables approach that exploits partially overlapping networks. To solve the third issue, I follow Caeyers and Fafchamps (2016) and include the student's value of the instrument as an additional control.

The first source of identifying variation is the exposure of networks $j(i)$ to protests in their networks, i.e. the "excluded network," which directly addresses the reflection problem. The second source is a restriction to the set of students in the "excluded network"; I focus only on those attending a *different* school than i in 2011, approximately 2,000 students. This restriction addresses concerns regarding unobservable variables and shocks affecting students in the same school. The third source of variation corresponds to school absenteeism on the first national protest day, May 12, organized outside of the network of high-school students (see Section 2). This final source of variation can be thought of as similar to the "partial population approach" in Dahl et al. (2014) in which a subset of the population is exogenously exposed to participation in a program (a protest in this case). All in all, this strategy is a variant of the "partially overlapping networks" approach proposed by Bramoullé et al. (2009) and De Giorgi et al. (2010).

To gain intuition about the strategy, recall student i's network is $j(i)$. The exposure of students in $j(i)$ is measured by how much *their* networks, say $J(i) \equiv j(j(i))$, skipped school on

May 12, with $i \notin J(i)$. Students in the set $J(i)$ may however still have unobservables similar to those of i. To deal with this concern, I restrict attention to a subset of students. Given the predetermined switching across schools, many students in $J(i)$ are attending a different school than i in 2011. Let $M(i)$ be the set of students that attend a different school than i in 2011, with $M(i) \subset J(i)$ and $j(i) \cap M(i) = \emptyset$. The identification assumption is thus that school absenteeism on May 12 among students in the set $M(i)$ only affects student i's absenteeism on June 16 through the absenteeism of $j(i)$.

The first stage using the previously described plausibly exogenous variation is strong (see Table A.1), with coefficients having the expected positive sign – higher exposure to initial protests fosters future school absenteeism – and corresponding F-stats that are always far from a weak instrument problem (Stock and Yogo, 2005). Reassuringly, the value of the instrument on non-protest days before May 12 does not predict networks' absenteeism on June 16 (see Figure A.6), suggesting that unobservable variables that affect absenteeism in non-protest days are unlikely to be affecting the results.

3.3 Partially overlapping networks in panel data

The second strategy exploits the multiple protest days observed in the data. I focus on all national protest days before the winter break of July. This decision is motivated by a potential change in the structure of networks after the break, but given the large number of observations it does not affect the statistical power of the analysis. In particular, I estimate versions of the following equation:

$$A_{isct} = f(A_{j(i)t}) + \sum_t \left(\delta_{1t}x_i + \delta_{2t}x_{j(i)} + \delta_{3t}x_s\right) + \xi_i + \zeta_{ct} + \epsilon_{isct} \tag{3}$$

where A_{isct} is an indicator that takes the value of one if student i, in school s, located in city c, decides to skip school on day t, a day of national protest. In addition, $f(A_{j(i)t})$ is a function of a vector of absenteeism decisions in i's network $j(i)$ in day t, and x_i, $x_{j(i)}$ and x_s are flexible functions of observable variables by students, networks, and schools, that may affect absenteeism decisions. These controls are the same as those in equation (1). Finally, ξ_i is a student fixed effect, ζ_{ct} is a city by day fixed effect, and ϵ_{isct} is an error term clustered by city. As in equation (1), I employ the functional form in equation (2) to test for non-linear network effects.

Note that, when using an OLS approach, the assumption for a consistent estimation of the parameters $\beta_1, \ldots, \beta_{10}$ is different than in the previous strategy. Indeed, because I am now using *within student* variation in absenteeism decisions, the main threat is the reflection problem and unobservable variables that vary over time. To deal with the reflection problem I again use the partially overlapping networks approach, restricting attention to students in *other* schools. In addition, to control for potential unobservable variables I interact protest day indicators with (1) student, network and school characteristics; and (2) include the following two-way fixed effects: protest day by city, protest day by neighborhood, and protest day by school.

4 Main Results

4.1 Linear estimates

Table 2 presents OLS estimates of a linear-in-means model using different specifications of equation (1).[10] Note that, because the mean of the dependent variable and the network variable of interest are similar (0.49 and 0.50), point estimates can be interpreted as an elasticity. For brevity I only discuss cross-sectional estimates, but their panel counterparts are similar in this linear-in-means model.

Column 1 of Table 2-A presents network estimates without including controls, a regression that explains almost two-thirds of the variation in June 16 absenteeism.[11] The coefficient implies that a one standard deviation increase in network absenteeism (0.31) is associated with an increase of 38 percentage points ($0.31\times1.23=0.38$) in the probability of skipping school. In terms of elasticities, a 10 percent increase in network absenteeism is associated with a 12 percent increase in student absenteeism. Columns 2-5 progressively include controls for student, network, and school characteristics, and city fixed effects. As a result, the coefficient of networks remains stable. Although regressions control for a large set of observable variables at multiple levels, reflection and unobservable variables could cause a comovement of decisions between students and their networks.

A leading concern with estimates in Table 2-A are neighborhood unobservable variables causing a spurious positive correlation between students and their networks. To explore this possibility, I geo-coded 50,000 home addresses of students in Santiago, the capital of Chile, and construct neighborhood fixed effects using latitude and longitude coordinates, creating areas of

[10] The number of observations is presented at the bottom of Table 2. Differences in observations are due to missing values, which are more common for small schools located in rural areas.

[11] In contrast, a regression of daily absenteeism on characteristics of students, networks, and schools explains less than one-third of the variation, suggesting network effects are important.

approximately 10×10 blocks (see Figure A.7 for a map). Column 6 includes these 714 indicators and estimated coefficients remain unchanged, providing some evidence that neighborhood-level variables are unlikely to be a concern. Column 7 includes school fixed effects and the coefficient for networks decreases to 0.07 (s.e. 0.03, first-stage F-stat of 77.7).

To deal with the reflection problem and the remaining possibility of unobservable variables confounding my estimates, I implement the two-stage estimation strategy. Table 2-B presents results while Table A.1 presents first-stages and reduced forms. As expected, estimated coefficients are positive and smaller in magnitude than their OLS counterparts. Columns 1-6 show the coefficient is robust, with an elasticity of 0.6–0.8, and column 7 shows a smaller elasticity of 0.07. Importantly, F-statistics in first-stages are always strong. Given the meaningful variation in absenteeism across schools, column 5 in Table 2-B is my preferred specification as it includes school-level controls but allows for absenteeism at this level to affect individual decisions. The estimated network coefficient using the two-stage strategy is robust to excluding schools closed by students on June 16, with a coefficient of 0.53 (s.e. 0.14) and a first-stage F-stat of 30.2 (see Table A.2).

4.2 Non-linear estimates

4.2.1 Exposure to the first national protest

To test for non-linear networks effects, I estimate equation (1) with the functional form in equation (2), and use the non-parametric approach proposed by Newey et al. (1999). This estimation strategy corresponds to a control function, and the coefficients of interest are associated with indicators for different values of absenteeism in the network. In all of the following regressions the coefficient associated with the control function parameter is statistically different from zero, as expected given the difference between previous two-stage linear estimates and their OLS counterparts. In addition, the following cross-sectional and panel data estimates are robust to using more modern estimation techniques, such as the method proposed by Rau (2013) that allows for the unobservables from the first-stage to enter non-linearly and interacted with observables in the second stage.[12]

Figure 2-a presents estimated coefficients $\left(\widehat{\beta}_1, \ldots, \widehat{\beta}_{10}\right)$ with their corresponding 95 percent

[12] I use these estimation techniques because of their simplicity, but there are other nonparametric estimations available (e.g. Newey and Powell 2003; Chetverikov and Wilhem 2017). A comprehensive comparison between different estimation strategies, which rely on somewhat different assumptions, is beyond the scope of this paper.

confidence interval using the Newey et al. (1999) approach and the corresponding OLS estimates for comparison. The exact econometric specification corresponds to that in Table 2 column 5, which includes student, network, and school controls, and also city fixed effects. Similar to before, this estimation technique delivers estimates that are lower in magnitude than their OLS counterparts, as expected.

The estimated two-stage coefficients are consistent with the threshold model of collective behavior proposed by Granovetter (1978) in the following sense. The school absenteeism decision of a student seems to not be affected by low values of school absenteeism in her networks. In contrast, large values of network absenteeism do seem to have strong effects on her decision to skip school. To more clearly show the marginal contribution of additional absenteeism in networks, Figure 2-b plots the sequential difference between estimated coefficients:

$$\varphi_k \equiv \widehat{\beta}_k - \widehat{\beta}_{k-1} \qquad \text{with } k = 1, \ldots, 10 \tag{4}$$

where $\beta_0 = 0$, and $\widehat{\beta}_1, \ldots, \widehat{\beta}_{10}$ correspond to the estimated coefficients in equation (2). Results in this figure suggest that the influence of networks on individual decisions is positive only after absenteeism reaches 40 percent of a network, and reaches a maximum around the 50-60 percent mark. Figure 2-c plots the distribution of absenteeism in networks, a bimodal distribution centered around 50-60 percent. Overall, patterns in Figure 2 are consistent with Granovetter's threshold model of collective behavior in which networks start to positively influence individual decisions after a threshold of 40 percent participation.

These non-linear network patterns are robust. Panels A, B, and C in Figure 3 show that estimated coefficients are also non-linear when I include school or neighborhood fixed effects and are also similar when I omit the set of schools that were closed by students from the equation. In addition, the maximum influence of networks around 40 percent should be interpreted as the average threshold in Granovetter's model. Indeed, panels D, E, and F in Figure 3 show that students in larger schools, with smaller networks, and in smaller cities have lower thresholds, results broadly in line with Mancur Olson's theory of groups.[13]

4.2.2 Panel data estimates

Two-stage panel data estimates of equation (3), which allow for non-linear network effects following equation (2), confirm previous results. For estimation, I again use the non-parametric

[13] Figure A.8 shows additional patterns of heterogeneity by gender, age, students who switched school, and by type of school (public vs. private).

control function approach proposed by Newey et al. (1999).

Figure 4 presents panel estimates of $\left(\widehat{\beta}_1, \ldots, \widehat{\beta}_{10}\right)$. Panel (a) presents two-stage estimates of protest day absenteeism on network absenteeism using the baseline non-linear "partially overlapping network" approach, which includes student and protest day fixed effects, and includes controls for student, network, and school time-invariant characteristics interacted with protest day indicators. The following panels include additional two-way fixed effects: city by day in panel (b), neighborhood by day in panel (c), and school by day in panel (d). All of these regressions employ more than six million observations, coming from approximately 800 thousand students for seven protest days.

The estimated coefficients in Figure 4 reveal the same non-linear network patterns found in the previous section and, in that sense, confirm previous results. However, the magnitude of coefficients and the shape of non-linearities are somewhat different across panels. Coefficients are smaller and non-linearities more clear when I include city-day, neighborhood-day, and school-day fixed effects. Despite of these differences, I again estimate that networks begin to influence individual decisions after 40 percent absenteeism and the marginal contribution of additional absenteeism – φ in equation (4) – is again maximized around 50-60 percent.

4.3 Heterogeneity

This last section of main results studies heterogeneity in network effects, in particular homophilic influence. The empirical regularity of individuals forming links with other individuals of similar characteristics is known as homophily (Jackson, 2010, chapter 6). Empirical work testing differential *influence* following homophily patterns within networks is, however, more limited. Conditional on a network structure, does the strength of influence follows homophily patterns? I present evidence focusing on three variables: gender, internet connection, and household income.

Figure 5 presents two-stage $\left(\widehat{\beta}_1, \ldots, \widehat{\beta}_{10}\right)$ coefficients for the non-linear model. Panels (a) and (b) test for gender homophily patterns of influence by estimating equation (1), restricting attention to males or females, and splitting the network into males and females. For the estimation, I use the same control function approach as before. Under the null hypothesis of equal influence we should observe similar coefficients for the male and the female networks. Results, however, indicate strong homophily patterns: same gender influence is more than ten times stronger than cross gender influence.

Panels (c) and (d) use the same estimation strategy but restrict attention to students with and without internet access, again splitting the network into two: students with and without internet access. The influence of students with internet access on other students with access is almost three times larger. The influence of students without internet access on students also without access is two times larger.[14] Similar patterns of influence arise when restricting attention to the position of students' parents in the income distribution. Panels (e) and (f) show that students from low-income households are more influenced by students also from low-income households, and students from high-income households are more influenced by students also from high-income households.[15]

Overall, non-linear network patterns are still clearly visible when allowing for differential influence, and the null hypothesis of equal influence is easily rejected.

4.4 Importance of results

The estimates presented so far constitute, to the best of my knowledge, one of the first empirical evidence supporting critical mass models of collective action in protest behavior. In this sense, my results are important because they suggest that this class of models seems more appropriate than other models as a foundation to understand individual decisions to protest. It is, however, important to recognize that the non-linear complementarities behind a critical mass do not imply that free-riding (i.e. substitution in actions) is unimportant. Indeed, it is empirically and theoretically possible that free-riding exists even in the presence of non-linear complementarities. Moreover, I find some evidence of substitution when school absenteeism in networks is high, where marginal increases in network absenteeism are associated with decreases in the individual probability of skipping school. See Figures 2 and 4, for above 70% network absenteeism.

To improve our understanding of previous estimates, the following section studies the consequences of protests. The motivation is to gain insights about the context in which school absenteeism decisions are taking place. Skipping classes without real costs for students nor political consequences would imply that school absenteeism decisions are of minor importance. The next section shows, however, that the protest behavior I have documented had significant

[14] This is a partial test for the hypothesis of stronger coordination with internet access because students may also have internet access at school. Manacorda and Tesei (2016) and Enikolopov et al. (2018) provide city-level evidence of stronger network coordination with increased access to cell phones and social media.

[15] High-income households are defined as those with reported annual income higher than US$16,000, low-income households with reported annual income lower than US$5,000, and the remainder is defined as the middle class.

educational costs for students, and helped to shift voting patterns in subsequent elections. In this sense, results in both of these sections suggest that these decisions were taking place in a high-stakes environment.

5 Consequences of Protests

This section estimates the private cost of skipping school on protest days and its effects on electoral outcomes. Results show that grade repetition increased by 60 percent among all high-school students in 2011, and school absenteeism on protest days is non-linearly associated with lower academic performance and higher grade retention. In addition, an analysis of the 2012 local elections suggests that the student movement successfully shifted votes towards non-traditional opposition parties, which were relatively more aligned with the movement's demands.

5.1 The cost of skipping school

An analysis of administrative data for the period 2007–2015 shows that skipping school led to increased grade retention, an outcome causally associated with higher dropout rates, lower educational attainment, and more criminal activities (Manacorda, 2012; Díaz et al., 2017). To estimate the change in grade retention among high-school students in 2011, I estimate the following regression:

$$y_{hst} = \beta_t \times (G_{hs} \times T_t) + \zeta_{hs} + \lambda_t + \varepsilon_{hst} \tag{5}$$

where y_{hst} is retention of students in grade h of school s in year t, with h representing either students in 1st-4th grade (non-protesters) or students in 9-12th grade (high-school, i.e. protesters). The indicator G_{hs} is equal to one for grades 9-12th and zero otherwise, T_t is a vector of indicator variables for years $t = 2007, \ldots, 2015$ (with 2010 as the omitted category), ζ_{hs} and λ_t are school-grade and year fixed effects, and ε_{hst} is an error term correlated within schools. An increase in grade retention among high-school students in 2011 translates into $\beta_{2011} > \beta_t$, with $t \neq 2011$.

Figures 6-A and 6-B present coefficients $\widehat{\beta_t}$. Figure 6-A uses absenteeism as dependent variable and Figure 6-B uses grade retention. High-school absenteeism increased by 4.5 percentage points in annual official statistics, a 60 percent increase from a base of 8 percent absenteeism in

2010.[16] Retention among high-school students increased by 3.5 percentage points in 2011, a 60 percent increase from a base of 6 percent in 2010.

For a better understanding of previous estimates I now focus on 2011 and estimate student-level costs. For the estimate, consider a version of equation (1) that includes school fixed effects and uses academic performance at the end of the 2011 academic year (December) as dependent variable. The coefficients of interest are again flexible estimates of network absenteeism on June 16. Figures 6-C and 6-D present estimates using grade point average (standardized GPA) and an indicator for grade retention as dependent variables. Estimated coefficients imply that 100 percent absenteeism in networks on June 16 is associated with (1) a decrease of 0.16 standard deviations in academic performance, and (2) an 38 percent increase in grade retention (from a base retention of 6 percent in 2010).

Finally, consider a similar regression using student-level absenteeism on the first massive protest day as the main independent variable. Estimated coefficients suggest that individual school absenteeism in June 16 – which is correlated with higher absenteeism in the following months – leads to (1) a decrease of 0.1 standard deviations in GPA (coefficient of -0.07, p-value<0.01), and (2) an 33 percent increase in grade retention (coefficient of 0.02, p-value<0.01). Results using annual school absenteeism as independent variable imply that a one standard deviation increase in absenteeism decreases GPA by 0.15 standard deviations and increases grade retention by 31 percent. Overall, estimates suggest sizable costs of skipping school on protest days.

5.2 The political effects of the student movement

The first election after the rise of the student movement was held on October 2012.[17] In these elections citizens elected mayors in all 345 counties in Chile. Traditional parties, organized into left and right wing coalitions, competed against each other and against candidates from "non-traditional" parties. Although with new leaders and lower participation rates, the student movement was still active and many anticipated it would have an effect on electoral outcomes. The movement showed its discontent with traditional politics and publicly supported non-traditional

[16]This increase in absenteeism needs to be interpreted with caution as both the denominator and the numerator are changing. The central government decreased the total number of official days of school in 2011.

[17]There was an informal plebiscite previously organized by citizens, in October 2011. Figure A.4 shows that participation was higher and people agreed more with students' demands in counties with higher school absenteeism.

parties.[18]

Despite its contemporary relevance, there is no research on the impact of the student movement on these elections. To estimate the effect of the student movement in the 2012 local elections, I estimate versions of the following regression equation:

$$V_{c,2012} \quad = \quad \alpha + \beta \cdot \text{Student Movement}_{c,2011} + \gamma V_{c,2008} + \delta X_{c,2009} + \varepsilon_c \qquad (6)$$

where $V_{c,2012}$ and $V_{c,2008}$ are electoral outcomes in the 2012 and 2008 local elections in county c and $X_{c,2009}$ is a vector of controls available for 324 counties, i.e. population, average household income, and average years of education. Student Movement$_{c,2011}$ is the county-level average *increase* in high-school absenteeism after the beginning of strikes, calculated as high-school absenteeism between May and November minus high-school absenteeism in March and April. By measuring absenteeism all days after May, I am able to capture absenteeism in schools that were closed by students for the whole year. Finally, ε_c is a robust error term. The dependent variable I examine are the vote shares for non-traditional candidates, left and right-wing candidates, the percentage of voters in the county population, number of non-traditional candidates competing, and total number of candidates.[19]

The main concern with an OLS estimation of β is the potential existence of omitted variables correlated with the student movement and electoral outcomes. Three exercises suggest this is unlikely to be a major concern. First are regressions controlling for electoral outcomes in previous elections, which captures cross-sectional variation in political preferences. Second are placebo checks using school absenteeism and elections in previous years, which support the previous results. Third, I use the method proposed by Altonji et al. (2005) to construct bounds for estimates and conclusions remain.

Table 3 presents regression estimates. Column 1 indicates that a one standard deviation increase in the intensity of the student movement is associated with a 5 percentage point increase in the vote share for non-traditional candidates, an increase of 15 percent (base of 34 percent in 2008). Columns 2 and 3 show that this increase in vote shares is mostly explained by a decrease in vote shares for right-wing candidates, the coalition of the incumbent president.[20] Column 4

[18]One popular election involved the non-traditional (independent) candidate Josefa Errázuriz – explicitly supported by the student movement – competing against the traditional (right-wing) candidate Cristián Labbé, mayor of *Providencia* county between 1996 and 2012. Errázuriz won that election.

[19]Electoral outcomes are administrative data reported by the Electoral Service of Chile. Population data come from censuses. Figure A.9 plots the student movement variable for all counties.

[20]More speculatively, Figure A.10 provide some suggestive evidence of non-linear effects that are consistent

suggests that the same increase in the movement intensity is associated with a decrease of 0.6 percentage points in votes. Column 5 and 6 suggest there were little changes in the number of competitors at these elections. As placebo checks, I create fake local movements using the differential increase in county-level school absenteeism between 2008 and 2007, i.e. before the rise of student movement, and examine their impact in the 2008 local elections. I also re-estimate equation (6) using 2008 vote shares as dependent variable and 2004 vote shares as controls. Reassuringly, the "fake movements" do not have an effect on electoral outcomes and the 2011 student movement does not predict 2008 electoral outcomes.

Local elections are a natural setting to use the Altonji et al. (2005) method to study a potential bias due to unobservable variables because past electoral outcomes are powerful predictors of outcomes at the county level. Oster (2017) emphasizes that changes in the r-squared from an uncontrolled to a controlled regression can be used to obtain an adjusted coefficient that accounts for unobservables. This "coefficient stability approach" confirms previous results and suggests the effect of the movement on votes for non-traditional candidates is in the range $[0.050, 0.086]$.[21]

6 Conclusion

Studying the Chilean student movement of 2011, this paper showed that students were influenced by their networks to skip school on a protest day only when a "critical mass" of 40 percent of their networks also skipped school. Overall, results support the popular idea of a tipping point in behavior (Gladwell, 2000) and the importance of strong ties to promote political activism (McAdam, 1986).

The findings in this paper have at least two implications. First, results are relevant for the modeling of collective action in networks. Theoretical work has emphasized that protest participation may be modeled as a game of strategic complements or strategic substitutes. The "critical mass" type of influence found in this paper suggests that complementarities are relevant for at least some participation levels. Results also point towards the possibility of protest participation as strategic substitute – i.e. individuals free-riding on the participation of others – but only for high participation levels in network groups.

---

with the previous "critical mass" patterns.

[21] Bounds use $\widehat{\beta} = \beta_c - (\beta_{nc} - \beta_c)\frac{R_{max}-R_c}{R_c-R_{nc}}$, where β_c and β_{nc} are coefficients from a regression with and without controls with corresponding R-squared of R_c and R_{nc}, and R_{max} is an unknown parameter in the interval $[R_c, 1]$. I use the conservative assumption of $R_{max} = 1$. See Oster (2017) for details.

Second, complementarities in protest behavior imply that individuals with larger networks are more influential. This corollary is potentially extremely important for both the organization of a social movement and its disruption. For example, imagine a group of individuals organizing a social movement to bring down a dictatorship, as the Otpor! movement in Serbia in the 1990s. The findings presented in this paper suggest that the marginal return of enrolling one additional citizen in the movement is higher for individuals with larger networks. In addition, an organization may exploit the "critical mass" patterns by exerting effort to go beyond the threshold. In the same way, a state could decrease participation in a social movement by preventing central individuals from participating or by exerting effort to avoid reaching a "critical mass."

Two additional remarks are necessary to interpret results more broadly. Firstly, students may be subject to more or less influence from their networks than the non-student population. This is more than a passing concern – after all many important movements have been started by students – the setting may restrict the external validity of results to interpret social movements originating in non-student populations. In the second place, the lack of a precise identification of the mechanisms behind the results may also hinder their external validity. The lack of emphasis on beliefs about the actions of others and the missing dynamics in network structure also prevent us from a full understanding of the decision to participate in a social movement.

My findings suggest that social dimensions in protest behavior are important, and open new and interesting questions to explore. For example, future empirical studies of social movements may explore how protests create network links among participants and the consequences, how police violence in protests disrupt (or foster) participation, and how habit formation contributes to the escalation of a mobilization. For now, we have evidence that networks amplify the effect of protests in non-linear ways with potentially significant consequences for institutional change.

References

Acemoglu, D., Hassan, T., and Tahoun, A. (2018). The power of the street: evidence from Egypt's Arab Spring. *Review of Financial Studies*, 31(1):1–42.

Acemoglu, D. and Robinson, J. (2000). *Economic origins of dictatorship and democracy*. Cambridge University Press.

Adimark (2011). Encuesta: evaluación gestión del gobierno. *Adimark GfK*.

Aidt, T. S. and Franck, R. (2015). Democratization under the threat of revolution: evidence from the great reform act of 1832. *Econometrica*, 83(2):505–547.

Alghamdi, Z., Jamour, F., Skiadopoulos, S., and Kalnis, P. (2017). A benchmark for betweenness centrality approximation algorithms on large graphs. *Proceedings of the 29th International Conference on Scientific and Statistical Database Management*, (6).

Altonji, J. G., Elder, T. E., and Taber, C. R. (2005). Selection on observed and unobserved variables: assessing the effectiveness of catholic schools. *Journal of Political Economy*, 113(1):151–184.

Angrist, J. D. (2014). The perils of peer effects. *Labour Economics*, 30:98–108.

Araos, C., Cea, M., Fernández, M., and Valenzuela, E. (2014). The role of school context on marijuana use in Chile: a classroom-level analysis. *Deviant Behavior*, 35:412–432.

Barbera, S. and Jackson, M. O. (2017). A model of protests, revolution, and information. *Working Paper*.

Bramoullé, Y., Djebbari, H., and Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics*, 150(1):51–55.

Brandes, U. and Pich, C. (2007). Centrality estimation in large networks. *International Journal of Bifurcation and Chaos*, 17(7).

Brock, W. A. and Durlauf, S. N. (2001). Discrete choice with social interactions. *Review of Economic Studies*, 68(2):235–260.

Bueno de Mesquita, E. (2010). Regime change and revolutionary entrepreneurs. *American Political Science Review*, 104(3):446–466.

Caeyers, B. and Fafchamps, M. (2016). Exclusion bias in the estimation of peer effects. *NBER working paper 22565*.

Cantoni, D., Yang, D. Y., Yuchtman, N., and Zhang, Y. J. (2017). Protests as strategic games: experimental evidence from Hong Kong's democracy movement. *Working Paper*.

Card, D., Mas, A., and Rothstein, J. (2008). Tipping and the dynamics of segregation. *Quarterly Journal of Economics*, 123(1):177–218.

Chetverikov, D. and Wilhem, D. (2017). Nonparametric instrumental variable estimation under monotonicity. *Econometrica*, 85(4):1303–1320.

Chwe, M. S.-Y. (2000). Communication and coordination in social networks. *Review of Economic Studies*, 67(1):1–16.

Confech (2011). Confederation of chilean students to Joaquín Lavín. Letter.

Dahl, G., Loken, K., and Mogstad, M. (2014). Peer effects in program participation. *American Economic Review*, 104(7):2049–2074.

De Giorgi, G., Pellizzari, M., and Redaelli, S. (2010). Identification of social interactions through partially overlapping peer groups. *American Economic Journal: Applied Economics*, 2(2):241–275.

Díaz, J., Grau, N., Reyes, T., and Rivera, J. (2017). The impact of grade retention on juvenile crime. *Working Paper*.

Durlauf, S. N. and Ioannides, Y. M. (2010). Social interactions. *Annual Review of Economics*, 2:451–478.

Edmond, C. (2013). Information manipulation, coordination, and regime change. *Review of Economic Studies*, 80(4):1422–1458.

Enikolopov, R., Makarin, A., and Petrova, M. (2018). Social media and protest participation: evidence from Russia. *Working Paper*.

Enikolopov, R., Makarin, A., Petrova, M., and Polishchuk, L. (2017). Social image, networks, and protest participation. *Working Paper*.

Figueroa, F. (2012). *Llegamos para quedarnos: crónicas de la revuelta estudiantil*. LOM ediciones.

Gladwell, M. (2000). *The Tipping Point: How Little Things Can Make a Big Difference*. Little, Brown and Company.

Granovetter, M. (1978). Threshold models of collective behavior. *American Journal of Sociology*, 83(6):1420–1443.

Guryan, J., Kroft, K., and Notowidigdo, M. J. (2009). Peer effects in the workplace: evidence from random groupings in professional golf tournaments. *American Economic Journal: Applied Economics*, 1(4):34–68.

Halberstam, Y. and Knight, B. (2016). Homophily, group size, and the diffusion of political information in social networks: evidence from Twitter. *Journal of Public Economics*, 143:73–88.

Hsieh, C.-T. and Urquiola, M. (2006). The effects of generalized school choice on achievement and stratification: evidence from Chile's voucher program. *Journal of Public Economics*, 90(8-9):1477–1503.

Jackson, G. (2013). *El país que soñamos*. Debate.

Jackson, M. O. (2010). *Social and Economic Networks*. Princeton University Press.

Kuran, T. (1989). Sparks and prairie fires: a theory of unanticipated political revolution. *Public Choice*, 61(1):41–74.

Kuran, T. (1991). The East European revolution of 1989: is it surprising what we were surprised? *American Economic Review, Papers and Proceedings*, 81(2):121–125.

Larson, J. M., Nagler, J., Ronen, J., and Tucker, J. A. (2017). Social networks and protest participation: evidence from 130 million twitter users. *Working Paper*.

Little, A. (2016). Communication technology and protest. *Journal of Politics*, 78(1):152–166.

Lohmann, S. (1993). A signaling model of informative and manipulative political action. *American Political Science Review*, 87(2):319–333.

Madestam, A., Shoag, D., Veuger, S., and Yanagizawa-Drott, D. (2013). Do political protest matter? Evidence from the tea party movement. *Quarterly Journal of Economics*, 128(4):1633–1685.

Manacorda, M. (2012). The cost of grade retention. *Review of Economic and Statistics*, 94(2):596–606.

Manacorda, M. and Tesei, A. (2016). Liberation technology: mobile phones and political mobilization in Africa. *Working Paper*.

Manski, C. (1993). Identification of endogenous social effects: the reflection problem. *Review of Economic Studies*, 60(3):531–542.

Marwell, G. and Oliver, P. (1993). *The critical mass in collective action: a micro-social theory*. Cambridge University Press.

McAdam, D. (1986). Recruitment to high-risk activism: the case of Freedom Summer. *American Journal of Sociology*, 92(1):64–90.

Newey, W. and Powell, J. (2003). Instrumental variable estimation of nonparametric models. *Econometrica*, 71(5):1565–1578.

Newey, W., Powell, J., and Vela, F. (1999). Nonparametric estimation of triangular simultaneous equations models. *Econometrica*, 67(3):565–603.

OECD (2013). Education at a glance 2013: OECD indicators. *OECD publishing*.

Olson, M. (1965). *The logic of collective action. Public goods and the theory of groups*. Cambridge, MA: Harvard University Press.

Oster, E. (2017). Unobservable selection and coefficient stability: theory and evidence. *Journal of Business Economics and Statistics*, pages 1–18.

Qin, B., Strömberg, D., and Wu, Y. (2017). Why does China allow freer social media? Protests versus surveillance and propaganda. *Journal of Economic Perspectives*, 31(1):117–140.

Rau, T. (2013). Modeling structural equations with endogenous regressors and heterogeneity through derivative constraints. *Quantitative Economics*, 4(1):125–148.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology*, 1(2):143–186.

Simonsen, E. (2012). *Mala educación: historia de la revolución escolar*. Debate.

Stevenson, M. (2015). Test of random assignment to peers in the face of mechanical negative correlation: an evaluation of four techniques. *Working Paper*.

Stock, J. and Yogo, M. (2005). Testing for weak instruments in linear IV regression. In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, chapter 5, pages 80–108.

Tilly, C. (1978). *From mobilization to revolution*. McGraw-Hill.

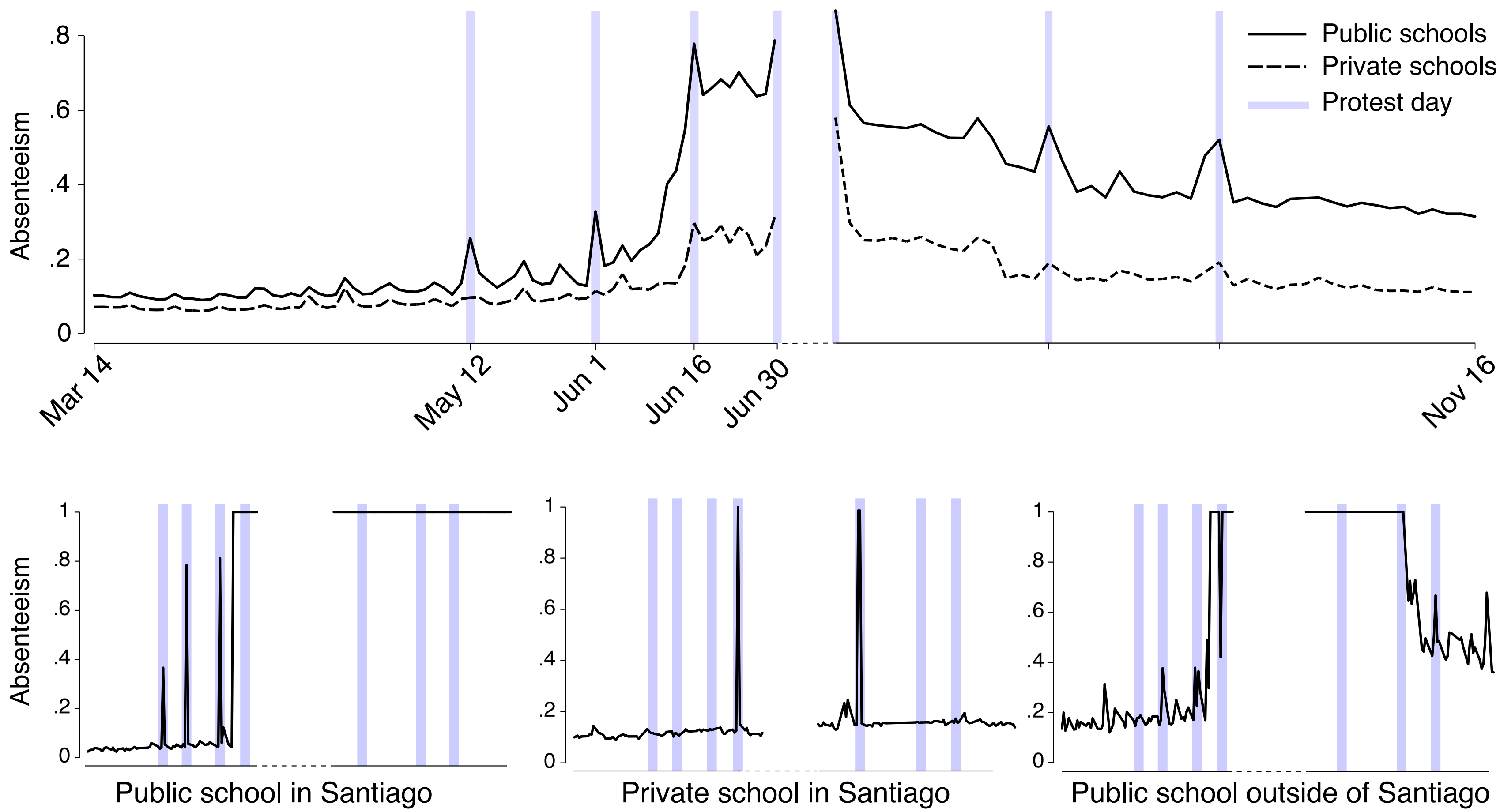
Tullock, G. (1971). The paradox of revolution. *Public Choice*, 11(1):89–99.

Vallejo, C. (2012). *Podemos cambiar el mundo*. Ocean Sur.

Yanagizawa-Drott, D. (2014). Propaganda and conflict: evidence from the Rwandan genocide. *Quarterly Journal of Economics*, 129(4):1–48.

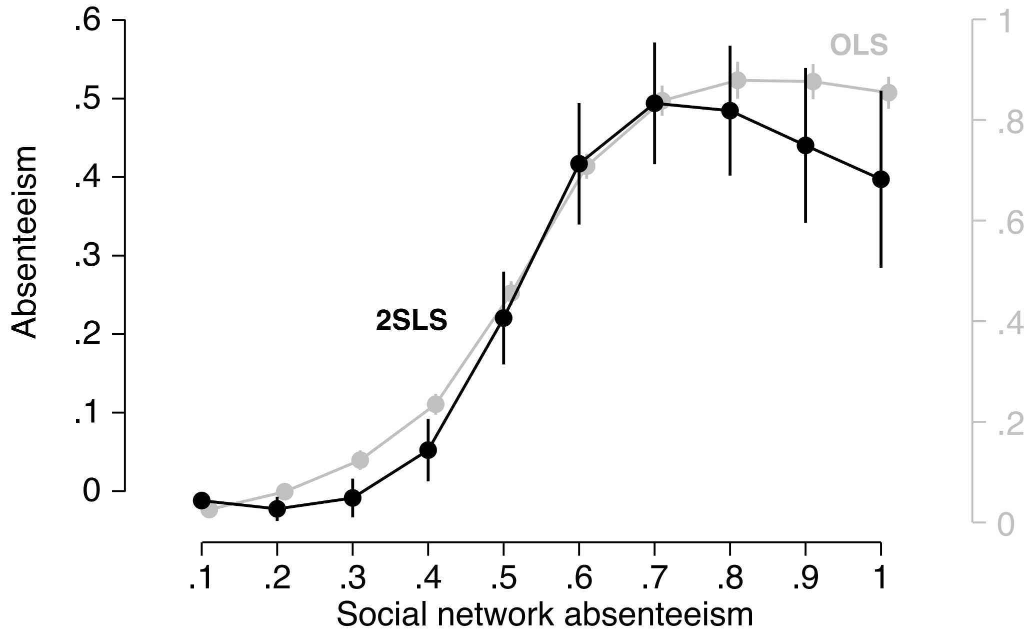
Yang, D. (2018). The impact of media censorship: evidence from a field experiment in China. *Working Paper*.

Figure 1: Absenteeism of high-school students in 2011

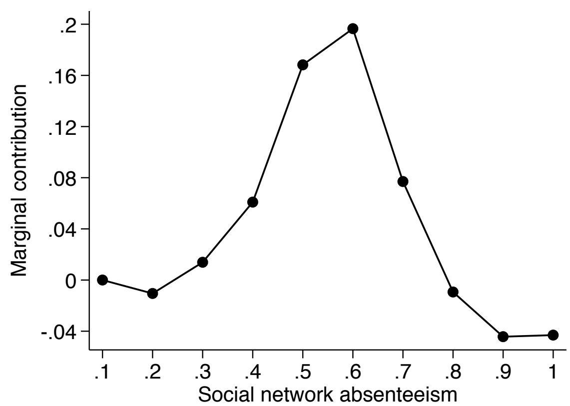


Notes: Own construction using administrative data. The y-axis is the average school absenteeism among high-school students (in percentages) and the x-axis represents days in 2011. Vertical lines denote the most important national protest days during week days (as measured by number of protesters in newspapers). The gap in the center of the figures corresponds to the winter break. More details in section 3.1.

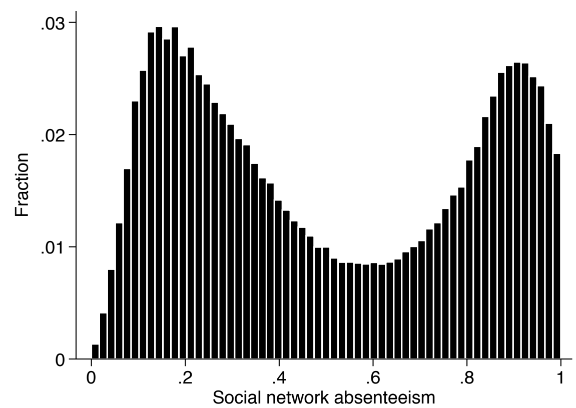
Figure 2: Main results



(a) OLS and 2SLS comparison



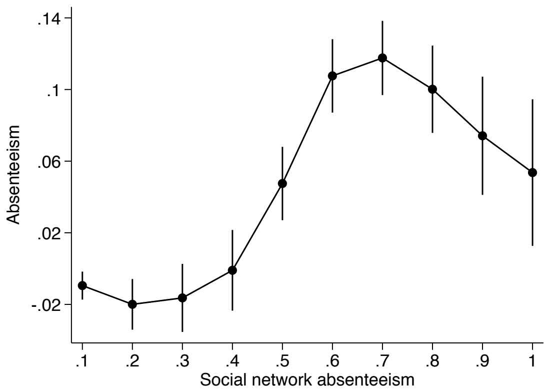
(b) Marginal contribution



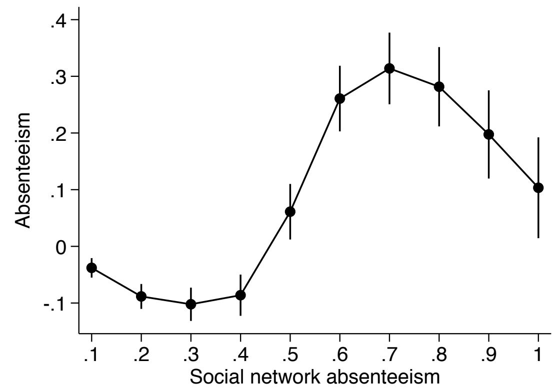
(c) Bimodal school absenteeism

Notes: Panel (a) plots estimates (OLS and 2SLS) from a regression of individual school absenteeism on 10 indicators of network absenteeism on June 16, controlling for school, network, and school characteristics, and city fixed effects (see equations 1 and 2). Vertical lines denote 95 percent confidence intervals (s.e. clustered at the city level). Panel (b) plots the difference in the estimated 2SLS coefficients in Panel (a). Panel (c) plots the empirical distribution of network absenteeism. More details in section 4.2.

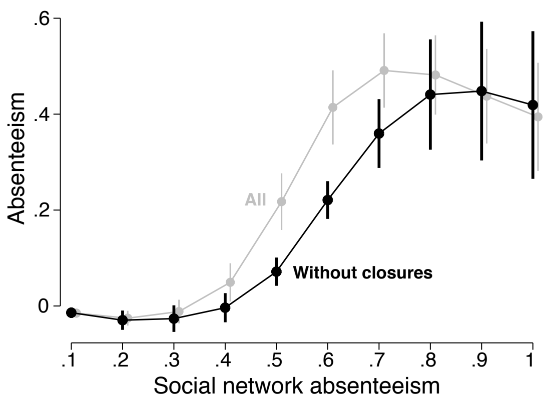
Figure 3: Additional results



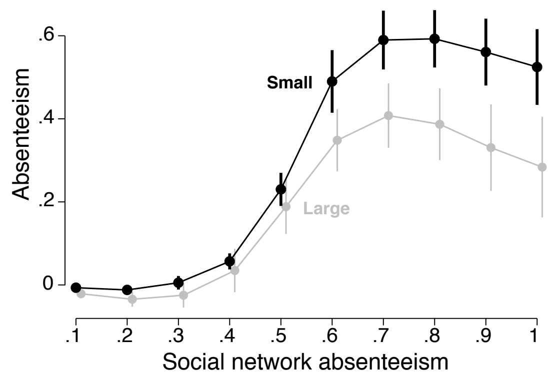
(a) Including school F.E.



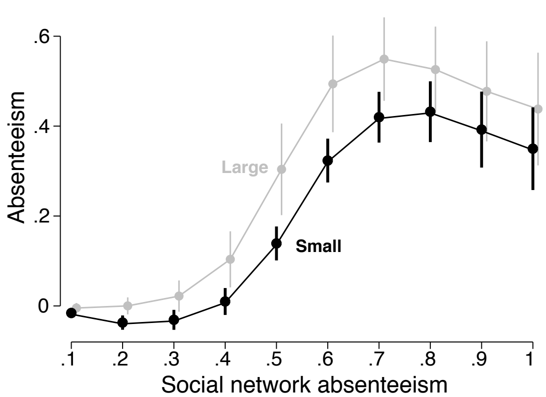
(b) Including neighborhood F.E.



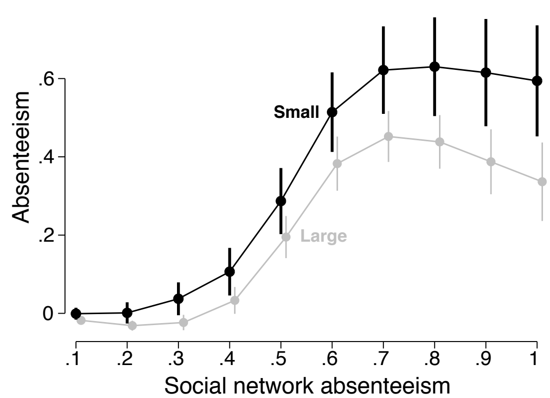
(c) School closures



(d) Network size



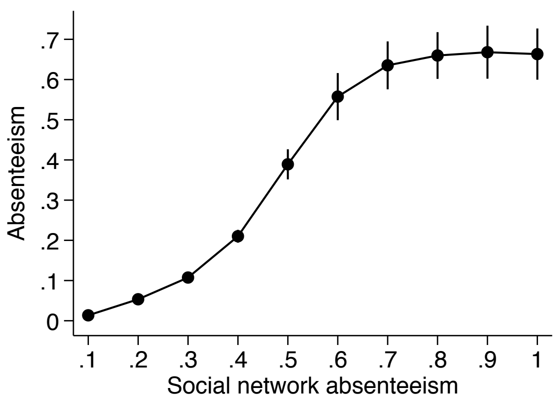
(e) School size



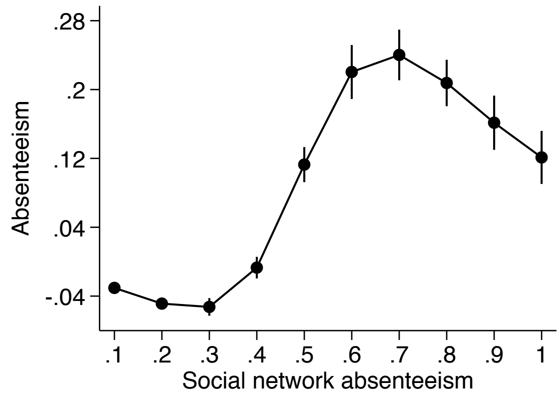
(f) City size

Notes: Panels (a) and (b) plot 2SLS estimates from a regression of individual school absenteeism on 10 indicators of network absenteeism, controlling for school, network, and school characteristics (see equations 1 and 2). Panels (c)-(f) present 2SLS estimates in sub-samples. More details in section 4.2.

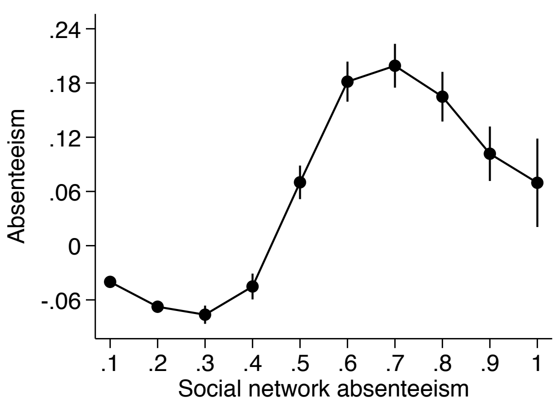
Figure 4: Two-stage panel data estimates



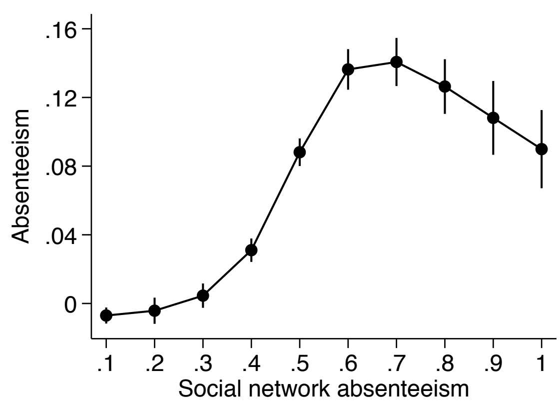
(a) Student and day F.E. (benchmark)



(b) Benchmark + city-day F.E.



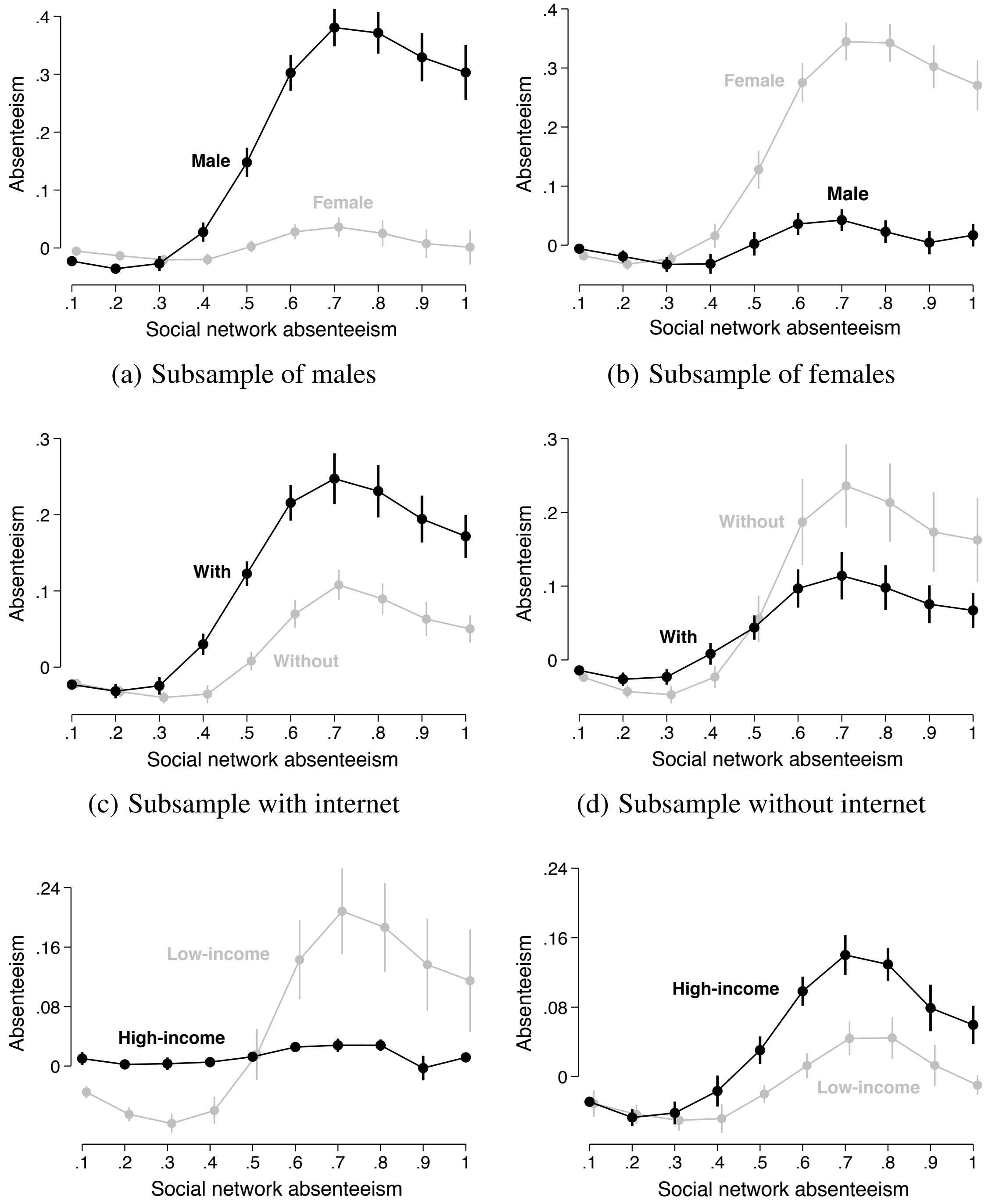
(c) Benchmark + neighborhood-day F.E.



(d) Benchmark + school-day F.E.

Notes: All regressions include fixed effects by student and protest day. Two-stage control function estimates of network effects in school absenteeism on protest days. Vertical lines denote 95 percent confidence intervals (s.e. clustered at the city level). More details in Section 4.2.2.

Figure 5: Differential influence within networks



(a) Subsample of males

(b) Subsample of females

(c) Subsample with internet

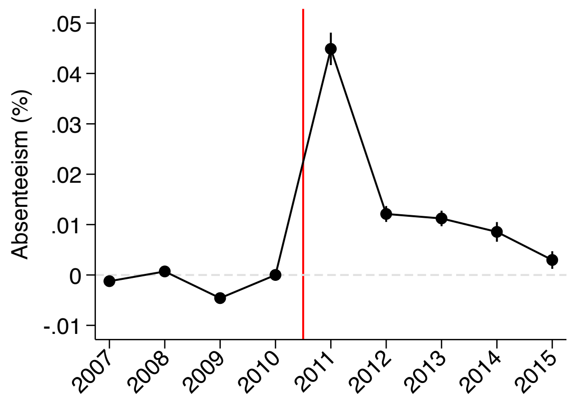
(d) Subsample without internet

(e) Low-income households

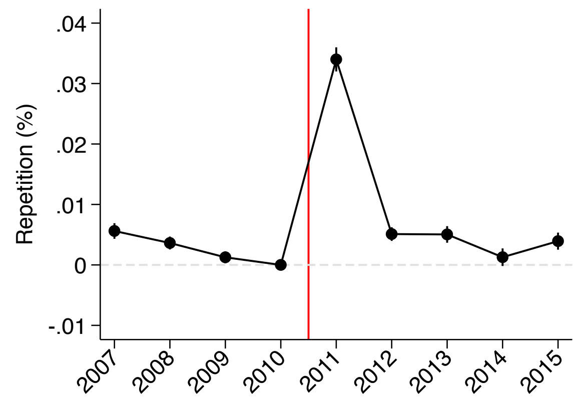
(f) High-income households

Notes: All panels plot 2SLS estimates from a regression of individual school absenteeism on 10 indicators of network absenteeism, controlling for school, network, and school characteristics, and city fixed effects. Regressions are in sub-samples and split the network in groups. More details in Section 4.3.

Figure 6: The cost of skipping school

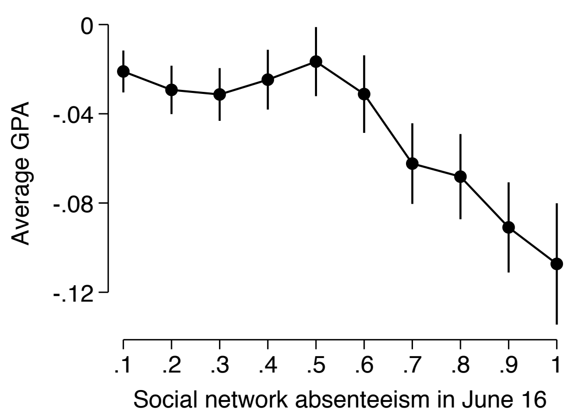


(a) Absenteeism 2002–2015

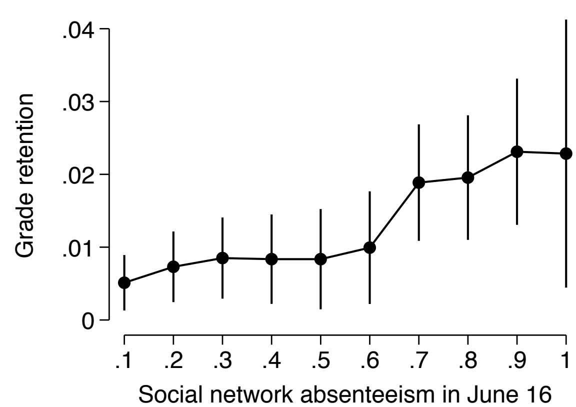


(b) Grade retention 2002–2015

Notes: Panels (a) and (b) plot differences-in-differences estimates of absenteeism/retention rates between high-school students (protesters) and primary students (non-protesters) in the period 2007-2015. Vertical lines denote 95 percent confidence intervals and standard errors are clustered at the school level. The omitted category is 2010. In both figures the *y*-axis is measured in percentage points. More details in section 5.1.



(c) GPA (N =755,874)



(d) Retention ($N = 780,306$)

Notes: Panels (c) and (d) plot OLS estimates from a regression of academic performance on network absenteeism in June 16, controlling for student controls, network controls, and school fixed effects. Vertical lines denote 95 percent confidence intervals and standard errors are clustered at the school level. The *y*-axis in Panel (c) is GPA. The standard deviation of GPA is 0.67. The *y*-axis in Panel (d) is in percentages. More details in section 5.1.

Table 1: Descriptive statistics

	Mean	St. Dev.	Min.	Max.	Observations
Students					
School absenteeism:					
May 12, 2011	0.15	0.36	0	1	760,801
June 1, 2011	0.19	0.39	0	1	
June 16, 2011	0.49	0.50	0	1	
Average in 2010	0.07	0.07	0	0.99	
Schools					
Indicator for public	0.30	0.46	0	1	2,224
Number of high-school students	342	325	1	2,835	
Cities					
High-schools in the city	7.7	44.3	1	729	290
High-school students in the city	2,623	16,134	1	267,803	

Notes: Own construction based on administrative data provided by the Ministry of Education. All variables are measured in 2011 unless otherwise stated. More details in section 3.1.

Table 2: Linear estimates

Dependent variable is absenteeism on June 16, first massive protest

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A – OLS estimates							
Network absenteeism on June 16	1.23***	1.22***	1.27***	1.21***	1.24***	1.47***	0.33***
	(0.02)	(0.02)	(0.02)	(0.04)	(0.05)	(0.03)	(0.05)
Panel B – 2SLS estimates							
Network absenteeism on June 16	0.80***	0.77***	0.69***	0.81***	0.69***	0.63***	0.07***
	(0.05)	(0.05)	(0.08)	(0.07)	(0.10)	(0.21)	(0.03)
Absenteeism before June 16	x	x	x	x	x	x	x
Other student controls		x	x	x	x	x	x
Network controls			x	x	x	x	x
School controls				x	x	x	
City fixed effects					x		
Neighborhood fixed effects						x	
School fixed effects							x
F-stat 1*st* stage	53.3	50.5	30.6	36.0	24.1	14.0	77.7
R-squared (Panel A)	0.626	0.629	0.638	0.645	0.652	0.583	0.718
Observations	779,327	779,251	771,121	760,801	760,801	49,273	771,121

Notes: "Absenteeism before June 16" includes student-level indicators for school absenteeism in May 12 and June 1. "Student controls" include academic performance, average school attendance in previous years and socioeconomic characteristics. "Network controls" include average student controls at the network level. "School controls" include indicators for publicly managed schools, average academic performance of students, and average socioeconomic characteristics of students. "Neighborhoods" are geographic areas where students live (size 10×10 blocks). See Figure A.5 for a map of cities. Neighborhood data is only available for a subset of students. See Figure A.7 for a map of neighborhoods. The instrument is past classmates of 2011 classmates. Standard errors clustered at the city level are reported in parentheses. Significance level: *** $p < 0.01$.

Table 3: The political effects of the student movement

Dependent variables are electoral outcomes

	Vote shares					
	Non traditional parties	Left wing	Right wing	Voters in population	Non traditional candidates	Total number of candidates
2012 local elections	(1)	(2)	(3)	(4)	(5)	(6)
Student movement	0.050**	-0.000	-0.044***	-0.006	0.10	0.09
	(0.025)	(0.018)	(0.013)	(0.004)	(0.12)	(0.10)
2008 local elections (placebo I)						
Δ school absenteeism 2008-2007	0.024	-0.028	0.003	0.002	0.39	0.22
	(0.020)	(0.017)	(0.010)	(0.001)	(0.37)	(0.41)
2008 local elections (placebo II)						
Student movement	-0.020	0.001	0.017	-0.002	0.09	-0.15
	(0.021)	(0.016)	(0.017)	(0.002)	(0.48)	(0.55)
Socio-economic controls	x	x	x	x	x	x
Dep. variable in previous election	x	x	x	x	x	x
Mean dep. variable (upper panel)	0.347	0.375	0.278	0.492	1.55	3.36
R-squared	0.23	0.13	0.49	0.88	0.12	0.18
Counties	324	324	324	324	324	324

Notes: Regressions are weighted by the total number of voters in 2008 (upper panel) and 2004 (lower panels). "Student movement" and "Δ school absenteeism 2008-2007" have been standardized to facilitate the interpretation of coefficients. Non-traditional parties correspond to parties that are different from the lef-wing and right-wing coalitions. The coefficients for "placebos I and II" come from separate regressions. Robust standard errors are reported in parentheses. Significance level: *** $p < 0.01$, ** $p < 0.05$.

Online Appendix

Collective Action in Networks: Evidence from the Chilean Student Movement

Felipe González

List of Tables

List of Figures

Table A.1: Linear estimates – reduced form and first stage

Dependent variable is network absenteeism (Panel A) and student absenteeism (Panel B) on June 16, 2011

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A – First stage							
Instrument	1.00***	0.96***	0.68***	0.79***	0.69***	0.62***	0.47***
	(0.14)	(0.14)	(0.12)	(0.13)	(0.14)	(0.16)	(0.05)
Panel B – Reduced form							
Instrument	0.80***	0.74***	0.47***	0.64***	0.48***	0.39*	0.03***
	(0.15)	(0.15)	(0.14)	(0.15)	(0.16)	(0.22)	(0.01)
Student controls		x	x	x	x	x	x
Network controls			x	x	x	x	x
School controls				x	x	x	
City F.E.					x		
Neighborhood F.E.						x	
School F.E.							x
F-stat 1*st* stage	53.3	50.5	30.6	36.0	24.1	14.0	77.7
Observations	779,327	779,251	771,121	760,801	760,801	49,273	771,121

Notes: "Student controls" include academic performance, average school attendance in previous days and years and socioeconomic characteristics. "Network controls" include average student controls at the network level. "School controls" include indicators for publicly managed schools, average academic performance of students, and average socioeconomic characteristics of students. "Neighborhoods" are geographic areas where students live. See Figures A.5 and A.7 for maps of cities and neighborhoods. Neighborhood data is only available for some students Standard errors clustered at the city level are reported in parentheses. Significance level: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

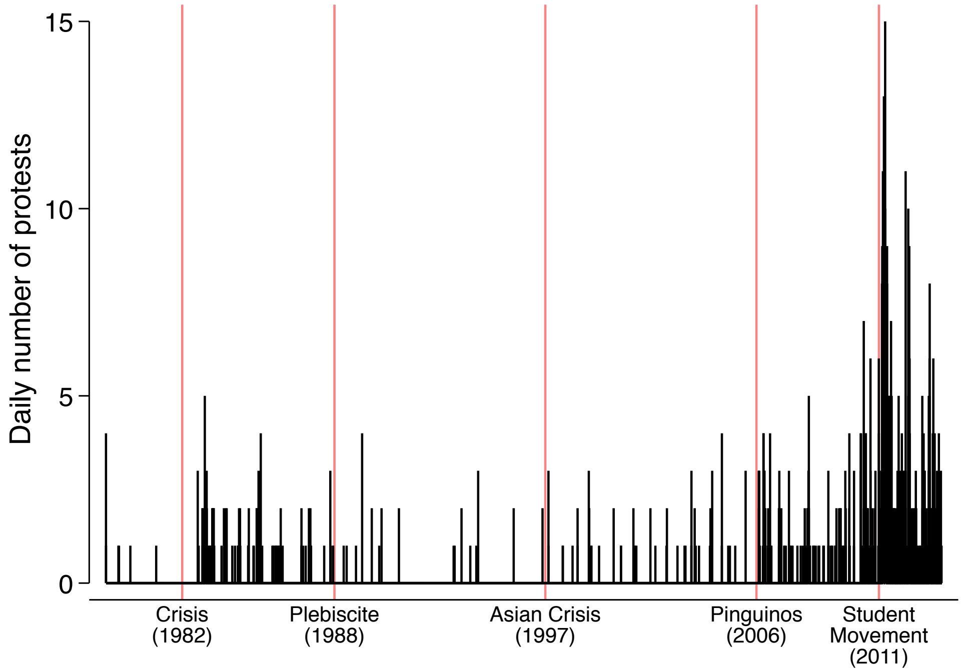
Table A.2: Two-stage least squares linear estimates – robustness to school closures

Dependent variable is student absenteeism on June 16, 2011

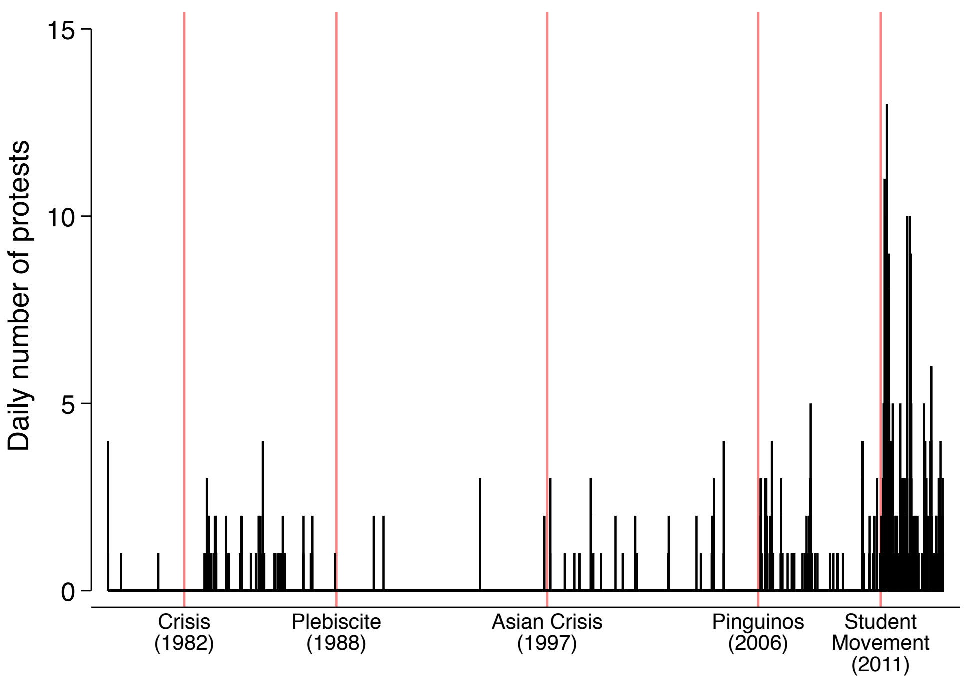
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Network absenteeism on June 16	0.66***	0.57***	0.56***	0.63***	0.53***	0.68***	0.08**
	(0.10)	(0.10)	(0.13)	(0.11)	(0.14)	(0.18)	(0.04)
Student controls		x	x	x	x	x	x
Network controls			x	x	x	x	x
School controls				x	x	x	
City F.E.					x		
Neighborhood F.E.						x	
School F.E.							x
F-stat 1*st* stage	23.7	23.7	23.8	24.7	30.2	28.5	183.3
Observations	505,643	505,610	500,834	492,903	492,902	34,145	500,798

Notes: Main estimation (two-stage least squares) of linear network effects restricting the data to schools that were not closed in June 16. "Student controls" include academic performance, average school attendance in previous days and years and socioeconomic characteristics. "Network controls" include average student controls at the network level. "School controls" include indicators for publicly managed schools, average academic performance of students, and average socioeconomic characteristics of students. "Neighborhoods" are geographic areas where students live. See Figures A.5 and A.7 for maps of cities and neighborhoods. Neighborhood data is only available for some students Standard errors clustered at the city level are reported in parentheses. Significance level: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Figure A.1: Protests in Chile 1979–2013



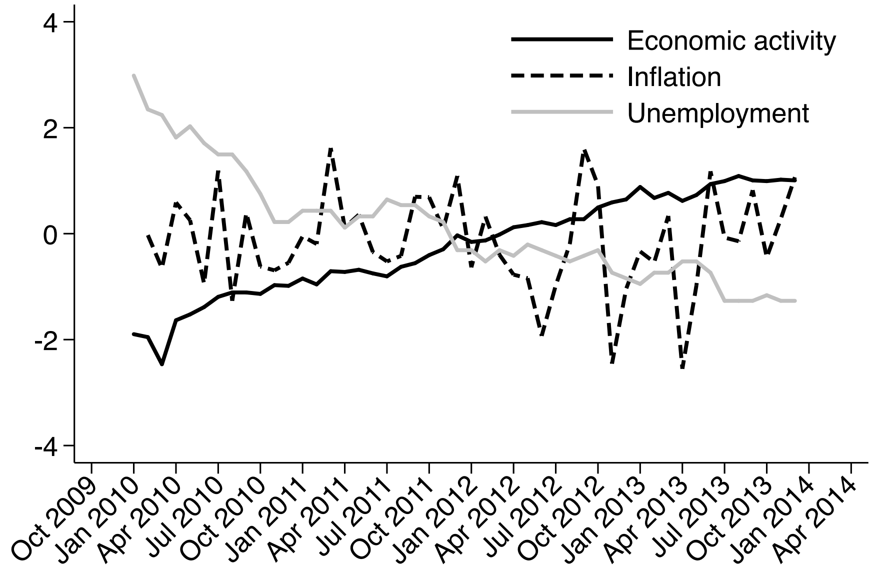
(a) Any type of protest event



(b) Protest events related to education

Notes: Data from the Global Dataset of Events, Language, and Tone.

Figure A.2: Economic indicators

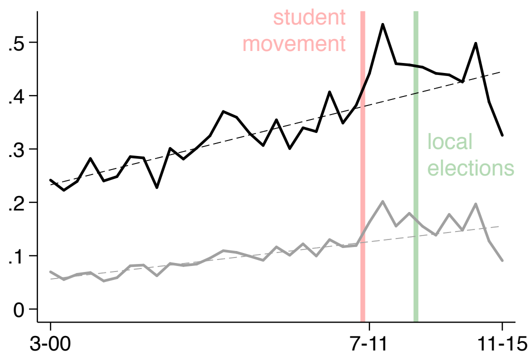


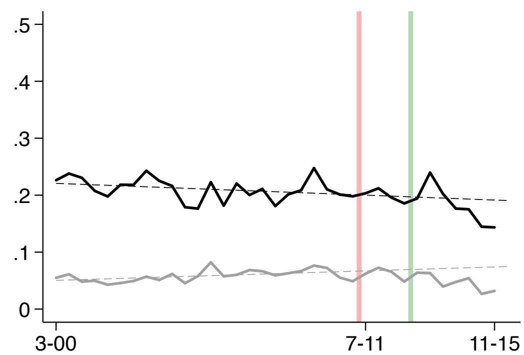
Notes: Data from the Central Bank of Chile. All variables have been normalized by substracting their average and dividing by their standard deviation in the time series. The vertical red line denotes the beginning of the student movement.

Figure A.3: Citizens' evaluation of incumbent politicians

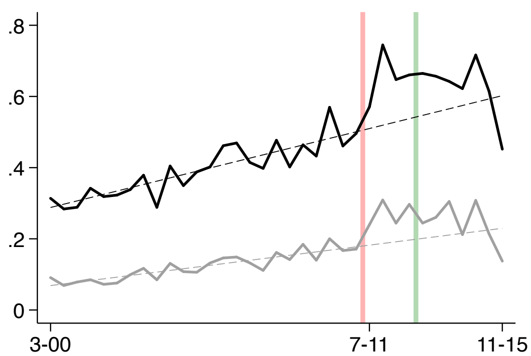
Notes: Normalized index (minus average and divide by standard deviation) for the approval of incumbent politicians. Data from the Centro de Estudios Públicos and Adimark.

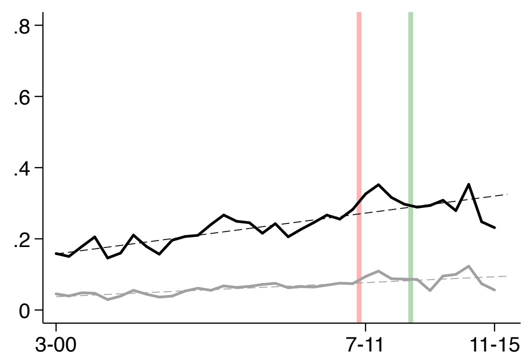
Figure A.4: Survey evidence for the impact of the student movement





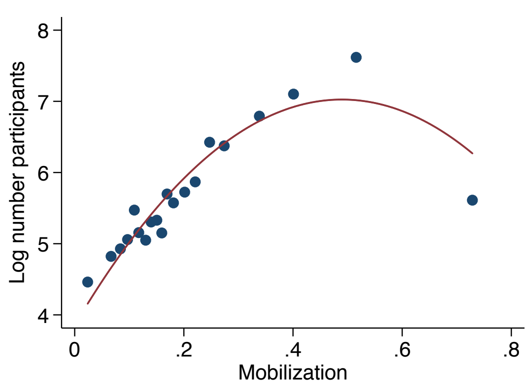
(a) Education should be priority (b) Placebo

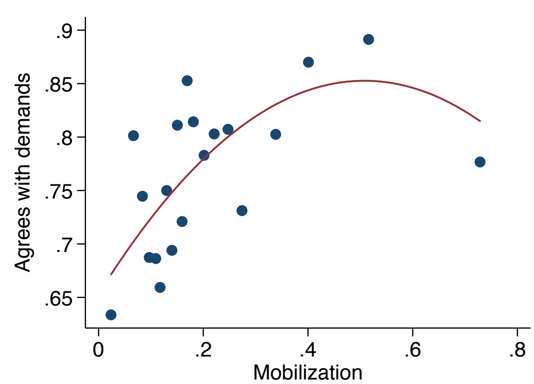




(c) Individuals 18–44 years old (d) Older than 44 years old

Notes: Panels (a)-(d) plot the percentage of people that answer the question "What should be the government's priority?" with "Education" ("Drugs" in Panel B). The gray line denotes the top 1 priority and the black line the top 3 priority.

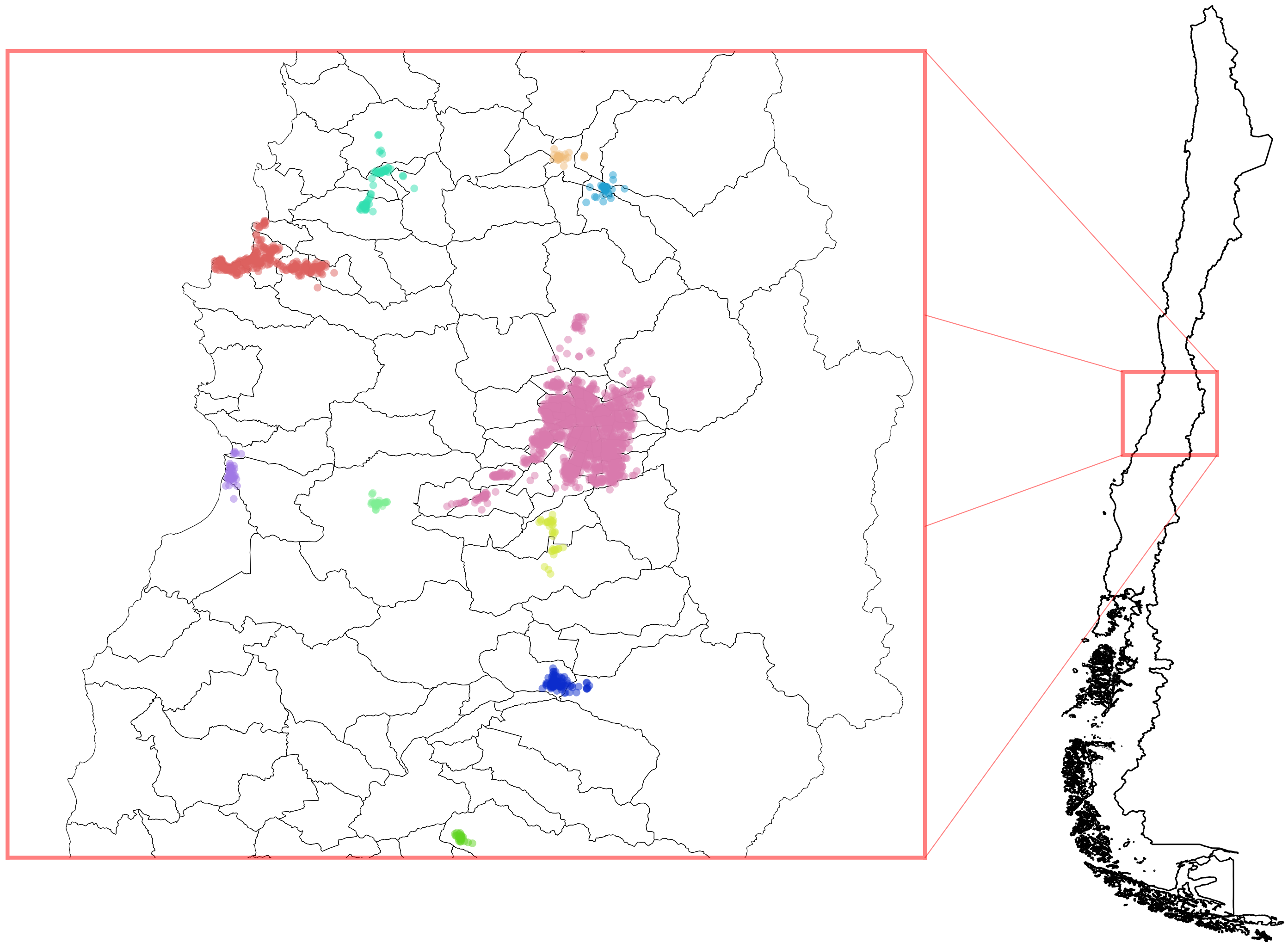




(e) Participation in plebiscite (f) Agrees with demands

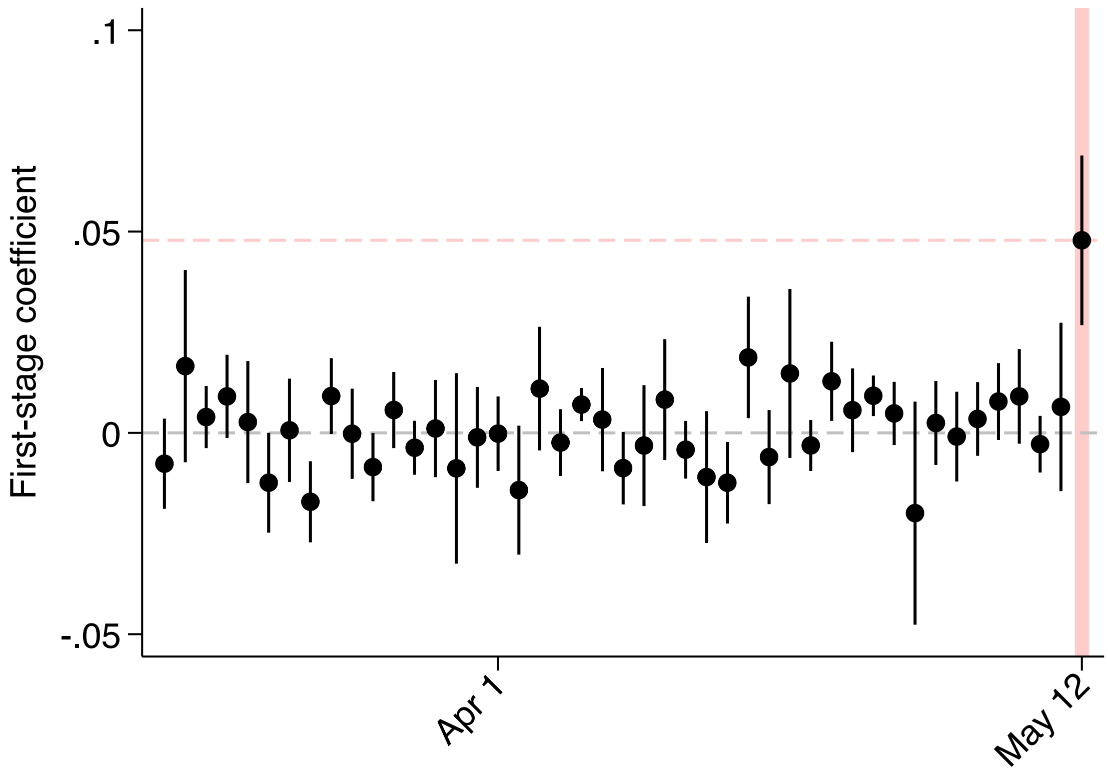
Notes: Panels (e) and (f) plot citizens' participation in the "National plebiscite for education" in October of 2011 at the county level and the percentage of people that agrees with the students' demands among those who participated.

Figure A.5: Cities



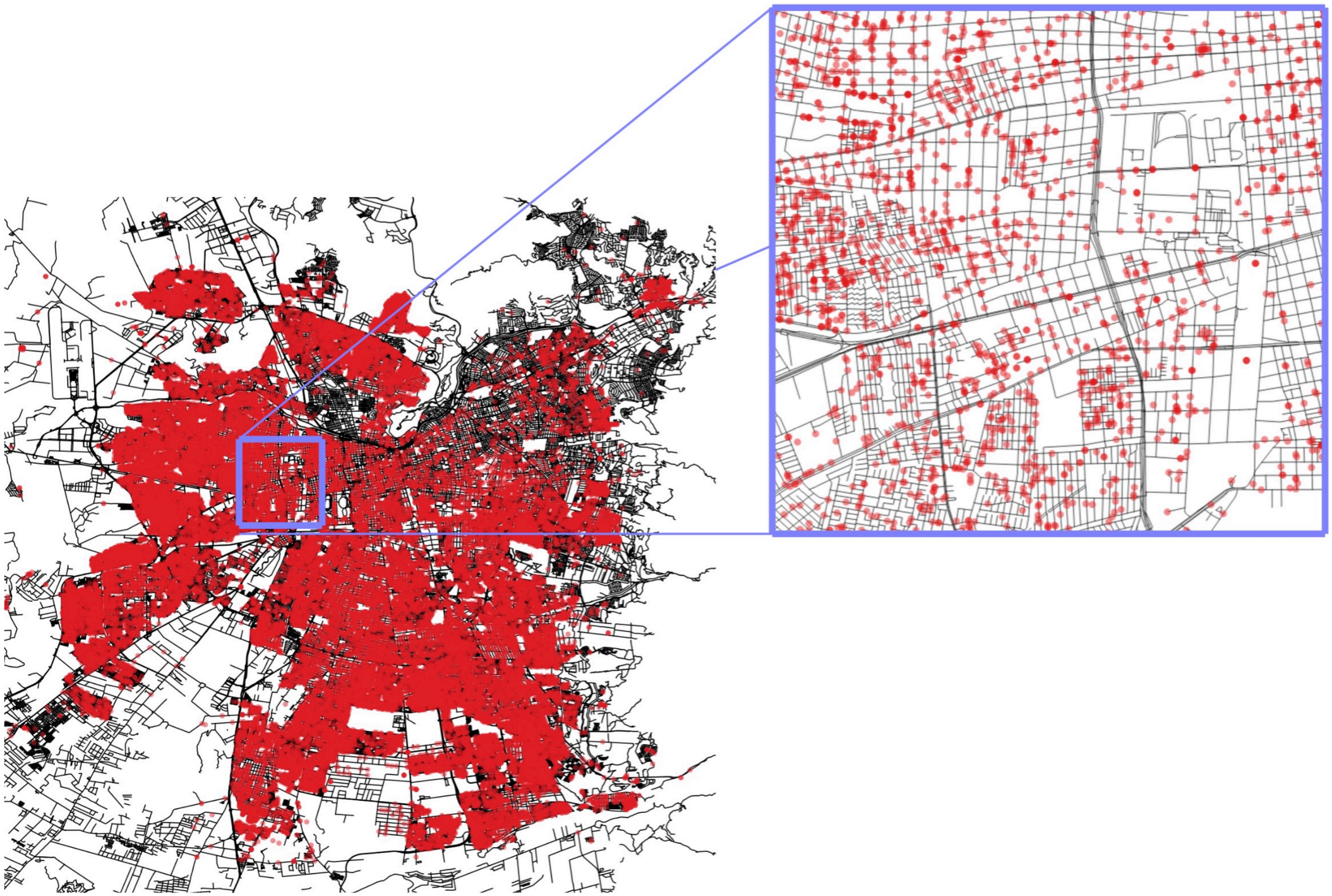
Notes: This map plots the ten largest cities in the most populated area of the country. Cities are defined as closed geographic polygons with schools closer than 5 kilometers.

Figure A.6: Placebos for first-stage



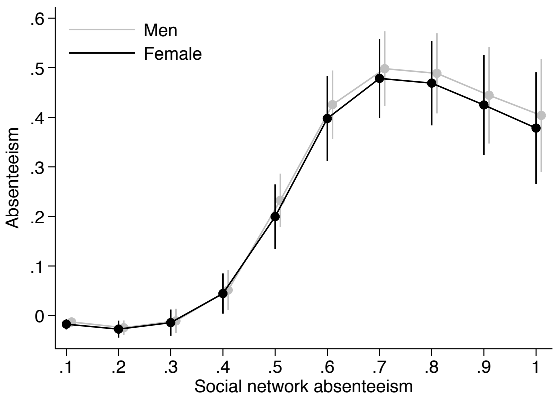
Notes: This figure plots OLS estimates from a single cross-sectional regression. The dependent variable is June 16 school absenteeism in students' social networks. The figure presents standardized coefficients for absenteeism in May 12 among out-of-school students in the "excluded network." Regression includes student absenteeism in May 12 and June 1, student controls, network controls, school controls, and city fixed effects. Vertical lines denote 95 percent confidence intervals with standard errors clustered at the city level. The coefficient highlighted in red (May 12) corresponds to the first-stage. All other coefficients are placebos for the first-stage. As expected, only 5 percent of coefficients are different from zero before May 12.

Figure A.7: Location of students in Santiago

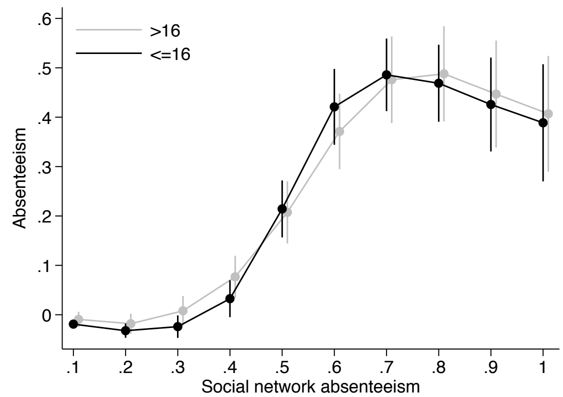


Notes: Home addresses for approximately 50,000 students in Santiago in 2011. The road network is plotted in black for geographic reference.

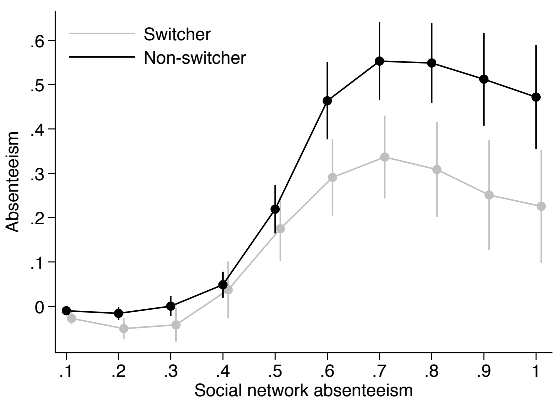
Figure A.8: Threshold model heterogeneity



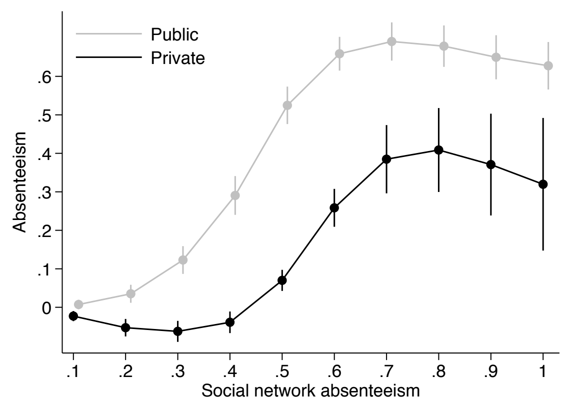
(a) Gender sub-samples



(b) Age sub-samples



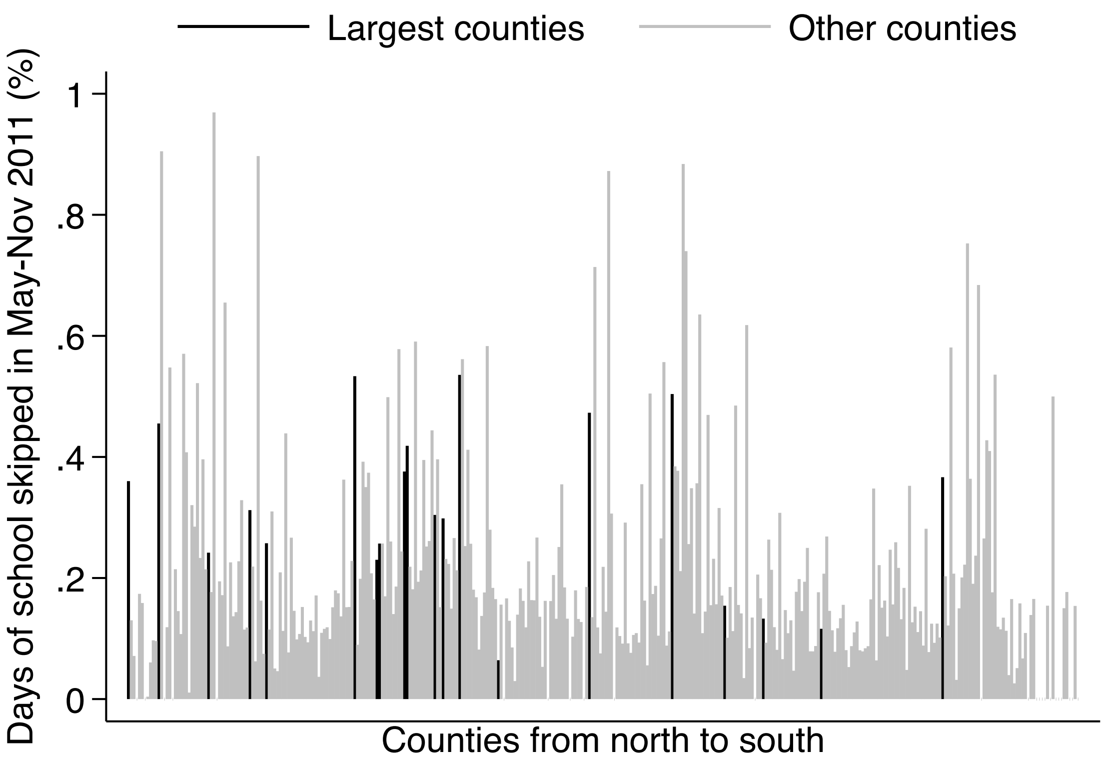
(c) Switcher sub-samples



(d) School type sub-samples

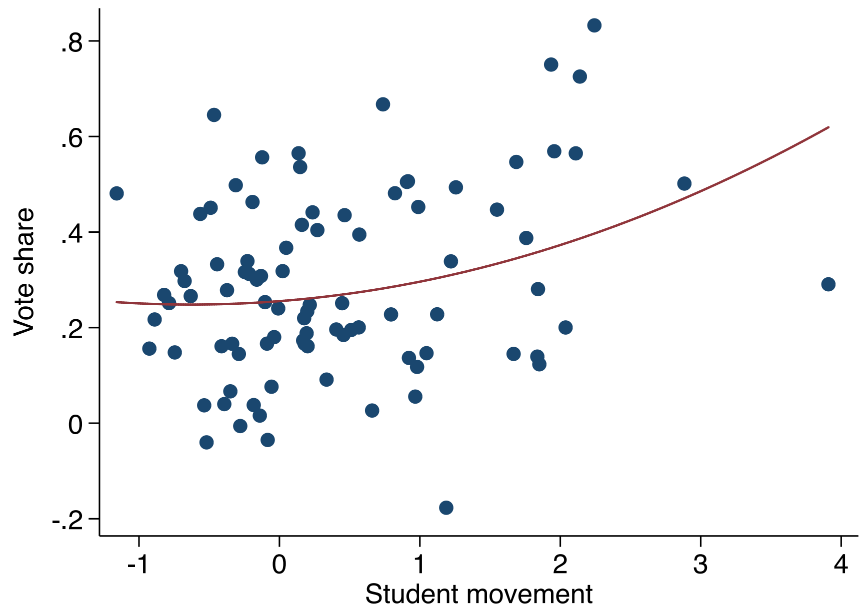
Notes: All panels plot 2SLS estimates from a regression of individual school absenteeism on 10 indicators of network absenteeism, controlling for school, network, and school characteristics in sub-samples.

Figure A.9: The intensity of the student movement by county

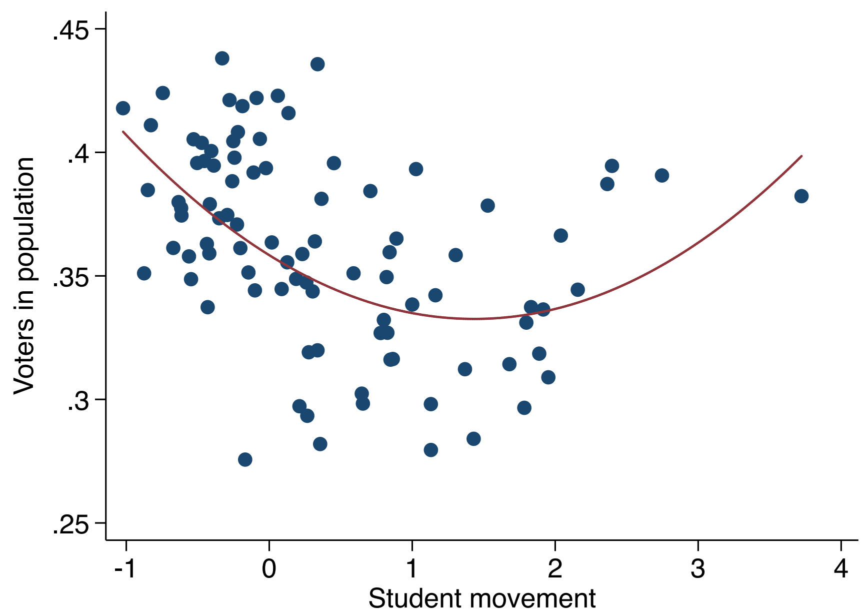


Notes: Own construction based on administrative data. Counties are ordered from north to south in the *x*-axis. The *y*-axis is defined as the percentage of additional days that high-school students skipped school between May and November 2011. There are 324 (out of 346) counties with non-zero intensity. "Large counties" are defined as counties with more than 10,000 students.

Figure A.10: The student movement and the 2012 local elections



(a) Vote share for non-traditional candidates



(b) Voters in population

Notes: This figure presents binned scatter plots and the quadratic fit of electoral outcomes in the 2012 elections (y-axis) on the intensity of the student movement in 2011 (x-axis, standardized). There are 345 counties in the country.