Collective action in networks: Evidence from the Chilean student movement*

Felipe González†

Hundreds of thousands of high-school students skipped school during the 2011 student movement in Chile to protest and reform educational institutions. Using administrative data of daily school attendance I present causal evidence of complementarities in school skipping decisions within student networks in national protest days. Identification relies on partially overlapping networks and within school exposure to an inaugural college protest. A structural estimation of a coordination game with incomplete information also supports the existence of these complementarities. Importantly, I show that skipping school imposed significant educational costs on students but it also helped to shift votes towards non-traditional candidates more aligned with their demands.

Keywords: collective action, networks, protests

*This version: May 2020. 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, Guillermo Marshall, Pablo Muñoz, Mounu Prem, Tomás Rau, Gerard Roland, Francisco Silva, Hugo Silva, Hans-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. Katia Everke, Sebastián Figari, Piera Sedini, and Cristine von Dessauer provided outstanding research assistance.

†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 body 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 are both the enormous data requirements, particularly important when studying protest behavior, and well-known identification problems (Manski, 1993; de Paula, 2013).

This paper studies the 2011 student movement in Chile, one of the largest mobilizations in the country's history, and provides causal evidence of complementarities in protest behavior. During days of national protests, hundreds of thousands of high-school students skipped school to protest with the goal of reforming the educational system. Using an administrative dataset of daily school attendance, I am able to observe the decisions of more than 500,000 high-school students on national protest days. In addition, students' social ties are mainly with classmates (Araos et al., 2014), and information about their *lifetime* history of classmates is available, allowing the construction of a countrywide network with more than 600 billion interactions across students, schools, and cities.

To guide the interpretation of results, and also to discuss the challenges imposed by the potential existence of multiple equilibria in this type of analysis (de Paula, 2013), I begin by introducing a coordination game with incomplete information. In the model students have to decide whether to skip school on a protest day, and they are perfectly informed about the network structure and the characteristics of students. When making a decision students use this public information to form an expectation about the skipping rate in their networks. The main insight comes from previous research and states that this coordination game has a *unique* equilibrium if the elasticity of network effects is lower than one (Xu, 2018). This result is crucial as it completes the econometric model and allows me to interpret the estimated parameters as network effects.

The first part of the analysis focuses on a national protest day to offer reduced-form estimates of network effects. The empirical strategy exploits the differential within school exposure of high-school students to an inaugural *college* protest and partially overlapping networks (Bramoullé et al., 2009; De Giorgi et al., 2010). The intuition behind this identification strategy is that the prede-

[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 (2020) among many others.

termined network structure made similar students, who were enrolled in the same school, to be exposed to different protest behavior in their networks for reasons unrelated to their characteristics. After providing some evidence for the validity of this variation, I use it in a non-parametric instrumental variables framework (Newey et al., 1999). The estimates constitute causal evidence of complementarities within student networks with an elasticity lower than one. Moreover, the estimates are consistent with Granovetter's (1978) "threshold model of collective behavior" in which complementarities increase markedly when the majority of students in the network skips school.[2] Moreover, despite recent evidence showing enhanced coordination in the presence of the internet (Manacorda and Tesei, 2020; Enikolopov et al., 2020), 2G-3G maps are unrelated to coordination within student networks in Chile. This result, combined with larger network effects in smaller networks, suggest that social effects (e.g. conformity) are more important than information flows.

The second part of the paper provides structural estimates of a coordination game with incomplete information based on Xu (2018). I use a two-step estimation procedure based on Hotz and Miller (1993). In the first step, I estimate the equilibrium conditional choice probabilities (CCP), which capture students' equilibrium behavior as a function of the network structure and the characteristics of students. In the second step, I estimate the primitives of the model using a discrete choice model, where a student decides whether to skip school using her beliefs about the skipping rate in her network, and the first-step CCP estimates are used to capture these beliefs. The model is estimated by maximum likelihood and results indicate the existence of complementarities within student networks with an elasticity that is lower than one.

The last part of the paper estimates the costs of skipping school and their impact on electoral outcomes. A differences-in-differences analysis among primary and high-school students in the period 2007–2015 reveals that grade repetition increased by 60 percent, from a base of 6 percent, among high-school students in 2011. Skipping school on the protest day under study – which led to higher absenteeism in the following months – decreased GPA by 0.1 standard deviations and increased grade repetition by 33 percent. In addition, I provide suggestive evidence that the student movement was able to shift votes towards political candidates more aligned with their 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, crowding-out mostly traditional right-wing candidates.

[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.

This paper contributes to the empirical understanding of participation in collective action. Only a few recent papers have studied *protest* behavior, and the role of social effects has been relatively overlooked.[3] Notable exceptions include Cantoni et al. (2019) and Hager et al. (2019), who show that beliefs about others' turnout to a protest affect participation in the context of recent rallies in Hong-Kong and Germany; Enikolopov et al. (2018) who show social image was an important motivation to participate in the 2011-2012 protests in Russia; and Larson et al. (2019) who use Twitter data to show that network position influenced attendance at the 2015 Charlie Hebdo protests in Paris. Recent research has also identified persistence in protest behavior mediated by social interactions (Bursztyn et al., 2020) and has found important interdependencies in other political behaviors (Fujiwara et al., 2016; Coppock and Green, 2016; Perez-Truglia and Cruces, 2017; Hensel et al., 2019). In contrast, my paper focuses on individual-specific networks and uses administrative data to test for complementarities in protest behavior.[4]

An important branch of the previous literature focuses on the role played by information communication technologies (ICT). One part of this research provides estimates of the aggregate contribution of ICT – such as mobile phones and social media – to the formation and spread of protests across cities (Manacorda and Tesei, 2020; Enikolopov et al., 2020). Authors have argued that ICT can help to spread grievances, particularly in times of difficult economic conditions, and enhance the ability of citizens to coordinate for a collective action such as a protest. Another part of this literature uses disaggregated data to estimate how political participation and information spreads through social networks (e.g. Twitter) depending on the network centrality of individuals (Halberstam and Knight, 2016; Larson et al., 2019). Related research has also estimated the direct impact of media censorship on political engagement (Yang, 2019) and ICT surveillance as a method to prevent the propagation of protests (Qin et al., 2017). In contrast to most research, this paper shows that coordination within student networks is similar in places *without* 2G-3G coverage, perhaps suggesting that ICT has little influence in offline networks with strong ties.

This paper also speaks to a literature estimating the impact of protests.[5] Madestam et al. (2013) uses rainfall shocks as exogenous variation affecting the number of protesters in the Tea

[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.

[4]Jackson and Storms (2019) also studies behavior in networks to identify communities using a structural approach. More generally, there is a large literature studying social interactions. See Durlauf and Ioannides (2010); Blume et al. (2010); de Paula (2013); Graham (2015); de Paula (2017) for important reviews. In relation to this literature, this paper is one of the first to show the existence of protest complementarities in partially overlapping networks.

[5]There is, of course, a large theoretical literature studying social unrest and political transformation. See, for example, Acemoglu and Robinson (2000) and Passarelli and Tabellini (2017).

Party movement across U.S. counties to show how the movement affected electoral outcomes and policies, while Larreboure and González (2020) uses a similar method to show that the Women's March in the U.S. empowered women and ethnic minorities. Aidt and Franck (2015) shows that the Swing riots in early 19th century Britain – credible signals of the threat of a revolution – facilitated democratic reforms. Finally, a recent literature also shows that protests can change public support for policies (Enos et al., 2019), racial attitudes (Mazumder, 2019a,b), and social norms such as reporting of sex crimes (Levy and Mattsson, 2019). 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.

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.[6] 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.[7]

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 express-

[6] 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).

[7] 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.

ing their discontent with the announcements (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 after the rise of the movement (Figure A.5). 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 movement, including the beginning of formal negotiations, the focus of media on violent protesters, and students' concerns about grade retention.[8]

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.6). 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. There were also changes to laws that regulate state guaranteed loans – used by most students to attend universities and technical schools – including a reduction in the interest rate paid by students, an increase in coverage, and caps to monthly repayments. In addition, the left-

[8]"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.

wing candidate Michelle Bachelet won the 2013 election with a platform that offered free tertiary education. Although with changes, this policy has been implemented gradually in the last years.

3 Theoretical framework

3.1 A coordination game with incomplete information

Consider the following setup based on Xu (2018), who studies college attendance decisions in a large network in the United States. There are N high-school students indexed by i and connected by a static network. The links in the network are public information, $j(i)$ represents the set of high-school students in i's friendship network, and x_i represents publicly known characteristics. Students choose simultaneously whether to skip school in a national protest day, decision denoted by $k \in A = \{0, 1\}$. Before choosing, student i observes a private information shock $\epsilon_i \equiv (\epsilon_{i0}, \epsilon_{i1})$. Therefore, the utility of student i from choosing $k \in A$ is given by:

$$U_{ik} = \beta_k(x_i) + \sum_{h \in j(i)} [f_k(A_h) + \gamma_k(x_h)] + \epsilon_{ik} \tag{1}$$

The unknown functions are f_k, β_k, γ_k with $k \in A$. The function f_k represents the impact of student i's network on i's decision i.e. network or social effects.

A strategy for student i is a mapping from his private shock and all the public information (i.e., the network structure and observable student characteristics) to an action. I use the notion of a Bayesian-Nash equilibrium, where the strategy of every student maximizes her expected utility given the strategies of all other students.[9] As argued in Xu (2018), the existence of a Bayesian-Nash equilibrium follows from Brouwer's fixed point theorem.

A concern with this setup is the possibility of multiple equilibria as is standard in coordination games (de Paula, 2013). Crucially, Xu (2018) provides conditions for equilibrium uniqueness, which depend on the magnitude of the social effects and the distribution of private shocks. For example, when the private shocks ϵ_{ik} are i.i.d. across actions and students and follow an extreme value distribution with density function $f(t) = \exp(-t)\exp[-\exp(-t)]$. Then, the restriction on the

[9] The uncertainty arises because student i does not observe the private shocks of other students, so student i must form expectations about their behavior using the distribution of private shocks.

magnitude of social effects that guarantees equilibrium uniqueness is:

$$\max_{k,m,\ell \in A} |f_k(\ell) - f_m(\ell)| \quad < \quad \frac{3}{2} \tag{2}$$

Intuitively, and given the decision is binary, this restriction means that if we take the maximum influences for each student, this number is bounded above. The bound is "similar to the requirement that all roots lie outside of the unit circle in spatial autoregressive models" and it implies weak dependence (Xu, 2018, p. 262). In my setting this assumption means that if the average protest participation among students in the network increases by one percentage point, then student i's probability of participation has to change less than one percentage point for the equilibrium to be unique. Unfortunately, it is not possible to test for the validity of this assumption. At the same time, I will show that the estimated elasticity of individual decisions with respect to network decisions is indeed lower than one, which is consistent with the key assumption of the model.[10]

3.2 *Discussion of mechanisms*

Several mechanisms can explain the existence of network considerations. On the benefits side, students might derive utility from having shared experiences with their networks. Then, when some people in their networks decides to attend a protest, they are pushed to protest as well to share that experience. This could also be the case if conversations in the network are grounded on past experiences and students derive benefits from joining the conversation, as suggested by recent evidence (Gilchrist and Sands, 2016). Similarly, students might update their information based on the actions of the network and this might cause a change in behavior. This could be the case if students assign a probability to the protest's success based on how many people in their social networks participate, as in theoretical models emphasizing the importance of group size in intergroup conflict (Blattman and Miguel, 2010). If many are protesting, then a student might believe the protest is likely to achieve change. In the case of the Chilean student movement, the success is associated to a higher probability of cheaper tertiary education.

Alternatively, the individual cost of protesting could be a function of network participants. There are two relevant cases in the context of student protests. First, a punishment from the network

[10]Besides showing the uniqueness of the equilibrium, Xu (2018) also demonstrates that this Bayes-Nash equilibrium satisfies a *Network Decaying Dependence* condition (Lemma 2), which means that social effects decay with network distance. This is, student i is affected by her first-degree friends more than she is affected by her second-degree friends (friends of friends who are not her friends). This result relates directly to the identification assumption in the empirical strategy, which restricts the direct social effects of second-degree friends to be zero.

for deviating from the social norm. Examples of punishments are shaming or feelings of guilt (Elster, 1989; Enikolopov et al., 2018). Second, an action from the school such as teachers less likely to teach the syllabus in the absence of many students, or teachers deliberately punishing protesters. In my context a deviation from the social norm is more likely to be relevant because I study a few protest days at the very beginning of the movement. Students presumably perceived the cost of missing a class to be low because it was difficult to anticipate that skipping school in the beginning would lead to significantly more absenteeism in the remaining of the academic year.

3.3 Extensions and predictions

The shape of the influence function $f(\cdot)$ speaks to an important literature studying individual decisions. Inspired by Schelling's (1971) tipping model, Granovetter (1978) proposed a theory of binary decisions based on thresholds. In particular, he argued that a person's decision can be influenced by the decision of others, but particularly so if there is a "critical mass" making a certain choice, e.g. a person might participate in a protest only if more than $x\%$ of others are also participating.[11] Two aspects of Granovetter's work are important to highlight for this research. First, the decision maker might care about the actions of everyone else or she might only care about the decisions of a "reference group." Empirically I follow Xu (2018) and estimate the response of a student to the decisions in her social network, and I flexibly control for decisions at more aggregate levels (e.g. schools and cities). Third, there might be multiple explanations behind the threshold behavior and, although empirically I cannot fully distinguish between these, I will use auxiliary results to discuss which one could be important.

All in all, Xu's (2018) framework coupled with the insights from previous research delivers three predictions. First, an increase (decrease) in the benefits (costs) of skipping school on a protest day – both coming from the actions of the network – means that the individual probability of skipping school should be a function of absenteeism in students' networks. Second, if smaller groups of students can coordinate more easily, as suggested by Olson (1965), then network effects should be easier to observe in smaller groups.[12] Third, if the internet improves coordination then social effects should also be easier to observe in student networks more connected to the internet,

[11]In this case $x\%$ represents the threshold. In the original model each person has a different threshold and the distribution of thresholds is exogenous. The canonical explanation is a non-linear decreased in the cost of apprehension when many people participate in a riot. But non-linear benefits are also theoretically possible.

[12]"In any event, size is one of the determining factors in deciding whether or not it is possible that the voluntary, rational pursuit of individual interest will bring forth group-oriented behavior. Small groups will further their common interests better than large groups." Olson (1965, p. 52).

as recent evidence suggests (Manacorda and Tesei, 2020).

4 Data

4.1 Daily school absenteeism and student networks

The analysis uses four administrative datasets. The first 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 (school, grade, and classroom) in 2011 and previous years. There were approximately 975,000 high school students enrolled in 2,700 high-schools in 2011. However, after restricting attention to students with all covariates this number decreases to 760,000. Moreover, when I focus on schools with daily attendance available for the June 16 protest there are 500,000 students in 1,700 high-schools. The third measures students' annual academic performance, i.e. GPA. The last dataset describes schools. In the final data approximately 20 percent of students were enrolled in public schools and 80 percent in private schools. School addresses are also available and I use these to construct geographic clusters that I refer to as "cities." There are more than 200 cities in the final data, with 7 high schools and 2,000 high-school students in the average city.[13] Table 1 presents descriptive statistics for the sample of students in high-schools opened in June 16 (column 1) and for all students for comparison (column 3).

Student 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, including the current ones. As shown in Figure A.8, past classmates in $j(i)$ live in i's neighborhood, hence their interactions are likely to remain. As of 2011, each high-school student had a unique set of past classmates that I identify from enrollment information in previous years. 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. 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

[13] 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. These can also be thought as clusters of counties that approximate conurbations. Figure A.7 presents a map of cities.

2011.[14] Importantly, 88% of private schools in 2011 had students who attended a public schools in previous years and hence public and private schools are highly connected in this network.

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 daily absenteeism in 2011 for three schools.

4.2 High-school absenteeism and rally attendance

How does high-school absenteeism in a protest day relates to protest participation? By far the most common definition of protest participation is rally attendance (Fisher et al., 2019), and the usual location for this rally is a city's main square. In the case of Chilean protests, however, this definition is incomplete as there were also protest activities (different from the rally) taking place inside and around schools. It is important to keep in mind these other activities when interpreting the following calculations that relate high-school absenteeism and rally attendance.

To estimate how many high-school students were at a rally I proceed in three steps. First, using data for the June 16 rally – the main protest day to be studied in the following sections – and comparing this day to a "business-as-usual" (non-protest) day, I can tell that there were approximately 100,000 additional high-school students skipping school in Santiago. Second, I gathered the reported number of people attending the June 16 rally in Santiago from different sources, including police reports and data collected by organizers, and use the average of estimates as the best guess for rally size: 87,500 people.[15] Third, I use a sample of 24 images taken from a 13-minute video of the June 16 rally, together with a crowd counting method, to calculate the

[14]For computational reasons I only consider classmates in years 2007-2011. The calculation of student-specific network variables takes substantially more computational time when including more years. In addition, the network is unfortunately 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).

[15]The four estimates I was able to collect are: two of 100,000 people, one of 80,000, and one of 70,000. Similar methods to calculate the size of rallies can be found in Acemoglu et al. (2018) and Enikolopov et al. (2020).

percentage of high-school students in the rally. Those in charge of identifying high-school students in the "crowd" were fifth and sixth year university students in 2019 – i.e. high-school students at the time of these protests – who performed the image analysis task a total of 520 times. High schoolers are identifiable because most of them wore school uniforms during the protest. This crowd-counting method delivers that 25% of people at the rally were high-school students.[16]

The previous calculations imply that 22% of students who skipped school attended the rally in Santiago's main square ($[0.25 \times 87,500]/100,000 = 0.22$). The remaining high-school students either protested in a different way or stayed at home. In what follows I use these numbers to provide some intuition for the relationship between network absenteeism and protest behavior.

5 Reduced-form analysis

5.1 Empirical strategy

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:

$$A_{isc} = f(A_{j(i)}) + g_1(x_i) + g_2(x_{j(i)}) + \zeta_s + \epsilon_{isc} \tag{3}$$

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, ζ_s is a full set of school fixed effects, and ϵ_{isc} is an error term clustered by city. The vector x_i includes 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 fully saturated bins for all controls. Student controls also include school absenteeism on previous (and smaller) protest days, i.e. May 12 and June 1.

The first part of the analysis uses a linear-in-means function f, i.e. the average absenteeism

[16] The link to the video is here. Six of the 24 images were selected at random, six were selected form the longest shot to maximize crowd flow, six from the two largest shots, and six were taken at random from the set of large shots. The 25% number may seem low but recall that many protesters were university students, workers from labor unions, and other citizens. In addition, this is likely to be a lower bound as some students might not have been wearing uniforms during the rally. Figure A.10 provides more details about the implementation of this method.

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:

$$f(A_{j(i)}) = \beta_1 \cdot 1\left[\overline{A}_{j(i)} \in [0.1, 0.2)\right] + \cdots + \beta_9 \cdot 1\left[\overline{A}_{j(i)} \in [0.9, 1)\right] + \beta_{10} \cdot 1\left[\overline{A}_{j(i)} = 1\right] \tag{4}$$

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 (3) 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. 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 $j(j(i))$ skipped school on May 12, with $i \notin j(j(i))$. Students in the set $j(j(i))$ 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 $j(j(i))$ are attending a different school than i in 2011. Let $s(i)$ denote the set of students attending the same school than i in 2011 and let A/B denote agents in A who are not in B. Then, the identification assumption is that school absenteeism on May 12 among students in the set $j(j(i))/s(i)$ only affects student i's absenteeism on June 16 through the absenteeism of $j(i)$.

A final remark regarding the identification of $f(\cdot)$ in equation (3) is necessary. As emphasized by Newey and Powell (2003), when the endogenous variable and the instrument are continuous the conditions for identification exist, but these are stronger than in linear models. Specification decisions are particularly important. The relative flexibility of bins in equation (4) and the use of Newey and Powell's (2003) series approximations help in this regard. However, the use of bins could also be problematic because it entails some discretization of the endogenous variable. Fortunately, as emphasized by Horowitz (2011) the continuous nature of the instrument helps to identify the parameters. Intuitively, it is crucial that the variation in the instrument shifts the endogenous variable from one bin to the other across the entire distribution. The lower panel in Table 2 shows that the first stage is strong, with coefficients having the expected positive sign – higher exposure to initial protests fosters future absenteeism – and corresponding F-stats that are always far from a weak instrument problem (Stock and Yogo, 2005). Reassuringly, the value of the instrument *before* the first protest (May 12) does not predict networks' absenteeism on June 16, suggesting that unobservables that affect absenteeism in non-protest days are unlikely to affect results. The first-stage is strong and absenteeism in social networks also varies significantly for different values of the instrument in the two cases (see Figure A.9 for details).

5.2 Reduced-form results

I begin by briefly describing linear estimates of equation (3). Table 2-A presents OLS estimates, panel B 2SLS estimates, and panel D the reduced forms.[17] Column 1 uses school fixed effects, and columns 2-4 progressively include controls for student and network characteristics. Column 4 is my preferred specification. Coefficients estimated using 2SLS are positive and smaller than their OLS counterparts. In terms of magnitude, column 4 suggests that a one standard deviation

[17]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. Table 2 also shows that standard errors are virtually unchanged when using the 53 provinces as clusters instead of the 240 cities.

in network absenteeism is associated to an increase of 2 p.p. in individual school absenteeism (0.20×0.09), a 10% increase over the mean. Put differently, when all members of a student's network skip school, then the student's probability of also skipping increases by 20 p.p. Hence, these linear estimates suggest that protest decisions are strategic complements within student networks.

Let me now discuss estimates of equation (3), now using the functional form in equation (4) and the approach proposed by Newey et al. (1999). This estimation corresponds to a control function, and the coefficients of interest are associated with indicators for different values of absenteeism in the network. Importantly, in the following regressions the coefficient associated with the control function parameter is always statistically different from zero, as expected given the difference between 2SLS and OLS estimates. Panel A in Figure 2 presents OLS estimates of $\left(\widehat{\beta}_1, \ldots, \widehat{\beta}_{10}\right)$ and panel B presents 2SLS estimates. The exact specification corresponds to that in Table 2 column 4. As before, 2SLS estimates are lower than their OLS counterparts.

The 2SLS estimates are consistent with Granovetter's (1978) threshold model of collective behavior 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-C plots the sequential difference between estimated coefficients, i.e. $\widehat{\beta}_k - \widehat{\beta}_{k-1}$ with $k = 1, \ldots, 10$, where $\beta_0 = 0$, and $\widehat{\beta}_1, \ldots, \widehat{\beta}_{10}$ correspond to the estimated coefficients in equation (4). This figure suggests that the influence of networks on individual decisions is positive only after absenteeism reaches 50 percent of a network, and reaches a maximum around the 60-70 percent mark.[18]

The difference between OLS and estimates 2SLS could be explained by the characteristics of the compliers, but also by other factors such as the reflection problem or some omitted variable. Regardless, a characterization of the compliers is always helpful. Using Abadie et al.'s (2002) method I calculate that the compliers are more likely to attend private schools but have similar levels of school absenteeism in their networks. Table A.1 presents the full characterization of compliers. If students in private schools are less likely to respond to social effects than students in public schools, then the characteristics of the compliers contribute to 2SLS estimates being smaller than OLS estimates. In addition, the difference between OLS and 2SLS estimates in the

[18] Additional coefficients might be of interest for the reader. School absenteeism in previous (smaller) protest days (May 12 and June 1) are highly predictive of school absenteeism on June 16 (coef. 0.09 and 0.12, p-values <0.01) and students of high-academic achievement are less likely to protest. In addition, students linked to students of high-academic achievement and low historical absenteeism are more likely to skip school on June 16.

linear and non-linear models can be explained by linear estimates placing relative more weight on observations in the median of the distribution (Yitzhaki, 1996).

How was *network* absenteeism related to the number of protesters in main square? On one side, section 4.2 showed that one of every four students who skipped school on June 16 was in Santiago's main square. On the other side, this section showed that when network absenteeism was close to 100% the individual probability of skipping school increased by 20 p.p. By combining both sets of numbers we conclude that 20 networks with full absenteeism created one additional protester in main square (i.e. $20 \times 0.20 = 4$). If we recall that the average network was composed by 80 students – and assume non-overlapping networks as a bound – then 1,600 students skipping school induced one additional student to attend the rally due to the network effects documented (i.e. $20 \times 80 = 1,600$). Considering that only 25% of students skipping school in the network attended the rally, we could also say that 400 high-school students in the rally induced one additional student to attend ($0.25 \times 1,600 = 400$). However, this number is likely to be a lower bound of social effects because (i) those in one network are likely appear in other networks (Jackson, 2019), and (ii) the estimate for the percentage of high-school students in the rally (25%) is presumably a lower bound.

Robustness of results. The non-linear reaction of students to decisions in their networks are similar when using different combinations of controls and when focusing on the sub-samples of students in 1st, 2nd, 3rd, and 4th grade of high-school (Table A.2, Figure A.11). Results are also robust to the use of alternative implementations of the Newey et al.'s (1999) estimator, in particular to splines for $f(\cdot)$ and flexible functions for the error term from the first-stage (Table A.3, Figure A.13). Finally, results are robust to the use of Newey and Powell's (2003) and Rau's (2013) estimators. The former methodology uses basis functions for the first-stage and polynomials of the instrument, while the latter uses saturated interaction terms between the error term from the first-stage and control variables and fixed effects from the second-stage (Figure A.13).

5.3 Discussion of mechanisms

This section discusses the underlying mechanism using heterogeneity of previous 2SLS non-linear results. I begin testing for enhanced coordination in student networks arising from the internet. I gathered administrative data from the Subsecretary of Telecommunications measuring the number of antennas operating in April of 2011 with their corresponding geographic location. These antennas emit radio electric signals that make connection to the internet possible using mobile phones. In practice, I constructed the number of antennas per 1,000 students in a $1km \times 1km$ gridded dataset spanning the entire country. Then I estimated the baseline specification in two sub-samples, (i) students located in cells without antennas (N =131,691) and, (ii) students located in cells with antennas ($N = 364,584$). Results are presented in Figure 3-A. These estimates reveal that, if anything, students in places without antennas are *more* likely to be influenced by their networks. Although suggestive, this result points towards social influence being more important in the absence of information-communication technologies, perhaps due to a substitution between online and offline network interactions.

Results appear to be stronger in groups with fewer students and among students with historically high attendance rates. Panel B in Figure 3 shows that students in smaller networks are more likely to respond to the network's decisions. In contrast, school or city size have smaller impacts on student networks (panels C and D), and students in private and public schools react similarly (panel E). Potential heterogeneous results by baseline absenteeism are also particularly important. If students with historically low school attendance are reacting more to social effects, then skipping school as a reaction to network absenteeism is more likely to be a pretext to simply skip school instead of a behavior related to the protests. The data, however, suggest this is unlikely to be the case. I estimated the baseline 2SLS specification in two groups: (i) students above the median of school attendance in 2010 (above 94% of attendance), and (ii) students below the median. If anything, estimates in panel F of Figure 3 suggest that students who historically skipped school *less* often ("High" school attendance) are the ones reacting more to social effects.[19]

What is the mechanism behind the observed complementarities? Although the data prevents me from providing one explanation, the collection of evidence suggests that social effects are likely to be important for at least four reasons. First, complementarities appear in early protests, when the educational cost of skipping school was low, both in terms of absenteeism itself and the likelihood of getting punished by teachers. Early protests were also unlikely to have changed the probability of achieving the desired policy change, i.e. cheaper tertiary education. Second, the similarity of results in places with differential access to the internet suggests *offline* connections were more important than *online* connections. Social effects such as reputation, retaliation, conformity, and reciprocity are arguably more salient in offline relationships. Additionally, the similarity of results across low-income (public) and high-income (private) schools also suggests social effects are likely to be more important than information about the social dimension of the protests. Third, the use

[19] Figure A.4 presents additional results suggesting the presence of homophily patterns of influence. In particular, I show that males (females) are more likely to influence males (females), and students are more likely to influence others of similar income and internet connectedness.

of within school variation makes information unlikely to be the explanation since students in the same school are presumably similarly informed. And fourth, complementarities are more clearly visible in small networks where social effects are probably stronger.

6 Structural estimation

This section presents a structural estimation of the game in section 3. The motivation is to take the model to the data and compare the parameter estimates with the reduced-form results. In the model students are imperfectly informed about the protest decisions in their networks but are perfectly informed about the network structure and students' characteristics. Following the assumptions in Xu (2018), we can interpret these results structurally as decisions that are part of a *unique* equilibrium played by students in the context of a static game with incomplete information.

6.1 From theory to estimation

To estimate the model, we first need to specify several functional forms. Let the utility that student i gets from skipping school during the June 16 national protest day be denoted by:

$$U_{i1}\left(\Omega, \epsilon_i\right) = \Theta_{s(i)} + x_i'\beta + \frac{\delta}{N_i}\sum_{h\in j(i)} \mathbb{E}\left[A_h = 1|\Omega\right] + \epsilon_{i1} \tag{5}$$

where A_h is the school skipping rate of a student in i's network – unobserved from i's perspective – and Ω represents public information. In addition, N_i denotes the number of students in i's network, the vector x_i represents i's characteristics, $\Theta_{s(i)}$ are school-level intercepts, and ϵ_{i1} is a random utility term drawn from an extreme value distribution with density function $f(t) = \exp(-t)\exp[-\exp(-t)]$.

Because students do not observe A_h, they form an expectation about the equilibrium probability that each student in their network skips school based on public information. Instead of implementing a full-solution method, where the equilibrium is computed for every trial vector of parameters, I estimate the model using a two-step estimation procedure based on Hotz and Miller (1993). In the first step, I estimate the conditional choice probabilities (CCPs) dictating whether a student chooses to attend school as a function of public information:

$$\mathbb{E}\left[A_h = 1|\Omega\right] = \frac{\exp\left[g(\mathbf{x}) + \Theta_{s(h)}\right]}{1 + \exp\left[g(\mathbf{x}) + \Theta_{s(h)}\right]} \tag{6}$$

where $g(\mathbf{x})$ are flexible functions (second order polynomials) of the characteristics of h, the characteristics of students in their first degree network, and the characteristics of students in their second degree network.[20] In addition, $\Theta_{s(h)}$ are intercepts for schools in which students h were enrolled. I estimate equation (6) by maximum likelihood. Table A.4 presents estimation results of two specifications: including h's characteristics and the characteristics of students in the first degree network (MLE I), and adding the characteristics of students in the second degree network (MLE II).

In the second step, I use the CCP estimates to form each student's expectations about the skip rate in their networks in equation (5). Specifically, student i aggregates the expected actions following the canonical linear-in-means model of network influence where all students in the network exert similar influence in his decision:[21]

$$a_{j(i)} \equiv \frac{1}{N_i} \sum_{h \in j(i)} \mathbb{E}[A_h = 1|\Omega] \tag{7}$$

I then proceed to estimate the parameters in equation (5) using a simple maximum likelihood estimator:

$$\widehat{\Gamma} = \underset{\tau}{\operatorname{argmax}} \sum_{i \in \mathcal{I}} \left\{ A_i \ln\left(\frac{\exp[\boldsymbol{\tau}\mathbf{W}]}{1+\exp[\boldsymbol{\tau}\mathbf{W}]}\right) + (1-A_i)\ln\left(1 - \frac{\exp[\boldsymbol{\tau}\mathbf{W}]}{1+\exp[\boldsymbol{\tau}\mathbf{W}]}\right)\right\} \tag{8}$$

where $\boldsymbol{\tau}\mathbf{W} \equiv \Theta_{s(i)} + x_i'\beta + \delta a_{j(i)}$ and $\widehat{\Gamma} = \left\{\widehat{\Theta}_s, \widehat{\beta}, \widehat{\delta}\right\}$ are the estimates of the model primitives. The main parameter of interest is δ and indicates whether students' protest decisions are strategic complements ($\delta > 0$) or substitutes ($\delta < 0$). Given that equation (8) uses the generated variable a, I present asymptotic MLE standard errors as well as p-values calculated using the score bootstrap proposed by Kline and Santos (2012) with 50 replications and schools as clusters.

6.2 *Estimation results*

Table 3 presents estimation results. Columns 1-2 present maximum likelihood estimates of $\widehat{\delta}$ using the two specifications for equation (6). In both cases we observe evidence of complementarities in protest decisions. In this sense, these structural estimates support the reduced-form results, which also suggested the existence of complementarities in protest decisions.

[20] The characteristics of students are: indicators for skipping school in May 12 June 1, age, attendance in 2010, GPA in 2010, retention in 2010, gender, and school switching in 2010.

[21] Note that the function in equation (4) cannot be used in this estimation because the existence of an equilibrium is based on the continuity of $f(a)$. Recall that reduced-form results are robust to the use of (continuous) splines.

To get a sense of the magnitude of these complementarities, the bottom of this table presents the network elasticity, defined as the change in a student's probability of skipping school, as a response to a change in the network's skipping rate from 0 to 100 percent. I calculate this elasticity using the commonly-used marginal effects while holding all other covariates at their average values. Column 2 constitutes my preferred specification (MLE II). The results suggest a network elasticity of 0.04: when the skipping rate in i's network jumps from 0 to 100 percent, the probability that student i skips school increases by 4 percentage points, an increase of 20% over the sample mean of 0.20.

This structural estimate is somewhat smaller in magnitude than the network elasticity of 0.09 estimated with the reduced-form strategy (column 4 of Table 2, panel B). However, estimates across estimation techniques are difficult to interpret because the underlying assumptions are different and the reduced-form estimates are identified based on the subpopulation of compliers. All in all, we observe robust evidence of complementarities in protest decisions within student networks, with an elasticity that is positive but small and significantly lower than one.

7 Consequences of protests

This section estimates the cost of skipping school in the second half of the 2011 academic year (June through November) and its effects on electoral outcomes.

7.1 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 dropout, 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:

$$y_{hst} \quad = \quad \beta_t \times (G_{hs} \times T_t) + \zeta_{hs} + \lambda_t + \varepsilon_{hst} \tag{9}$$

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 4-A and 4-B present coefficients $\widehat{\beta_t}$. Figure 4-A uses absenteeism as dependent variable and Figure 4-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.[22] 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.

Let me now estimate student-level costs using equation (3) and 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. This is a reduced form relationship but network absenteeism on June 16 is highly predictive of the percentage of days of school a student missed in 2011 (slope 0.41, p-value<0.01). Figures 4-C and 4-D present estimates using grade point average (standardized GPA) and an indicator for grade retention as dependent variables. A 100 percent absenteeism in networks on June 16 is associated with (i) a decrease of 0.16 standard deviations in academic performance, and (ii) a 38 percent increase in grade retention (from a base retention of 6 percent in 2010).

Finally, consider the same regression but using student-level absenteeism on June 16 as the main independent variable. Students who skipped school that day missed 24 percentage points more days of school in 2011 (p-value<0.01). The estimate suggest that absenteeism in June 16 leads to (i) a decrease of 0.10 standard deviations in GPA (coefficient of -0.07, p-value<0.01), and (ii) a 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.

7.2 The political effects of the student movement

The first election after the rise of the student movement was held on October 2012.[23] 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-

[22]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 (i.e. the denominator).

[23]There was an informal plebiscite previously organized by citizens, in October 2011. Figure A.6 shows that participation was higher and people agreed more with students' demands in counties with higher school absenteeism.

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 parties.[24]

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:

$$V_{c,2012} = \alpha + \beta \cdot \text{Student Movement}_{c,2011} + \gamma V_{c,2008} + \delta X_{c,2009} + \varepsilon_c \qquad (10)$$

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 temporarily closed by students. Finally, ε_c is a robust error term. The dependent variables 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.[25]

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 threat. 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. Third, I use the method proposed by Altonji et al. (2005) to construct bounds for estimates and conclusions remain.

Table 4 presents estimates. Column 1 indicates that a one standard deviation increase in the intensity of the student movement is associated with a 5 p.p. 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. Column 4 suggests that the same increase in

[24]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.

[25]Electoral outcomes are administrative data reported by the Electoral Service of Chile. Population data come from censuses. Figure A.14 plots the student movement variable for all counties.

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 increase in county-level school absenteeism between 2008 and 2007, i.e. before the student movement, and examine their impact in the 2008 local elections. I also re-estimate equation (10) 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 (2019) 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 $[0.050, 0.086]$.[26]

8 Conclusion

Studying the Chilean student movement of 2011, this paper showed evidence of complementarities in protest behavior within networks of high-school students. The results also constitute suggestive evidence supporting the popular idea of a tipping point in behavior (Gladwell, 2000) and the importance of strong ties to promote political activism (McAdam, 1986).

These findings 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. Ballester et al. (2006) provides a way to reconcile both approaches, allowing decisions to exhibit (local) complementarities within friendship networks but substitutability with the rest of the population. Second, complementarities in protest behavior imply that individuals with larger networks could be more influential. This is 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. My findings suggest that the marginal return of enrolling one additional citizen in the movement is higher for individuals with larger networks. Similarly, a

[26] 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 (2019) for details.

state could disrupt a social movement by preventing central individuals from participating.

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 and the missing dynamics in network structure prevent us from a full understanding of the decision to participate in a social movement.

Finally, my findings suggest that social dimensions in protest behavior are important, and open new and interesting questions to explore. For example, future studies of social movements may explore how protests create network links among participants, how police violence disrupts (or foster) protest participation, or how habit formation contributes to the escalation of a mobilization.

References

Abadie, A., Angrist, J., and Imbens, G. (2002). Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica*, 70(1):91–117.

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.

Ballester, C., Calvó-Armengol, A., and Zenou, Y. (2006). Who's who in networks. Wanted: the key player. *Econometrica*, 74(5):1403–1417.

Barbera, S. and Jackson, M. O. (2020). A model of protests, revolution, and information. *Quarterly Journal of Political Science*.

Blattman, C. and Miguel, E. (2010). Civil war. *Journal of Economic Literature*, 48(1):3–57.

Blume, L. E., Brock, W. A., Durlauf, S. N., and Ioannides, Y. M. (2010). Identification of social interactions. In Benhabib, J., Bisin, A., and Jackson, M., editors, *Social Economics*, chapter 18, pages 853–964. North Holland: Amsterdam.

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.

Bursztyn, L., Cantoni, D., Yang, D. Y., Yuchtman, N., and Zhang, Y. J. (2020). Persistent political engagement: Social interactions and the dynamics of protest movements. *Working Paper*.

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. (2019). Protests as strategic games: Experimental evidence from Hong Kong's democracy movement. *Quarterly Journal of Economics*, 134(2):1021–1077.

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

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.

Coppock, A. and Green, D. P. (2016). Is voting habit forming? New evidence from experiments and regression discontinuities. *American Journal of Political Science*, 60(4):1044–1062.

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.

de Paula, A. (2013). Econometric analysis of games with multiple equilibria. *Annual Review of Economics*, 5:107:31.

de Paula, A. (2017). Econometrics of network models. In Honore, B., Pakes, A., Piazzesi, M., and Samuelson, L., editors, *Advances in Economics and Econometrics: Theory and Applications*, chapter 8, pages 268–323. Cambridge University Press.

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.

Elster, J. (1989). Social norms and economic theory. *Journal of Economic Perspectives*, 3(4):99–117.

Enikolopov, R., Makarin, A., and Petrova, M. (2020). Social media and protest participation: Evidence from Russia. *Econometrica*.

Enikolopov, R., Makarin, A., Petrova, M., and Polishchuk, L. (2018). Social image, networks, and protest participation. *Working Paper*.

Enos, R. D., Kaufman, A. R., and Sands, M. L. (2019). Can violent protest change local policy support? Evidence from the aftermath of the 1992 Los Angeles riot. *American Political Science Review*, 113(4):1012–1028.

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

Fisher, D. R., Andrews, K. T., Caren, N., Chenoweth, E., Heaney, M. T., Leung, T., Perkins, N., and Pressman, J. (2019). The science of contemporary street protest: New efforts in the United States. *Science Advances*, 5(10):EAAW5461.

Fujiwara, T., Meng, K., and Vogl, T. (2016). Habit formation in voting: Evidence from rainy elections. *American Economic Journal: Applied Economics*, 8(4):160–188.

Gilchrist, D. S. and Sands, E. G. (2016). Something to talk about: Social spillovers in movie consumption. *Journal of Political Economy*, 124(5):1339–1382.

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

Graham, B. S. (2015). Methods of identification in social networks. *Annual Review of Economics*, 7:465–485.

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.

Hager, A., Hensel, L., Hermle, J., and Roth, C. (2019). Strategic interdependence in political movements and countermovements. *CESifo Working Papers*.

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.

Hensel, L., Hermle, J., Rink, A., and Roth, C. (2019). Political activists as free-riders: Evidence from a natural field experiment. *Working Paper*.

Horowitz, J. L. (2011). Applied nonparametric instrumental variables estimation. *Econometrica*, 79(2):347–394.

Hotz, V. J. and Miller, R. A. (1993). Conditional choice probabilities and the estimation of dynamic models. *Review of Economic Studies*, 60(3):497–529.

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. (2019). The friendship paradox and systematic biases in perceptions and social norms. *Journal of Political Economy*, 127(2):777–818.

Jackson, M. O. and Storms, E. C. (2019). Behavioral communities and the atomic structure of networks. *Working Paper*.

Kline, P. and Santos, A. (2012). A score based approach to wild bootstrap inference. *Journal of Econometric Methods*, 1:23–41.

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.

Larreboure, M. and González, F. (2020). The impact of the Women's March on the U.S. House Election. *Working Paper*.

Larson, J., Nagler, J., Ronen, J., and Tucker, J. (2019). Social networks and protest participation: Evidence from 130 million Twitter users. *American Journal of Political Science*, 63(3):690–705.

Levy, R. and Mattsson, M. (2019). The effects of social movements: Evidence from #metoo. *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. (2020). Liberation technology: Mobile phones and political mobilization in Africa. *Econometrica*, 88(2):533–567.

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.

Mazumder, S. (2019a). Black Lives Matter for white's racial prejudice: Assessing the role of social movements in shaping racial attitudes in the United States. *Working Paper*.

Mazumder, S. (2019b). The persistent effect of U.S. Civil Rights protests on political attitudes. *Working Paper*.

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. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business Economics and Statistics*, pages 1–18.

Passarelli, F. and Tabellini, G. (2017). Emotions and political unrest. *Journal of Political Economy*, 125(3):903–946.

Perez-Truglia, R. and Cruces, G. (2017). Partisan interactions: Evidence from a field experiment in the United States. *Journal of Political Economy*, 125(4):1208–1243.

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.

Roodman, D., MacKinnon, J. G., and Orregaard Nielsen, M. (2019). Fast and wild: bootstrap inference in Stata using boottest. *The Stata Journal*, 19(1):4–60.

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.

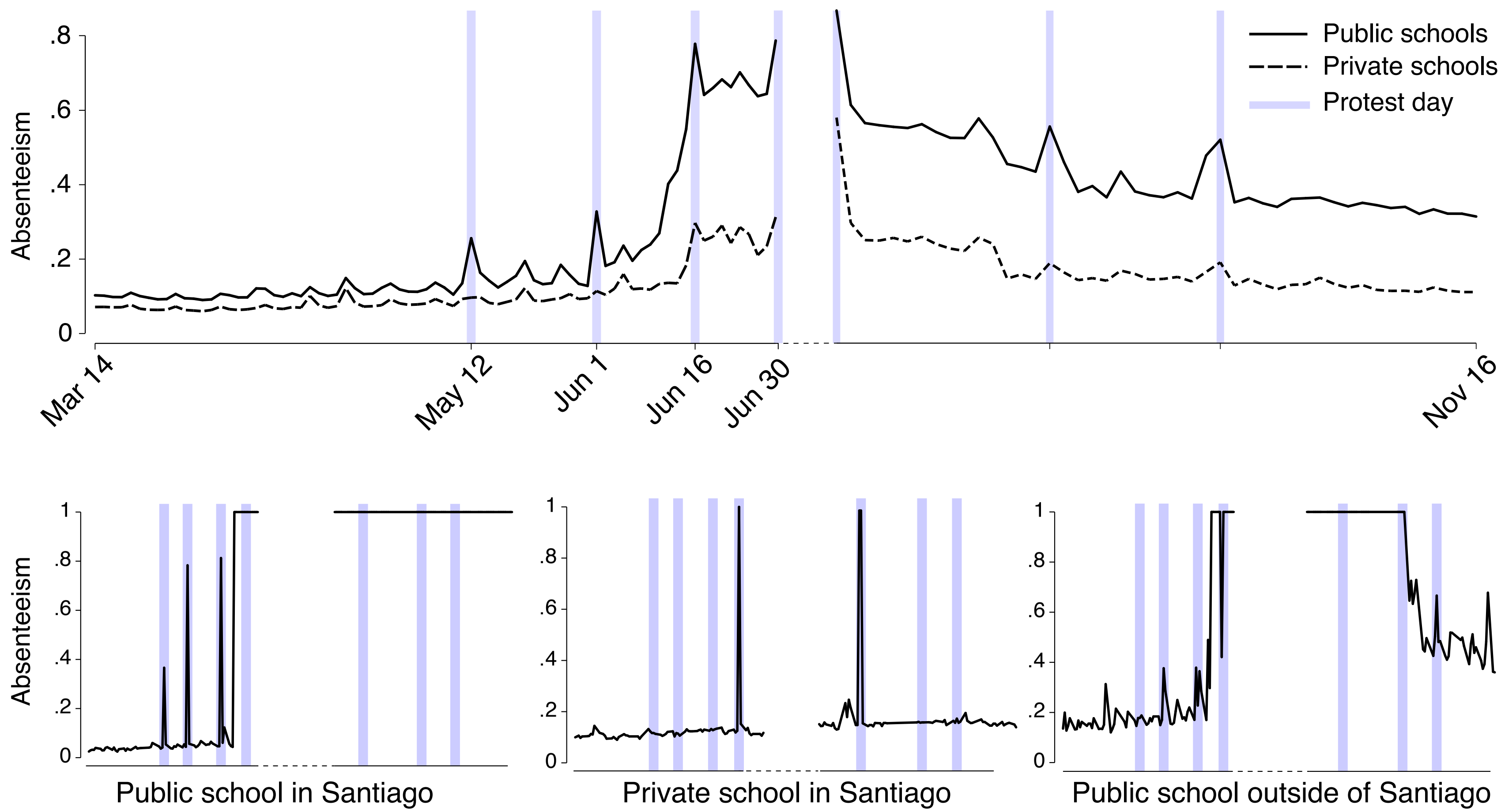
Xu, H. (2018). Social interactions in large networks: A game theoretic approach. *International Economic Review*, 59(1):257–284.

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

Yang, D. (2019). The impact of media censorship: 1984 or Brave New World? *American Economic Review*, 109(6):2294–2332.

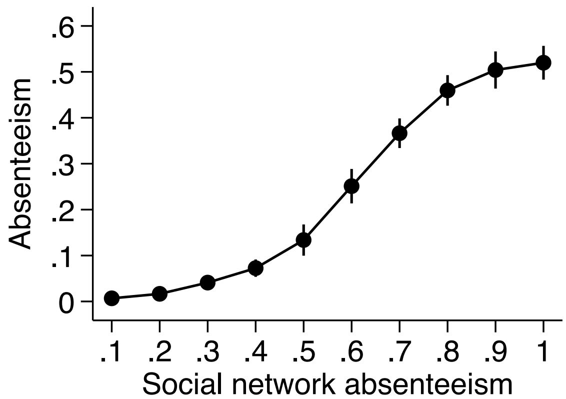
Yitzhaki, S. (1996). On using linear regressions in welfare economics. *Journal of Business and Economic Statistics*, 14(4):478–486.

Figure 1: Absenteeism of high-school students in 2011

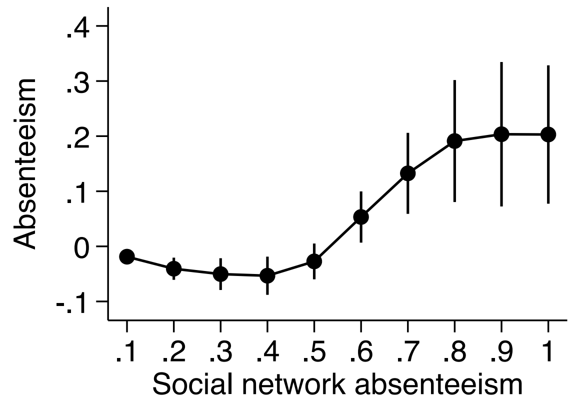


Notes: Own construction using administrative data for all students with school absenteeism data. The y-axis is the average school absenteeism among high-school students (in percentages) and the x-axis represents days in 2011. Vertical lines denote the most important national protest days during week 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 4.1.

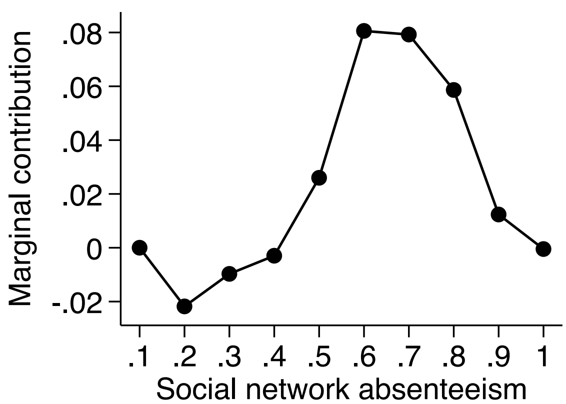
Figure 2: Reduced-form results



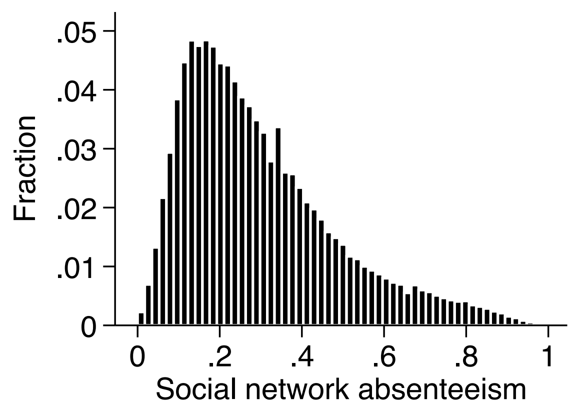
(a) OLS estimates



(b) 2SLS estimates



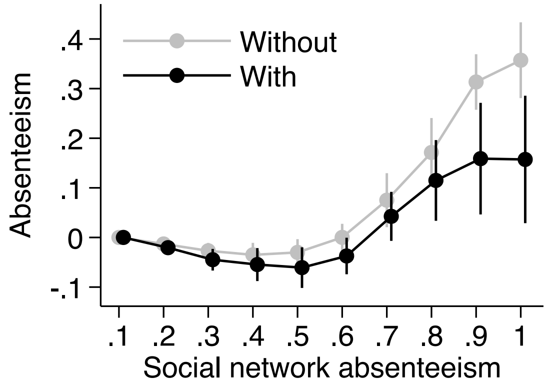
(c) Marginal contribution



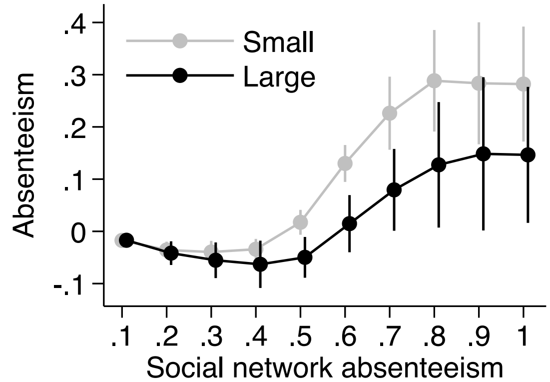
(d) Distribution of network absenteeism

Notes: Panels (a) and (b) plot OLS and 2SLS estimates from a regression of individual school absenteeism on 10 indicators of network absenteeism on June 16, controlling for network characteristics and school fixed effects (see equations 3 and 4). Vertical lines denote 95 percent confidence intervals (s.e. clustered at the city level). Panel (c) plots the difference in the estimated 2SLS coefficients in Panel (b). Panel (d) plots the empirical distribution of network absenteeism in June 16. More details in section 5.2.

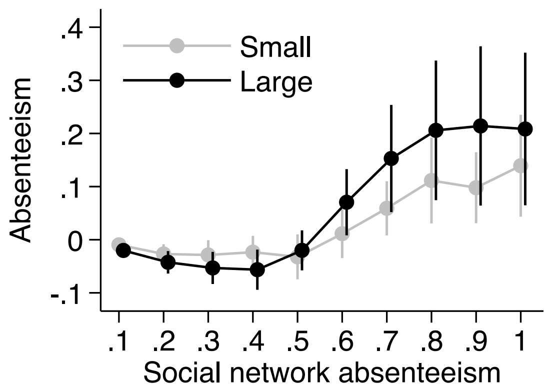
Figure 3: Additional reduced-form results



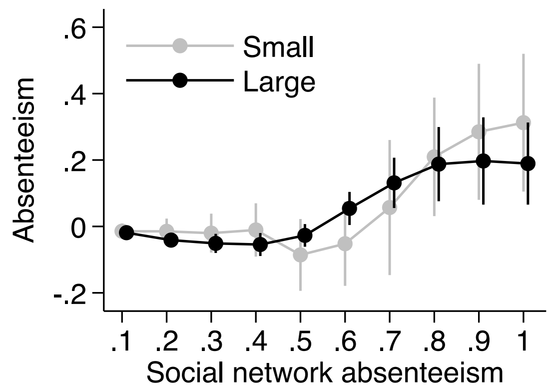
(a) 2G-3G coverage network



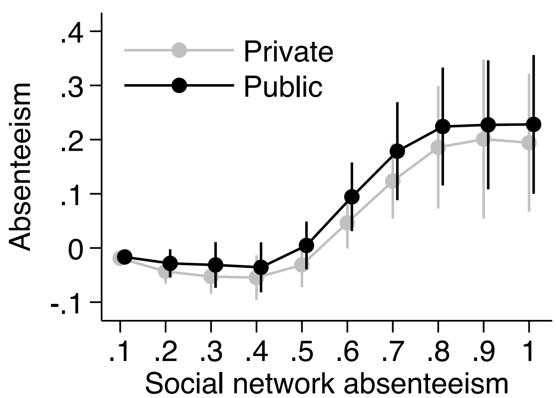
(b) Network size



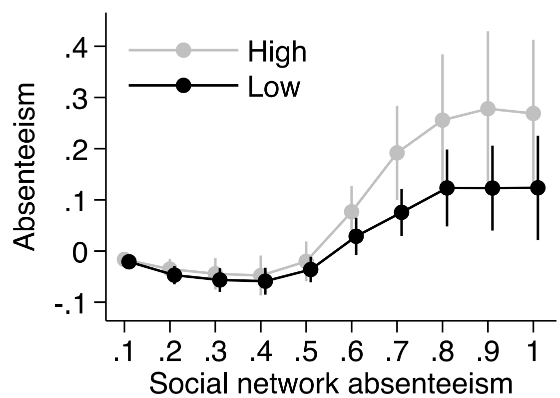
(c) School size



(d) City size



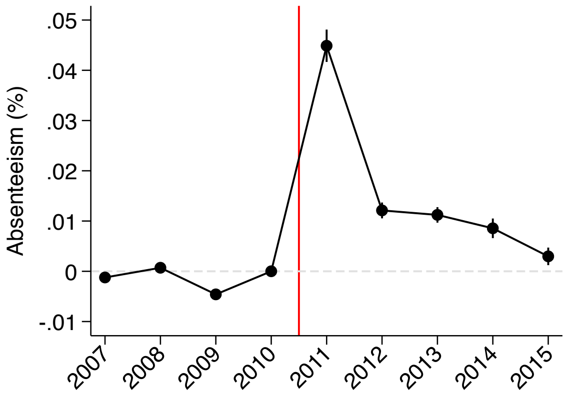
(e) School type sub-samples



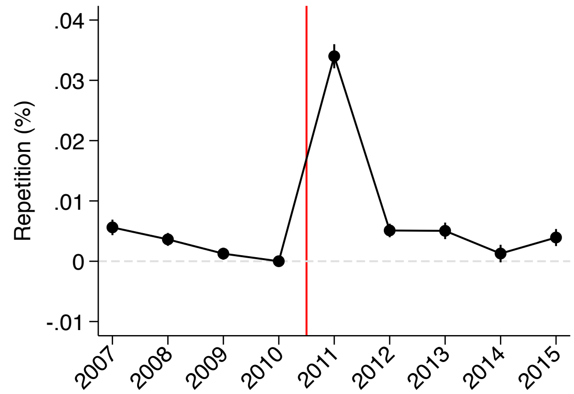
(f) Attendance in 2010 sub-samples

Notes: Panels (a) and (b) plot 2SLS estimates from a regression of individual school absenteeism on 10 indicators of network absenteeism, controlling for student and network characteristics, and school fixed effects (see equations 3 and 4). Panels (c)-(f) present 2SLS estimates in sub-samples. More details in section 5.3.

Figure 4: The cost of skipping school

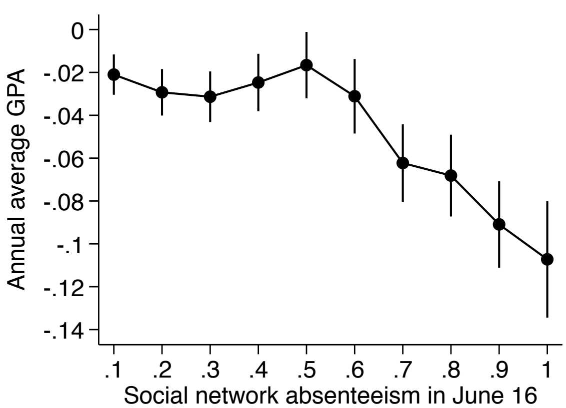


(a) Absenteeism 2002–2015

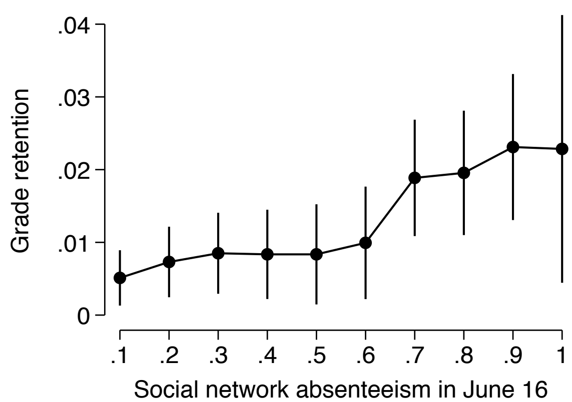


(b) Grade retention 2002–2015

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 7.1.



(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 7.1.

Table 1: Descriptive statistics

Students	Students in high-schools opened in June 16 (1)	Observations (2)	All high-schools (3)	Observations (4)
Absenteeism in May 12, 2011	0.10	500,935	0.15	760,801
	(0.30)		(0.36)	
Absenteeism in June 1, 2011	0.12		0.19	
	(0.32)		(0.39)	
Absenteeism in June 16, 2011	0.21		0.49	
	(0.40)		(0.50)	
Average absenteeism in 2010	0.07		0.07	
	(0.07)		(0.07)	
Schools				
Indicator for public	0.16	1,719	0.30	2,224
	(0.37)		(0.46)	
Number of high-school students	289		342	
	(283)		(345)	
Cities				
High-schools in the city	7.2	240	7.7	290
	(36.9)		(44.3)	
High-school students in the city	2,067		2,623	
	(11,958)		(16,134)	

Notes: Own construction based on administrative data provided by the Ministry of Education. Descriptive statistics for the sample of high-school students enrolled in schools opened in June 16 in column 1, and for all high-schools in column 3. All variables are measured in 2011 unless otherwise stated. More details in section 4.1.

Table 2: Linear estimates

Dependent variable is absenteeism on June 16, first massive protest

	(1)	(2)	(3)	(4)
Panel A – OLS estimates				
Network absenteeism on June 16	0.48***	0.46***	0.44***	0.49***
	(0.06)	(0.06)	(0.06)	(0.06)
	[0.07]	[0.06]	[0.06]	[0.07]
Panel B – 2SLS estimates				
Network absenteeism on June 16	0.26***	0.20***	0.05**	0.09**
	(0.04)	(0.04)	(0.03)	(0.04)
	[0.04]	[0.04]	[0.04]	[0.04]
Panel C – First-stage				
Instrument	0.69***	0.68***	0.68***	0.53***
	(0.05)	(0.05)	(0.05)	(0.04)
	[0.05]	[0.05]	[0.05]	[0.04]
Panel D – Reduced form				
Instrument	0.17***	0.14***	0.04*	0.05**
	(0.03)	(0.03)	(0.02)	(0.02)
	[0.03]	[0.03]	[0.02]	[0.02]
School fixed effects	X	X	X	X
Daily absenteeism before June 16		X	X	X
Student controls			X	X
Network controls				X
Kleibergen-Paap F-statistic	226.6	225.9	211.0	192.0
Mean of dependent variable	0.21	0.21	0.20	0.20
R-squared (Panel A)	0.30	0.33	0.34	0.34
Students	501,139	500,935	500,904	496,275

Notes: "Daily absenteeism before June 16" includes student-level indicators for school absenteeism in May 12 and June 1. "Student controls" include academic performance and average school attendance in previous years and predetermined socioeconomic characteristics. "Network controls" include average student controls at the network level. The instrument is past classmates of 2011 classmates. Standard errors clustered at the city level in parentheses and at the province level in square brackets (240 and 53 clusters respectively). Significance level: *** $p < 0.01$.

Table 3: Structural estimation

Estimation of network effects in coordination game		
	MLE I	MLE II
	(1)	(2)
Network effects $\left[E(A_{j(i)})\right]$	0.72***	0.86***
	(0.17)	(0.10)
Students $[i \in \mathcal{I}]$	504,105	504,105
Student characteristics $[x_i]$	X	X
School-level intercepts $[\Theta_{s(i)}]$	X	X
Network elasticity	0.03	0.04
Bootstrap p-value for network effects	< 0.01	< 0.01

Notes: Maximum likelihood (logit) estimates for the decision to skip school during the June 16 protest. The network effects are identified from variations in (expected) networks' decisions:

$$E(A_{j(i)}) = \frac{1}{N_i} \sum_{h \in j(i); \ell \in j(h)} \mathbb{E}\left[A_h = 1 | x_h, x_\ell, \Theta_{s(h)}\right]$$

The variable $A_{j(i)}$ represents the average decision to skip school in i's network, and (x_i, Θ_s) are observables that are public information, including students' characteristics x_i and the schools in which they are enrolled Θ_s. Columns 1-3 use the expected rate of skipping based on the public information and three ML logit models. Standard errors in parentheses are clustered by school. Bootstrap p-values for the null hypothesis that network effects are equal to zero are calculated using the score bootstrap proposed by Kline and Santos (2012) with 50 replications, schools as clusters, and the code by Roodman et al. (2019). More details in section 6.

Table 4: 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 left-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

List of Tables

List of Figures

A Additional reduced-form results

A.1 Partially overlapping networks in panel data

This 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:

$$A_{isct} \quad = \quad f(A_{j(i)t}) + \sum_t \left(\delta_{1t} x_i + \delta_{2t} x_{j(i)}\right) + \xi_i + \zeta_{st} + \epsilon_{isct} \tag{A.1}$$

where A_{isct} is an indicator that takes the value of one if student i, in school s, located in city c, skipped 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 and $x_{j(i)}$ are control variables by students and networks. The baseline specification includes student's past GPA and the average GPA in social networks, although results are robust to include more variables. Finally, ξ_i is a student fixed effect, ζ_{st} is a school by day fixed effect, and ϵ_{isct} is an error term clustered by city. As in equation (3), I employ the functional form in equation (4) 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 and network characteristics, and (2) include protest day by school fixed effects.

Figure A.3 confirms reduced-form results using 2SLS panel data estimates of equation (A.1), the functional form in equation (4), and Newey et al.'s (1999) estimation. These regressions employ more than five million observations, coming from more than 700 thousand students during eight protest days. The estimates in Figure A.3-B reveal the same non-linear network patterns from the previous section: networks begin to influence individual decisions after 50 percent absenteeism and the marginal contribution of additional absenteeism is again maximized at 60 percent. Finally. Figure shows that 2SLS estimates using panel data are also robust to the inclusion of one, two, or three lags of individual and network absenteeism, particularly important in the potential presence of habit formation in absenteeism decisions (Figure A.12).

A.2 Homophilic influence in reduced-form analysis

Does the strength of influence in student networks follows homophily patterns? Figure A.4 presents results. Panels A and B test for gender homophily patterns of influence by estimating equation (3), restricting attention to males or females, and splitting the network into males and females. 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. 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 (2020) and Enikolopov et al. (2020) provide city-level evidence of stronger network coordination with increased access to cell phones and social media.

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. 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.

Table A.1: Additional summary statistics and characteristics of compliers

	Treated compliers	Untreated compliers	Full sample
	(1)	(2)	(3)
Student enrolled in public school in 2011	0.12	0.08	0.21
			(0.41)
Student absenteeism May 12, 2011	0.10	0.06	0.10
			(0.30)
Student absenteeism June 1, 2011	0.12	0.09	0.12
			(0.33)
Student GPA in 2010	5.53	5.44	5.42
			(0.59)
Student retention in 2010	0.05	0.03	0.05
			(0.22)
Student attendance in 2010	92.3	93.8	93.1
			(6.72)
Student gender (female)	0.51	0.51	0.51
			(0.50)
Student age	15.4	15.4	15.7
			(1.27)
Student switched in 2010	0.32	0.18	0.23
			(0.42)
Network GPA in 2010	5.40	5.47	5.40
			(0.25)
Network retention in 2010	0.06	0.06	0.06
			(0.54)
Network attendance in 2010	91.7	93.6	92.9
			(2.26)
Network female in 2010	0.51	0.51	0.51
			(0.18)
Network age in 2010	15.8	14.9	15.7
			(1.1)
Network switcher in 2010	0.79	0.66	0.77
			(0.23)
Students			496,275

Notes: Columns 1 and 2 present the characteristics of compliers using the Abadie et al.'s (2002) method. Column 3 presents summary statistics (mean and standard deviation) for the full sample of students used in the analysis.

Table A.2: Robustness of 2SLS non-linear estimates (I)

Dependent variable is absenteeism on June 16 (columns 1-6) or several protest days (column 7)

	Empirical strategy:						
	Exposure to first protest						Panel data
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Social network absenteeism							
$\in [0.10, 0.20)$	-0.00	-0.01***	-0.02***	-0.02***	-0.02***	-0.02***	-0.01***
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)
$\in [0.20, 0.30)$	-0.01*	-0.02***	-0.04***	-0.04***	-0.04***	-0.04***	-0.01***
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
$\in [0.30, 0.40)$	0.00	-0.02**	-0.05***	-0.05***	-0.05***	-0.06***	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$\in [0.40, 0.50)$	0.02***	-0.01	-0.06***	-0.05***	-0.05***	-0.06***	0.01
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)
$\in [0.50, 0.60)$	0.07***	0.04***	-0.03***	-0.03*	-0.03*	-0.04**	0.06***
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.00)
$\in [0.60, 0.70)$	0.18***	0.13***	0.05***	0.05**	0.05**	0.04*	0.14***
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)
$\in [0.70, 0.80)$	0.28***	0.23***	0.13***	0.13***	0.13***	0.12***	0.20***
	(0.02)	(0.02)	(0.03)	(0.04)	(0.04)	(0.04)	(0.02)
$\in [0.80, 0.90)$	0.35***	0.30***	0.18***	0.19***	0.19***	0.18***	0.25***
	(0.03)	(0.04)	(0.04)	(0.06)	(0.06)	(0.06)	(0.03)
$\in [0.90, 1)$	0.39***	0.32***	0.19***	0.20***	0.20***	0.19***	0.26***
	(0.04)	(0.04)	(0.05)	(0.07)	(0.07)	(0.07)	(0.04)
= 100%	0.41***	0.33***	0.19***	0.20***	0.20***	0.18***	0.24***
	(0.04)	(0.04)	(0.05)	(0.06)	(0.06)	(0.06)	(0.05)
School fixed effects	X	X	X	X	X	X	X
Daily absenteeism before June 16		X	X	X	X		
Student controls			X	X	X		X
Network controls				X	X		X
LASSO-chosen controls						X	
Observations	496,275	496,275	496,275	496,275	496,275	496,275	5,133,035

Notes: Each observation corresponds to a student (columns 1-6) or a student-day (column 7). These estimates correspond to two-stage control function estimates of network effects in school absenteeism on June 16, day of the first massive protest, or a protest day (column 7). School, student, and network controls are interacted with protest day fixed effects in column 7. Standard errors clustered by city are reported in parentheses. Significance level: *** $p < 0.01$, ** $p < 0.05$.

Table A.3: Robustness of 2SLS non-linear estimates (II)

	Splines	
	(1)	(2)
Social network absenteeism	-0.58***	-0.29***
	(0.19)	(0.11)
Social network absenteeism^2	0.99**	-3.06***
	(0.41)	(0.40)
Social network absenteeism^3	-0.08	6.75***
	(0.31)	(0.78)
Social network absenteeism^4		-3.71***
		(0.49)
School fixed effects	X	X
Daily absenteeism before June 16	X	X
Student controls	X	X
Network controls	X	X
Observations	496,275	496,275

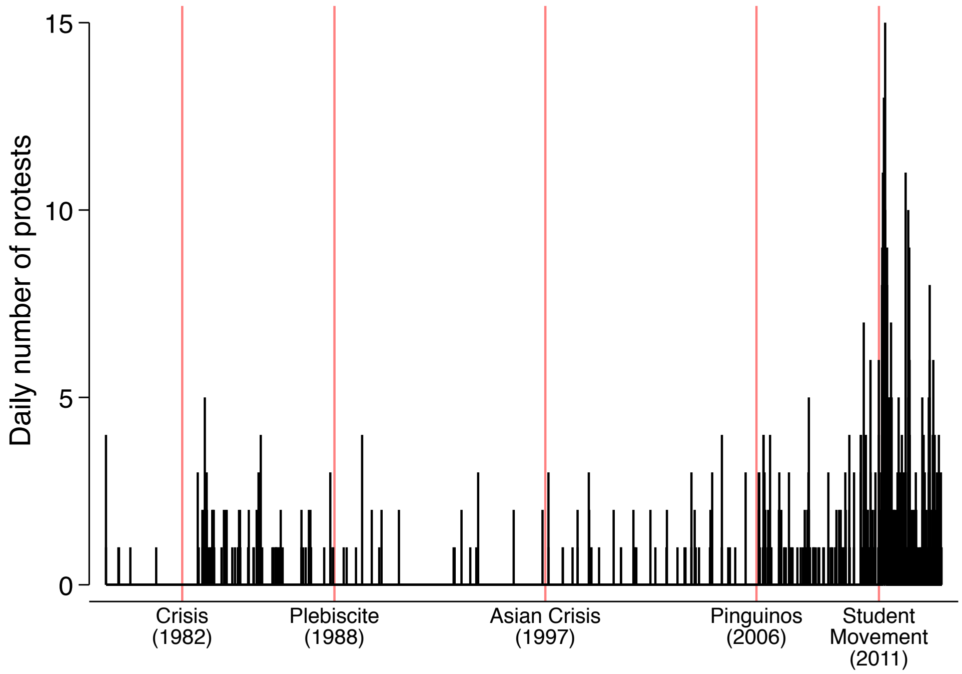
Notes: Each observation corresponds to a student. These estimates correspond to two-stage control function estimates of network effects in school absenteeism on June 16, day of the first massive protest. Standard errors clustered by city are reported in parentheses. Significance level: *** $p < 0.01$, ** $p < 0.05$.

Table A.4: Structural estimation, first step

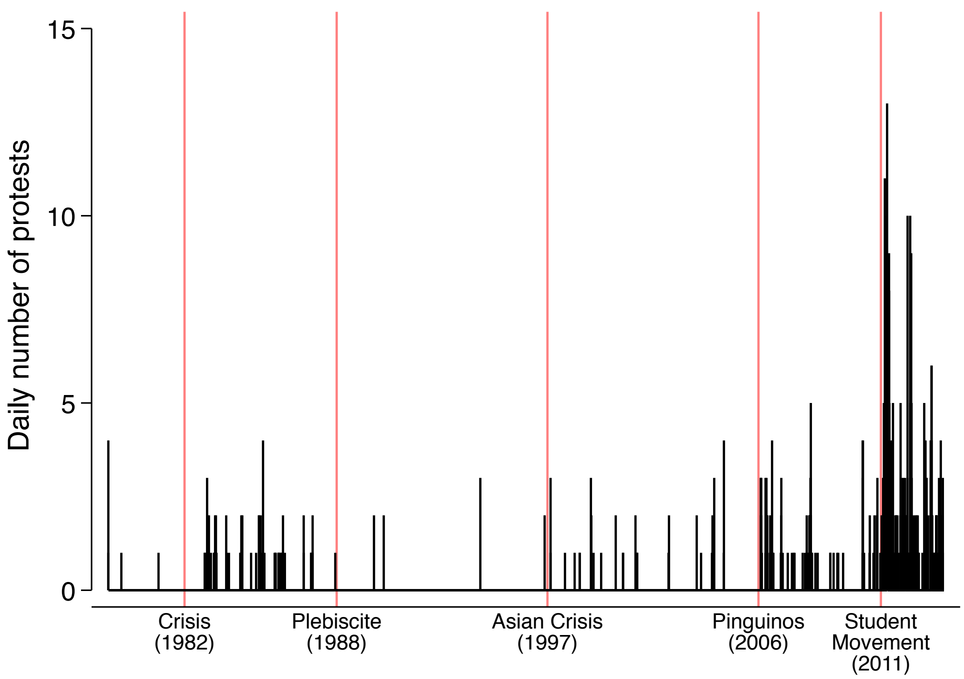
Dependent variable: Indicator skipped school in June 16, 2011			
	MLE I	MLE II	MLE III
Indicator student skipped school in May 12, 2011	0.79***	0.64***	0.63***
	(0.01)	(0.01)	(0.01)
Indicator student skipped school in June 1, 2011	1.02***	0.82***	0.81***
	(0.01)	(0.01)	(0.01)
Indicator student repeated grade in 2010	-0.16***	-0.19***	-0.20***
	(0.03)	(0.03)	(0.03)
Indicator student is female	0.14***	0.16***	0.16***
	(0.01)	(0.01)	(0.01)
Indicator student switched school in 2010	0.23***	0.14***	0.13***
	(0.01)	(0.01)	(0.01)
Indicator student age is 14	0.68	0.64	0.63
	(0.46)	(0.46)	(0.46)
Indicator student age is 15	0.88*	0.78*	0.74
	(0.46)	(0.46)	(0.46)
Indicator student age is 16	0.95**	0.85*	0.84*
	(0.46)	(0.46)	(0.46)
Indicator student age is 17	1.05**	0.96**	0.95**
	(0.46)	(0.46)	(0.46)
Indicator student age is 18	1.18**	1.09**	1.06**
	(0.46)	(0.46)	(0.46)
Indicator student age is 19	1.08**	1.02**	0.99**
	(0.47)	(0.47)	(0.46)
Indicator student age is 20	0.98**	0.92*	0.89*
	(0.47)	(0.47)	(0.47)
Indicator student age is 21	0.51	0.55	0.51
	(0.55)	(0.55)	(0.55)
Students	505,019	498,786	498,657
Avg. predicted skipping rate	0.21	0.21	0.21
School fixed effects	X	X	X
Student GPA and attendance in 2010 bins $[x_h]$	X	X	X
Characteristics of 1st degree network $\left[x_{j(h)}\right]$	–	X	X
Characteristics of 2nd degree network $\left[x_{j(j(h))}\right]$	–	–	X
Log-likelihood	-181,648	-176,622	-176,400

Notes: This table presents maximum likelihood (logit) estimates for the probability of skipping school in June 16. Network characteristics are included as a second-degreee polinomial, including all double-interactions. Standard errors in parentheses. Significance level: *** $p < 0.01$, ** $p < 0.05$.

Figure A.1: Protests in Chile 1979–2013



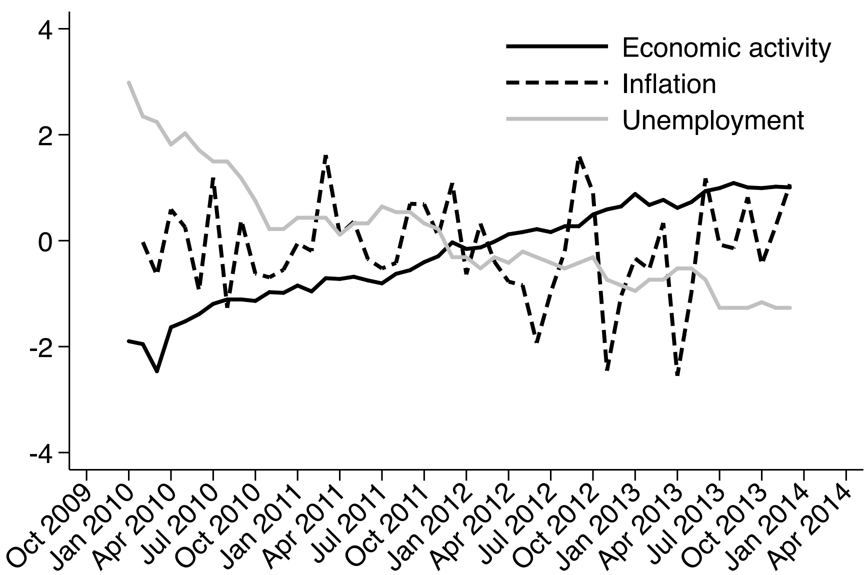
(a) Any type of protest event



(b) Protest events related to education

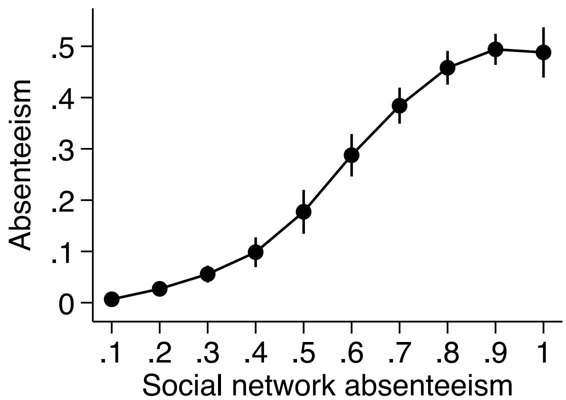
Notes: Data from the Global Dataset of Events, Language, and Tone.

Figure A.2: Economic indicators

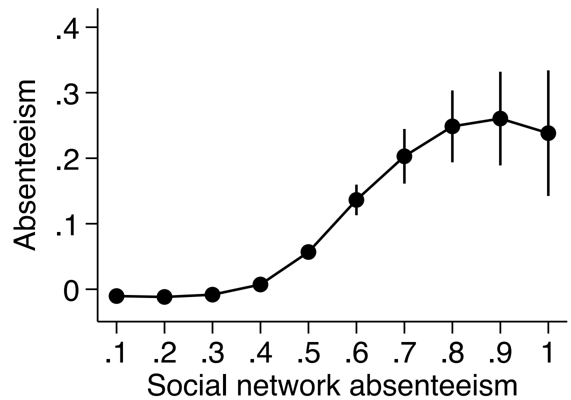


Notes: Data from the Central Bank of Chile. All variables have been normalized by subtracting their average and dividing by their standard deviation in the time series. The vertical red line denotes the beginning of the student movement.

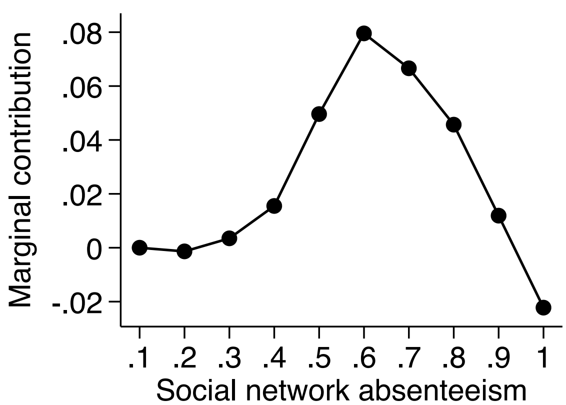
Figure A.3: Panel data estimates using multiple protest days



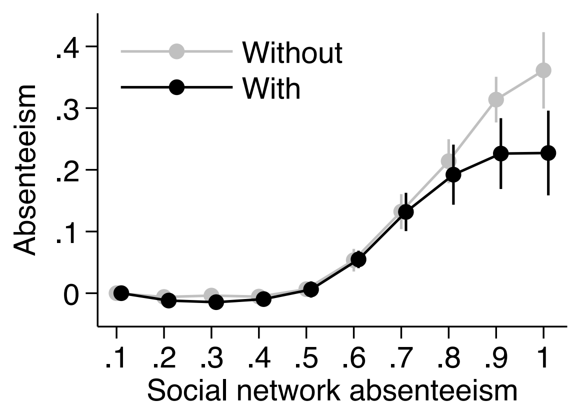
(a) OLS estimates



(b) 2SLS estimates



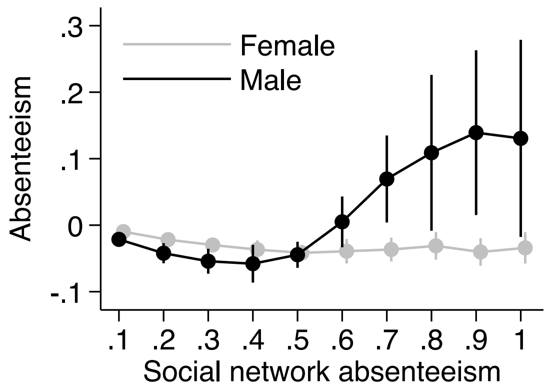
(c) Marginal contribution



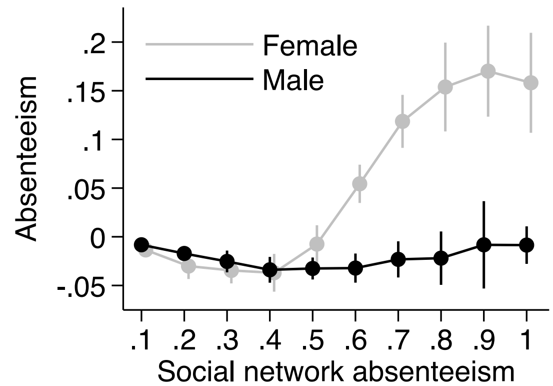
(d) 2G-3G coverage network

Notes: These estimates correspond to two-stage control function estimates of network effects in school absenteeism on protest days. The estimating sample includes a panel of students observed daily during protest days in schools that were opened that day (excludes school closures). The total number of observations is 5,140,042. All regressions include student and school-by-day fixed effects. For reference, the analogue linear estimate is 0.10 (s.e. 0.01). Vertical lines denote 95 percent confidence intervals (s.e. clustered at the city level). More details in Section 5.2.

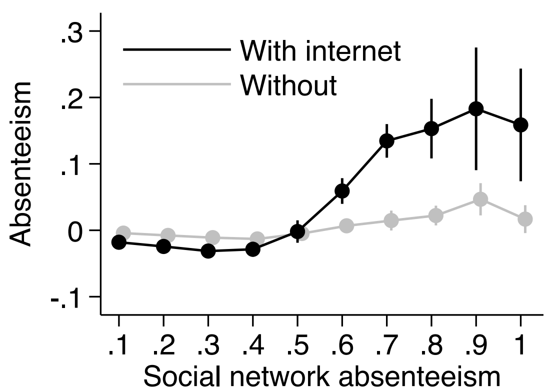
Figure A.4: Differential influence within networks



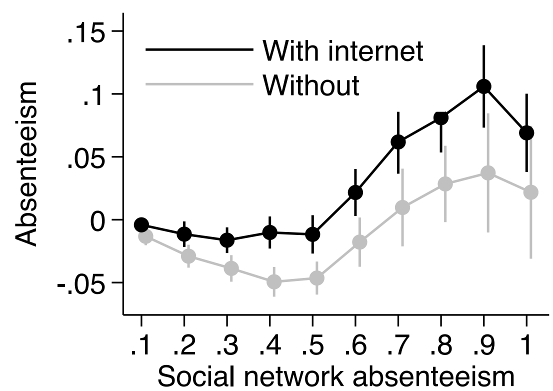
(a) Subsample of males



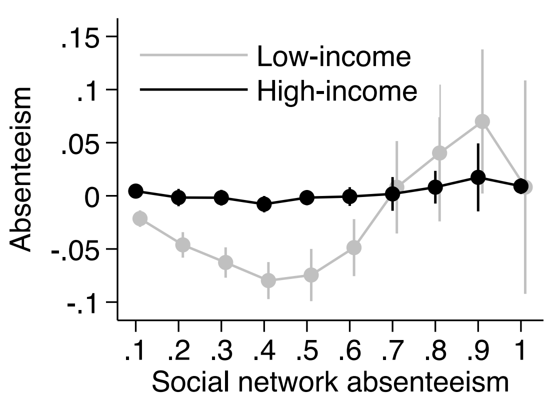
(b) Subsample of females



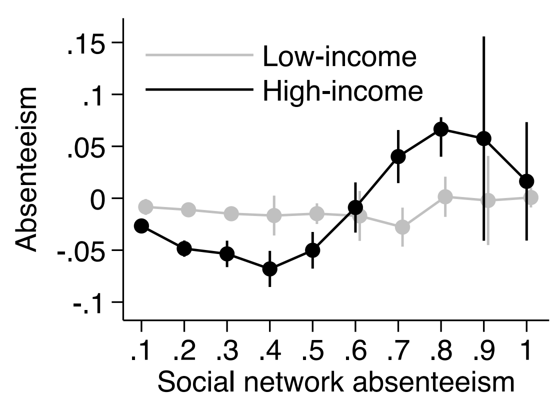
(c) Subsample with internet



(d) Subsample without internet



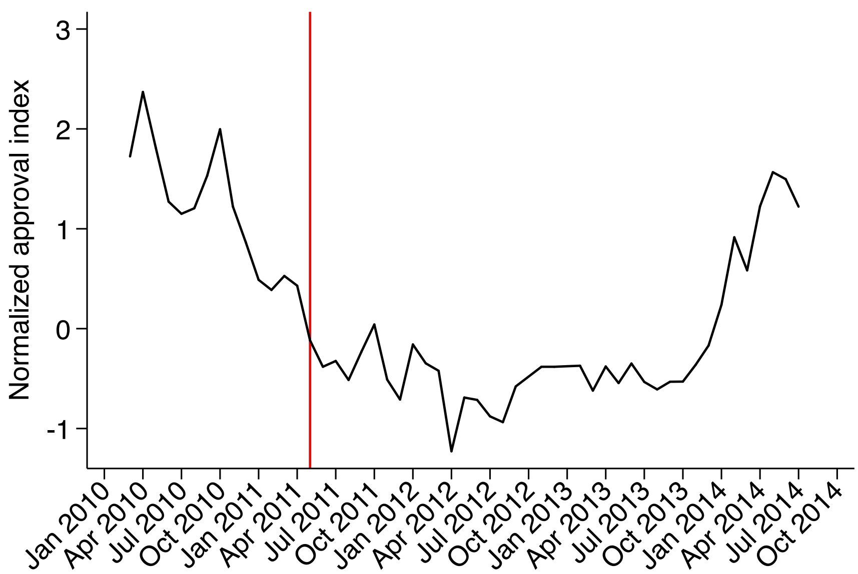
(e) Low-income households



(f) High-income households

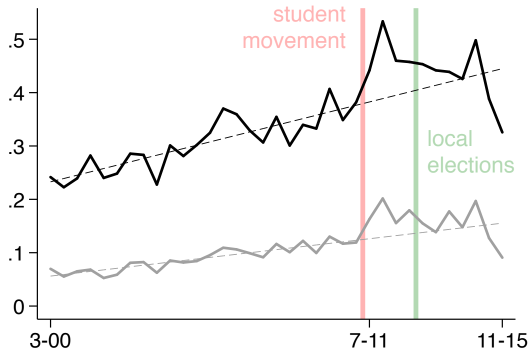
Notes: All panels plot 2SLS estimates from a regression of individual school absenteeism on 10 indicators of network absenteeism, controlling for student and network characteristics, and school fixed effects. Regressions are in sub-samples and split the network in groups. More details in Section **??**.

Figure A.5: Citizens' evaluation of incumbent politicians

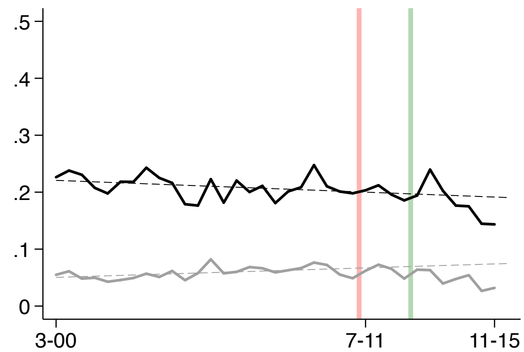


Notes: Normalized index (minus average and divide by standard deviation) for the approval of incumbent politicians. Data from the Centro de Estudios Públicos and Adimark.

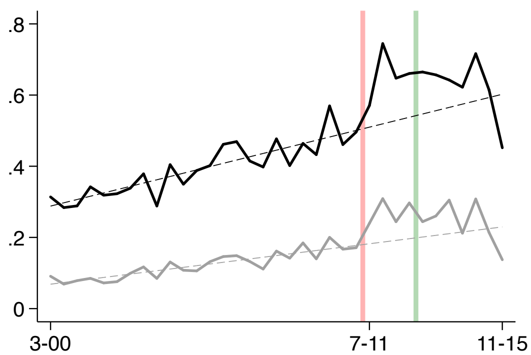
Figure A.6: Survey evidence for the impact of the student movement



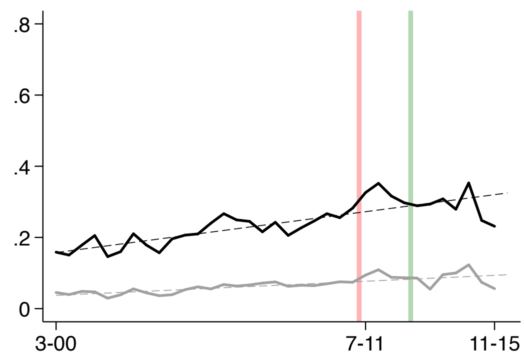
(a) Education should be priority



(b) Placebo

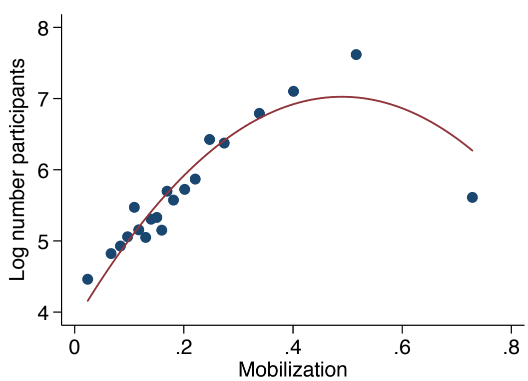


(c) Individuals 18–44 years old

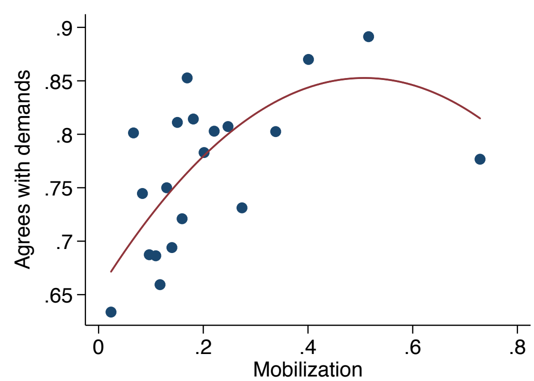


(d) Older than 44 years old

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.



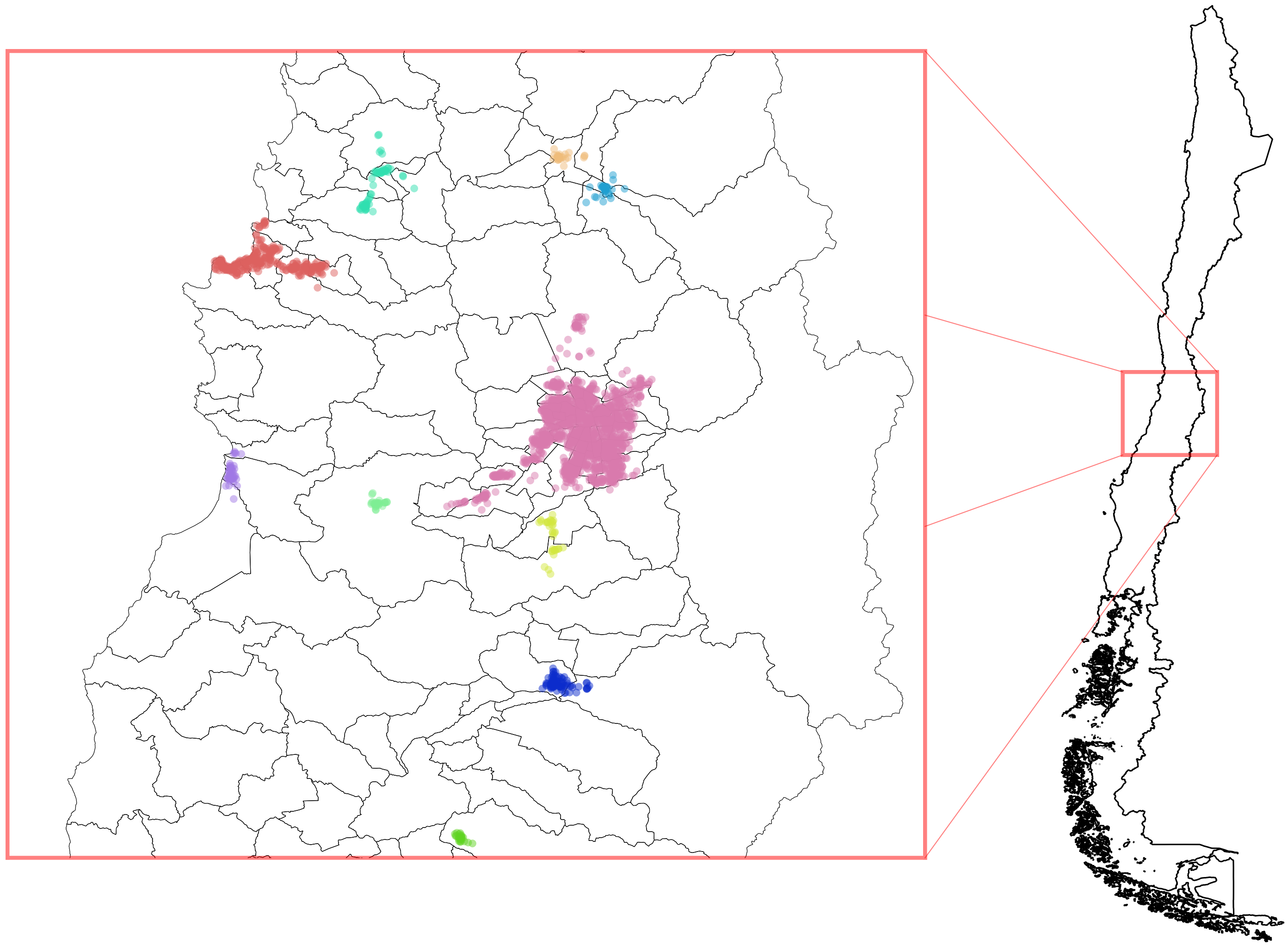
(e) Participation in plebiscite



(f) Agrees with demands

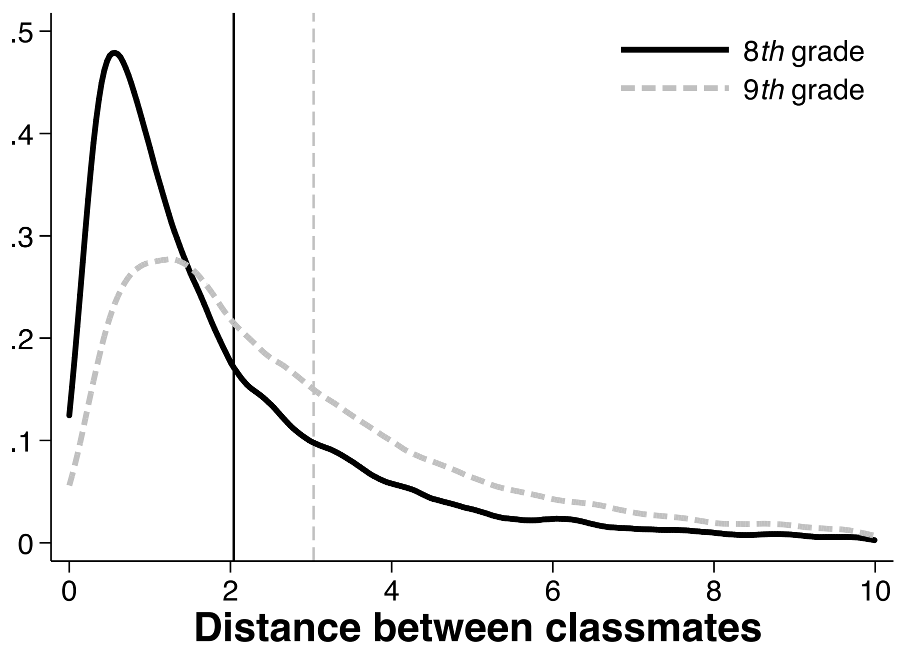
Notes: Panels (e) and (f) plot citizens' participation in the "National plebiscite for education" in October of 2011 at the county level and the percentage of people that agrees with the students' demands among those who participated.

Figure A.7: Cities



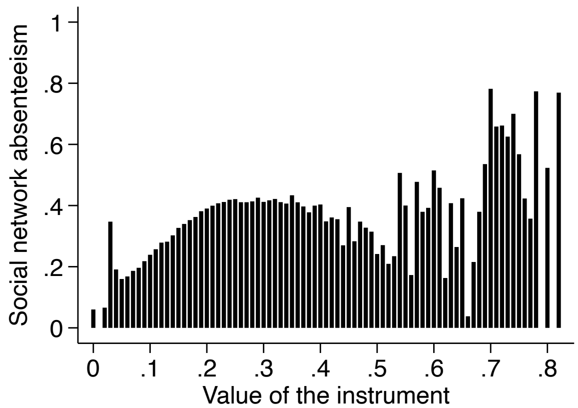
Notes: This map plots the ten largest cities in the most populated area of the country. Cities are defined as closed geographic polygons with schools closer than 5 kilometers.

Figure A.8: Social interactions with past classmates

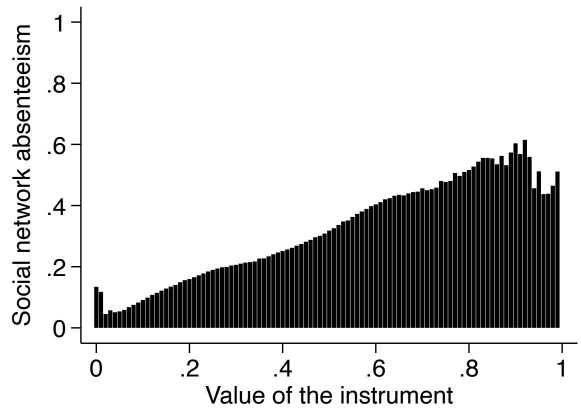


Notes: Distribution of Euclidean distances between the homes of contemporaneous classmates in 8th grade and 9th grade. The *y*-axis measures the density of the distribution and the *x*-axis the distance in kilometers. Each observation corresponds to the average distance between student *i*'s home and the homes of her current classmates. Students' home addresses is administrative data collected by the Ministry of Education. Most students live closer than 1 kilometer from their classmates in 8th grade, implying that they live mostly in the same neighborhood. The average distance between classmates increases by almost 50% from 8th to 9th grade.

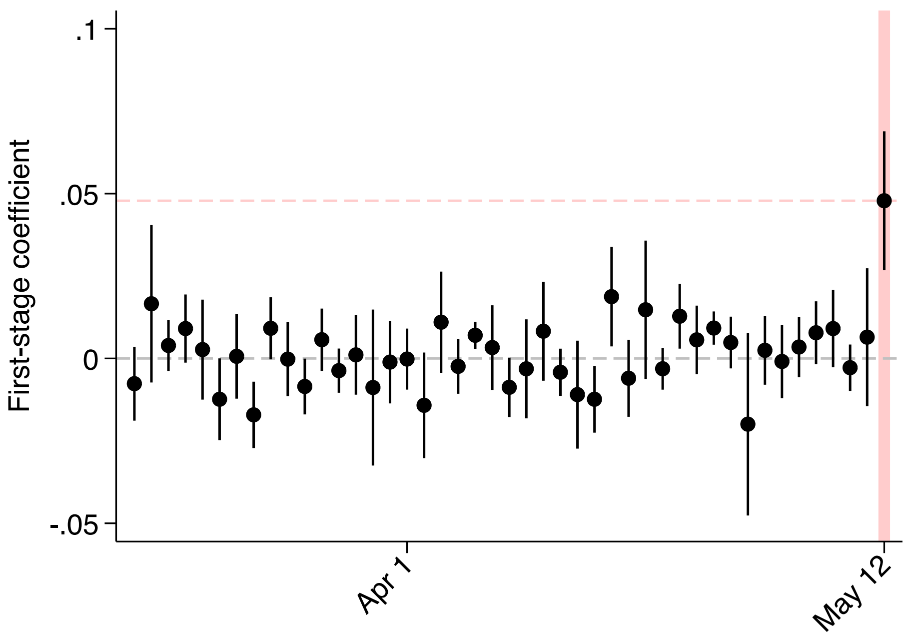
Figure A.9: First-stage



(a) June 16 protest



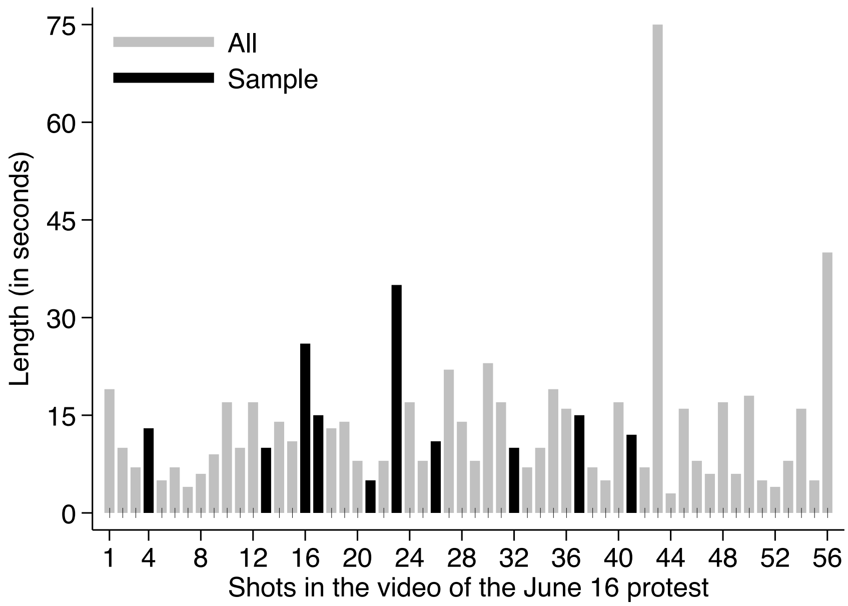
(b) Panel data



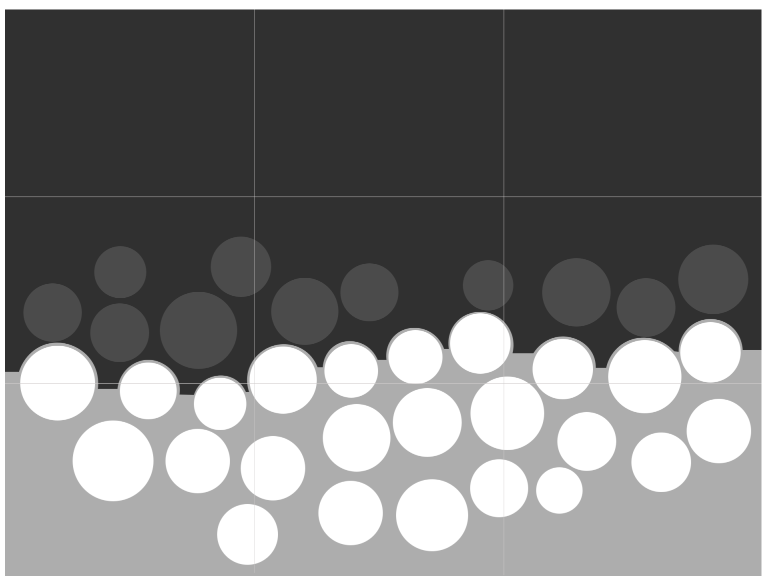
(c) Placebos

Notes: Panels (a) and (b) plot the average social network absenteeism for different values of the instrument in both econometric strategies. Panel (c) 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.10: Crowd counting high-school students in the June 16 protest



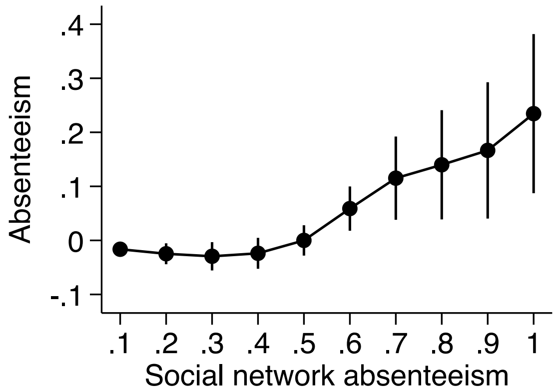
(a) Images in video of the protest



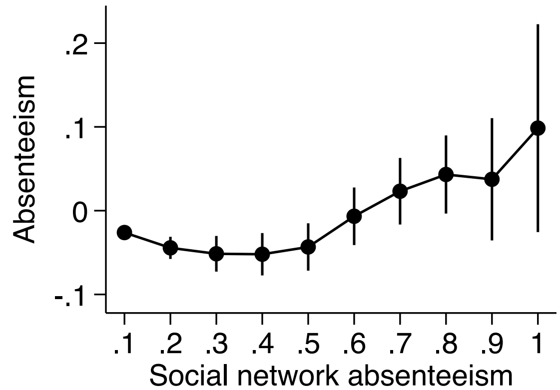
(b) Protocol for crowd-counting

Notes: Panel (a) presents a graphical description of the video of the June 16 rally. The video is composed by 56 shots (*x*-axis) of varying length (*y*-axis, from less than 5 to 75 seconds). Black bars represent the location of the images we use as a sample. Panel (b) shows the sketch of an image, where a crowd is identifiable in the front, and a non-identifiable crowd is located in the back. We asked 100 university students to count the number of high-school students in the front of the image using an economic incentive to do it right. High-schoolers were counted in a total of 520 images.

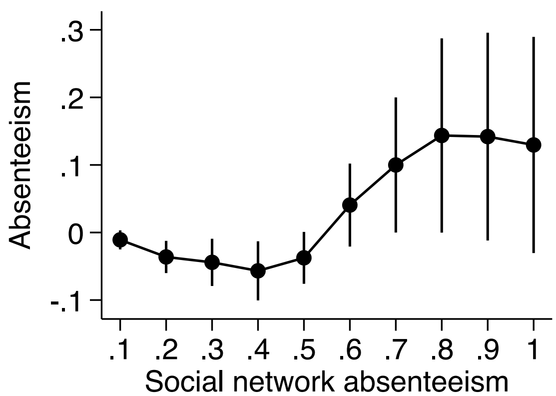
Figure A.11: Reduced-form results in sub-samples



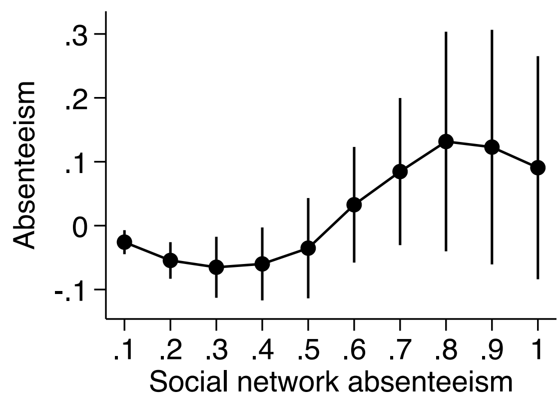
(a) Students in 1st year of high-school



(b) Students in 2nd year of high-school



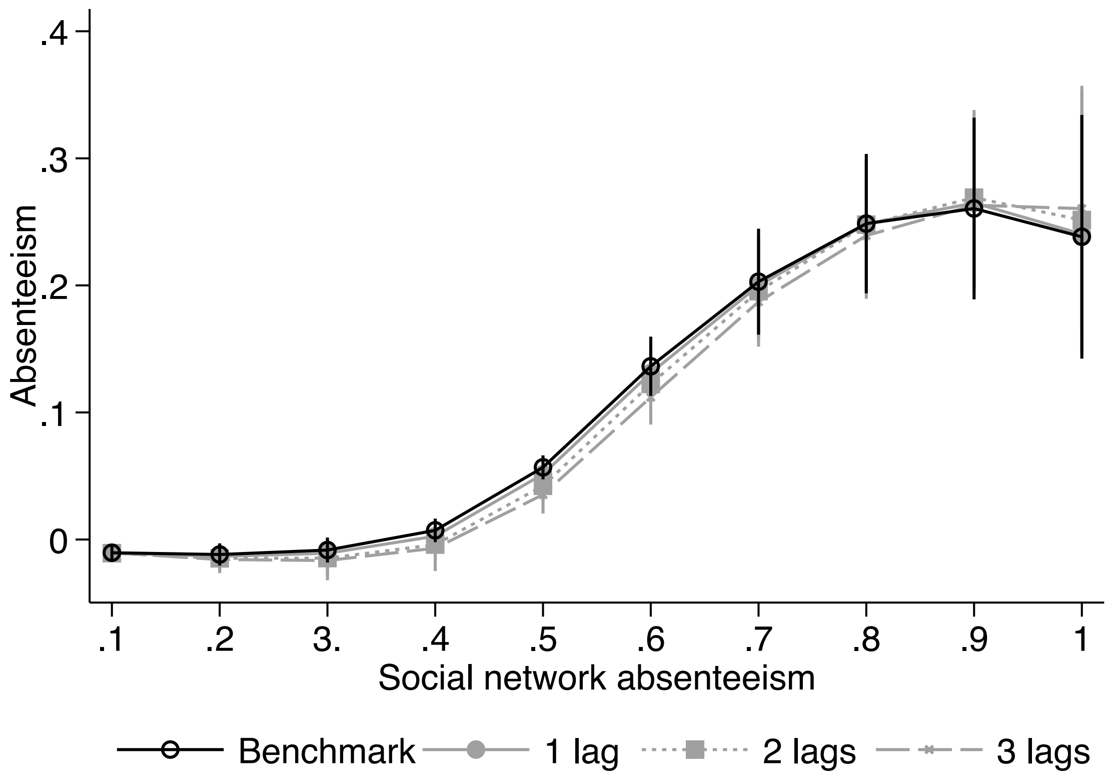
(c) Students in 3rd year of high-school



(d) Students in 4th year of high-school

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

Figure A.12: Panel data specification with lags



Notes: This figure presents β estimates of the following 4 specifications:

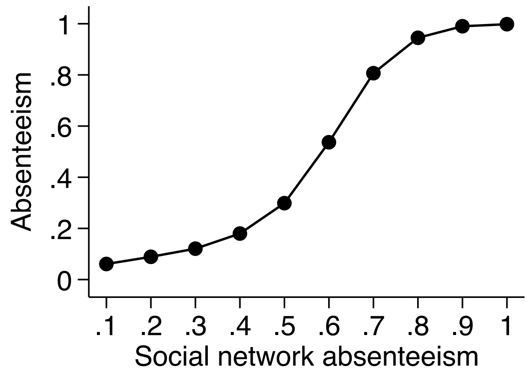
$$\begin{aligned}
A_{isct} &= f(A_{j(i)t}) + \xi_i + \zeta_{ct} + \epsilon_{isct} \\
A_{isct} &= f(A_{j(i)t}) + A_{isc,t-1} + A_{j(i),t-1} + \xi_i + \zeta_{ct} + \epsilon_{isct} \\
A_{isct} &= f(A_{j(i)t}) + A_{isc,t-1} + A_{isc,t-2} + A_{j(i),t-1} + A_{j(i),t-2} + \xi_i + \zeta_{ct} + \epsilon_{isct} \\
A_{isct} &= f(A_{j(i)t}) + A_{isc,t-1} + A_{isc,t-2} + A_{isc,t-3} + A_{j(i),t-1} + A_{j(i),t-2} + A_{j(i),t-3} + \xi_i + \zeta_{ct} + \epsilon_{isct}
\end{aligned}$$

where A_{isct} is an indicator that takes the value of one if student i, who attends school s, located in city c, is absent from school in day t. Similarly $A_{j(i),t} \in [0,1]$ is the percentage of students in i's social network who are absent from school in day t. Finally, the β estimates come from the following parameterization of $f(\cdot)$:

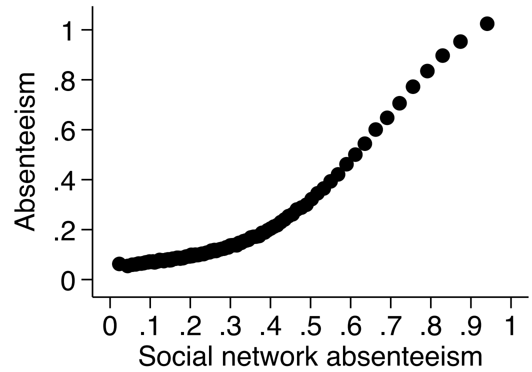
$$f(A_{j(i)}) = \beta_1 \cdot 1\left[\overline{A}_{j(i)} \in [0.1, 0.2)\right] + \cdots + \beta_9 \cdot 1\left[\overline{A}_{j(i)} \in [0.9, 1)\right] + \beta_{10} \cdot 1\left[\overline{A}_{j(i)} = 1\right]$$

where $1[\cdot]$ is an indicator function that takes the value of one when the statement within square brackets is true.

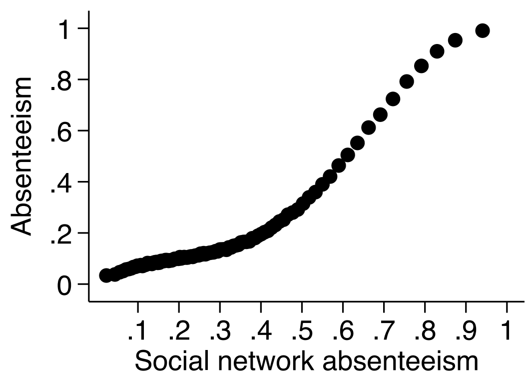
Figure A.13: Robustness of results to estimation method



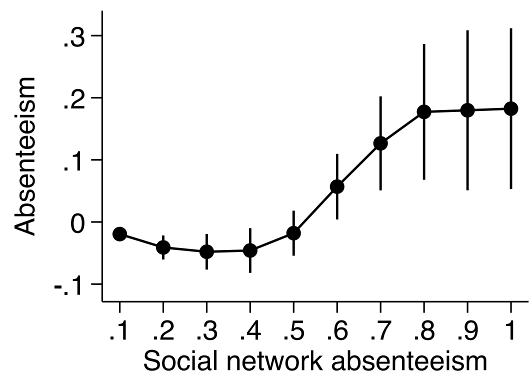
(a) Descriptive bivariate relationship



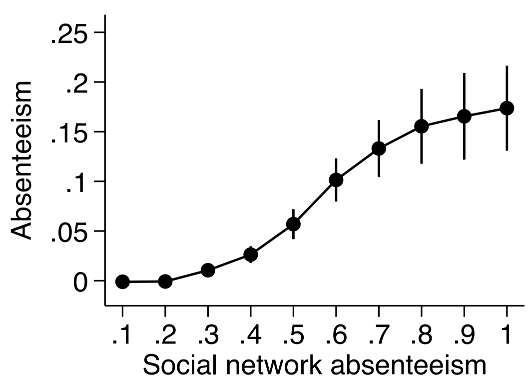
(b) Newey et al. (1999), 3rd degree pol.



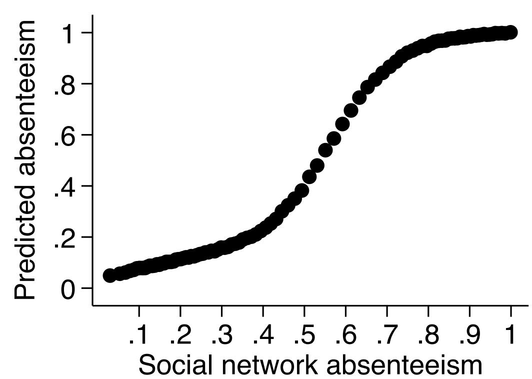
(c) Newey et al. (1999), 4th degree pol.



(d) Newey et al. (1999), robustness



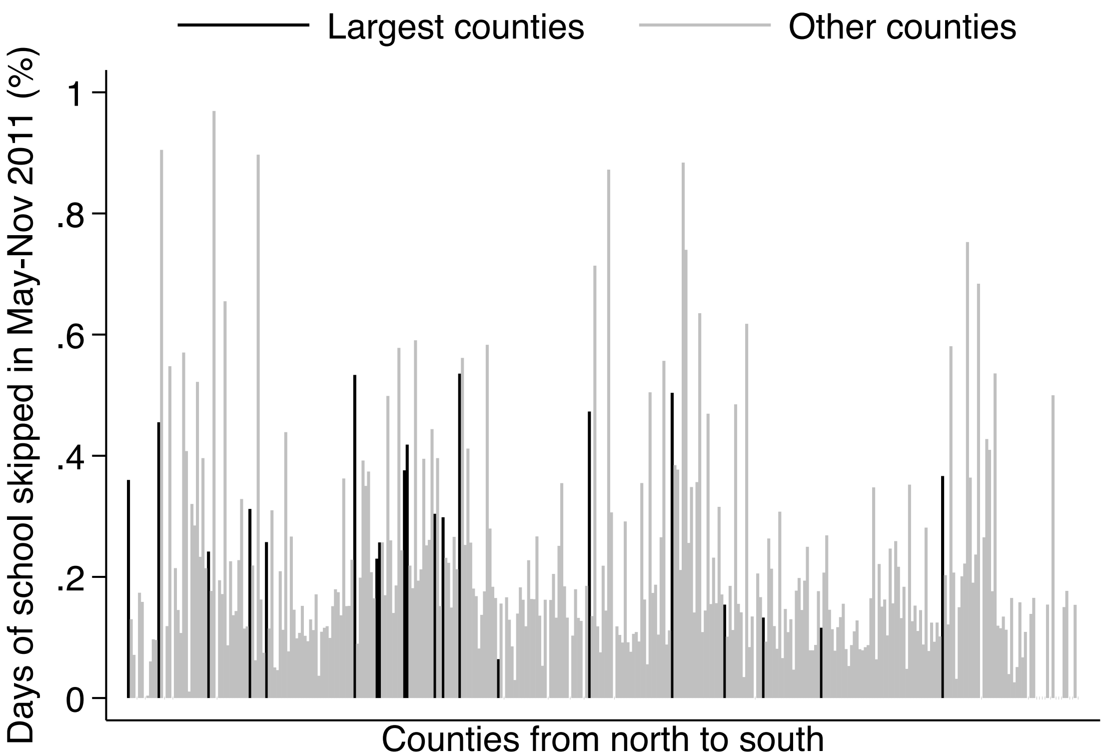
(e) Rau (2013)



(f) Newey and Powell (2003)

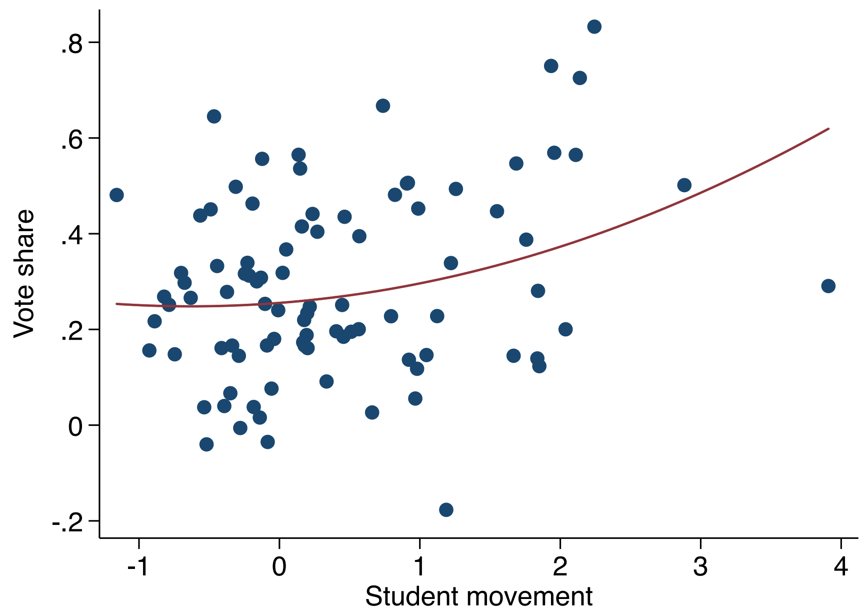
Notes: Each panel presents estimation results from an alternative nonparametric instrumental variables estimation. The exception is panel (a) in which the descriptive bivariate relationship between individual absenteeism and social network absenteeism is plotted. Panels (b), (c), and (f) present predicted values of individual absenteeism, and panels (d) and (e) present regression coefficients associated to indicators of social network absenteeism. More details in section **??**.

Figure A.14: The intensity of the student movement by county

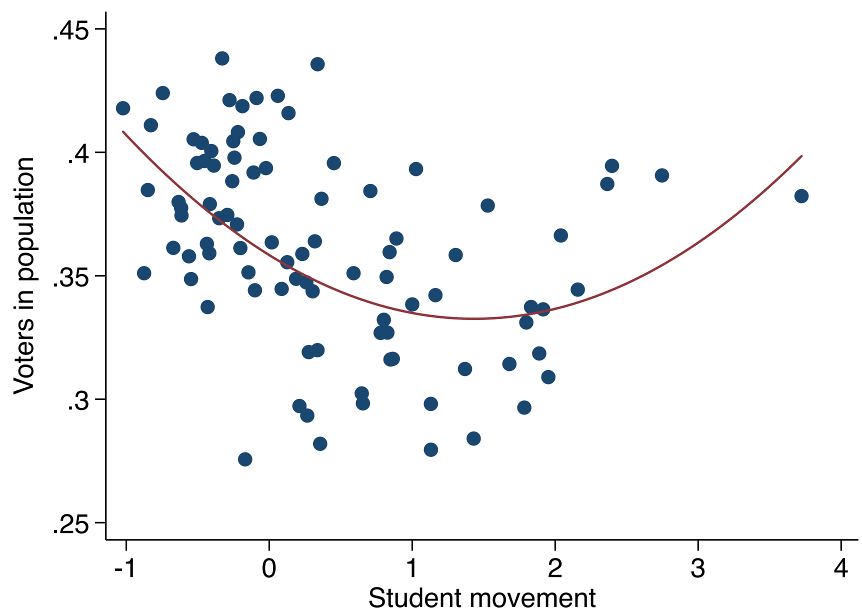


Notes: Own construction based on administrative data. Counties are ordered from north to south in the *x*-axis. The *y*-axis is defined as the percentage of additional days that high-school students skipped school between May and November 2011. There are 324 (out of 346) counties with non-zero intensity.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

Figure A.15: 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.