ONLINE APPENDIX

Collective action in networks: Evidence from the Chilean student movement

Felipe González[1]

List of Tables

A.1 Additional summary statistics and characteristics of compliers iv
A.2 Robustness of 2SLS non-linear estimates (I) v
A.3 Robustness of 2SLS non-linear estimates (II) vi
A.4 Structural estimation, first step vii

List of Figures

A.1 Protests in Chile 1979–2013 viii
A.2 Economic indicators ix
A.3 Panel data estimates using multiple protest days x
A.4 Differential influence within networks xi
A.5 Citizens' evaluation of incumbent politicians xii
A.6 Survey evidence for the impact of the student movement xiii
A.7 Cities xiv
A.8 Social interactions with past classmates xv
A.9 First-stage xvi
A.10 Crowd counting high-school students in the June 16 protest xvii
A.11 Reduced-form results in sub-samples xviii
A.12 Panel data specification with lags xix
A.13 Robustness of results to estimation method xx
A.14 The intensity of the student movement by county xxi
A.15 The student movement and the 2012 local elections xxii

[1]May 2020. Instituto de Economía, Pontificia Universidad Católica de Chile. Contact email: `fagonza4@uc.cl`

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:

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

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

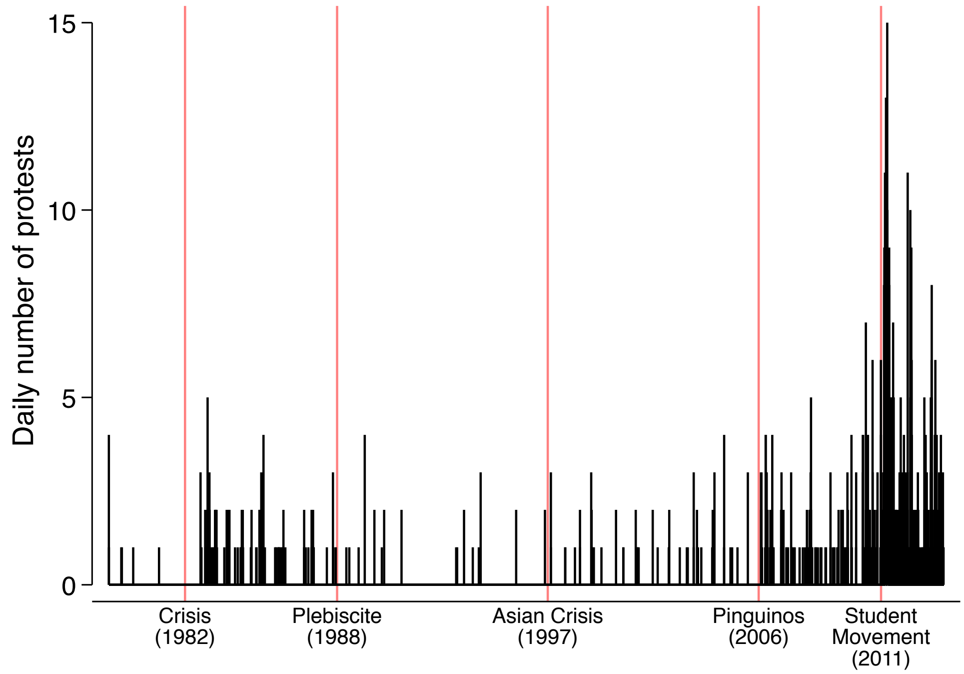
Notes: Each observation corresponds to a student. These estimates correspond to two-stage control function estimates of network effects in school absenteeism on June 16, day of the first massive protest. Standard errors clustered by city are reported in parentheses. Significance level: *** $p < 0.01$,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

Table A.4: Structural estimation, first step

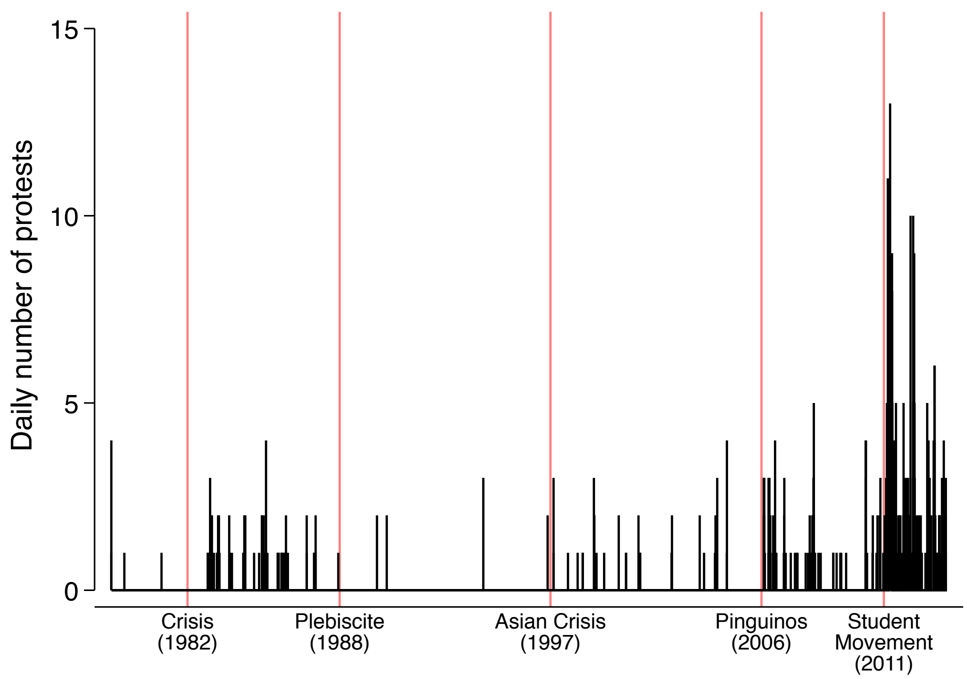
Dependent variable: Indicator skipped school in June 16, 2011		
	MLE I	MLE II
Indicator student skipped school in May 12, 2011	0.64***	0.63***
	(0.01)	(0.01)
Indicator student skipped school in June 1, 2011	0.82***	0.81***
	(0.01)	(0.01)
Indicator student repeated grade in 2010	-0.19***	-0.20***
	(0.03)	(0.03)
Indicator student is female	0.16***	0.16***
	(0.01)	(0.01)
Indicator student switched school in 2010	0.14***	0.13***
	(0.01)	(0.01)
Indicator student age is 14	0.64	0.63
	(0.46)	(0.46)
Indicator student age is 15	0.78*	0.74
	(0.46)	(0.46)
Indicator student age is 16	0.85*	0.84*
	(0.46)	(0.46)
Indicator student age is 17	0.96**	0.95**
	(0.46)	(0.46)
Indicator student age is 18	1.09**	1.06**
	(0.46)	(0.46)
Indicator student age is 19	1.02**	0.99**
	(0.47)	(0.46)
Indicator student age is 20	0.92*	0.89*
	(0.47)	(0.47)
Indicator student age is 21	0.55	0.51
	(0.55)	(0.55)
Students	498,786	498,657
Avg. predicted skipping rate	0.21	0.21
School fixed effects	X	X
Student GPA and attendance in 2010 bins $[x_h]$	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	-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



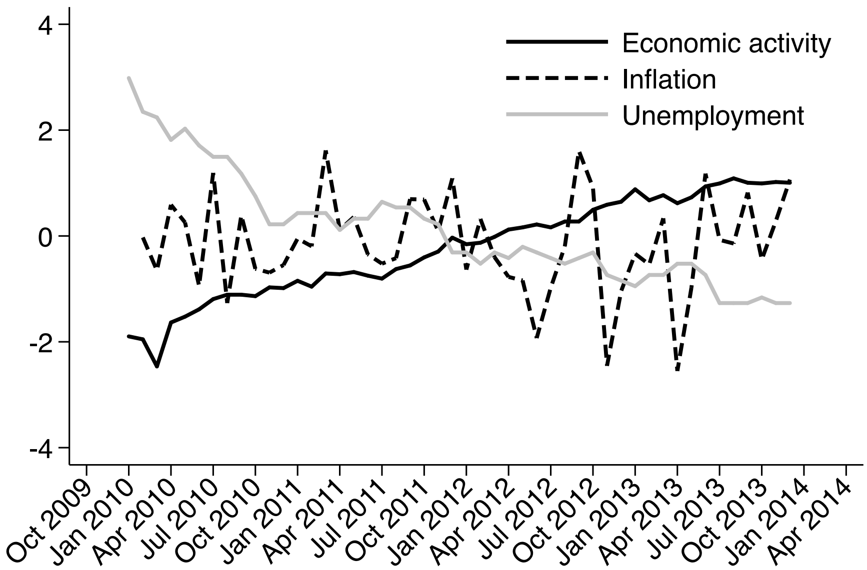
(a) Any type of protest event



(b) Protest events related to education

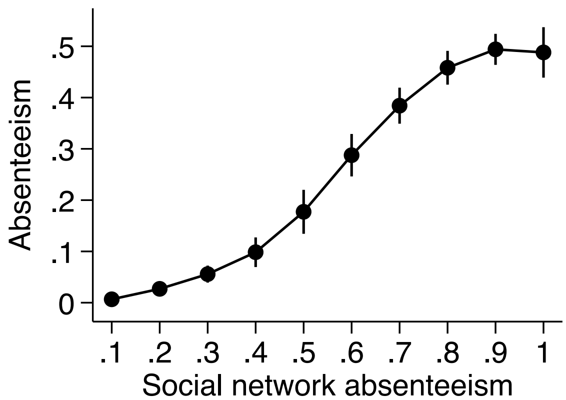
Notes: Data from the Global Dataset of Events, Language, and Tone.

Figure A.2: Economic indicators

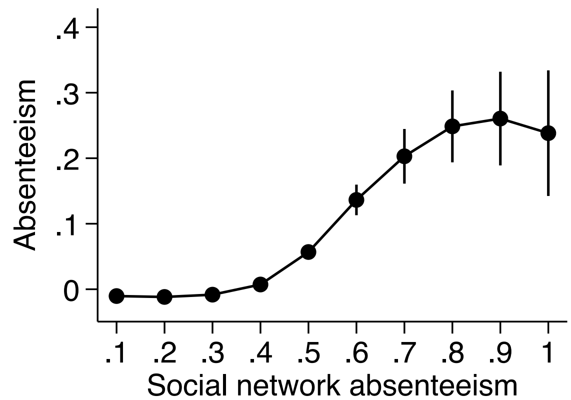


Notes: Data from the Central Bank of Chile. All variables have been normalized by subtracting their average and dividing by their standard deviation in the time series. The vertical red line denotes the beginning of the student movement.

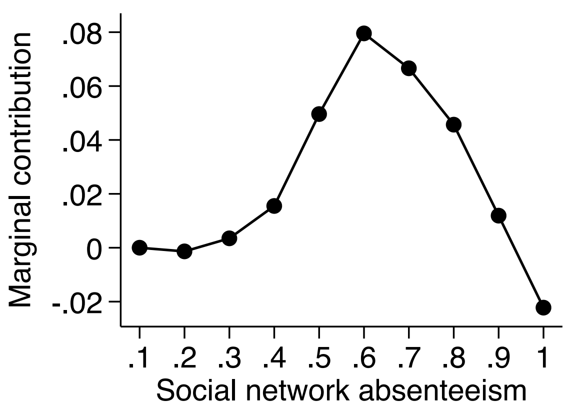
Figure A.3: Panel data estimates using multiple protest days



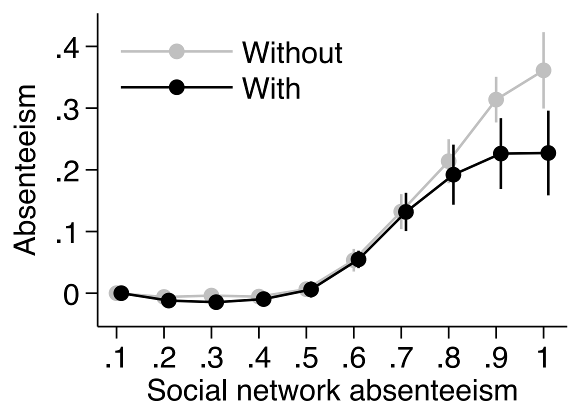
(a) OLS estimates



(b) 2SLS estimates



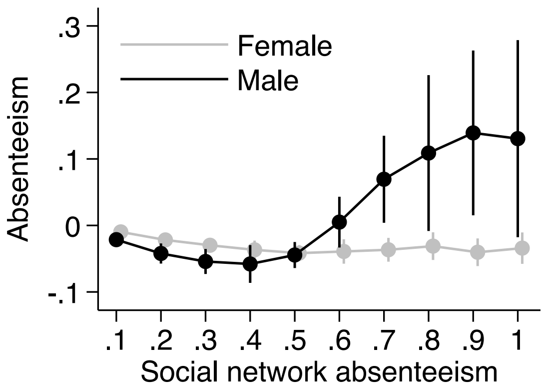
(c) Marginal contribution



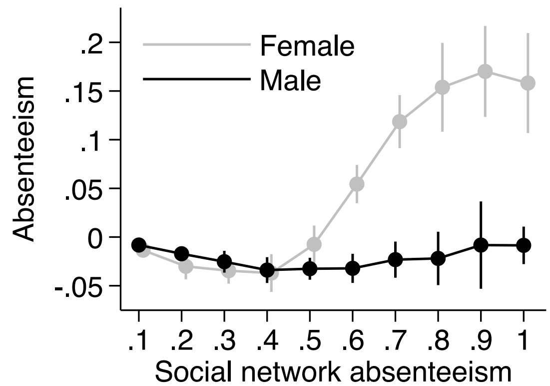
(d) 2G-3G coverage network

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).

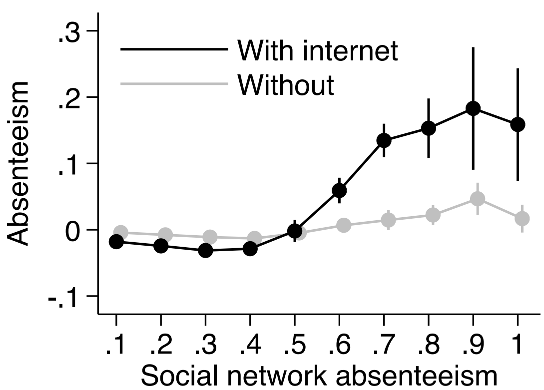
Figure A.4: Differential influence within networks



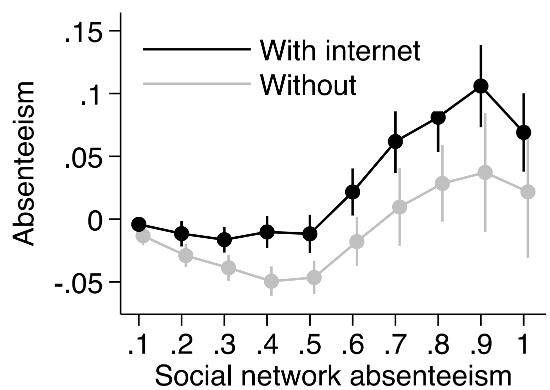
(a) Subsample of males



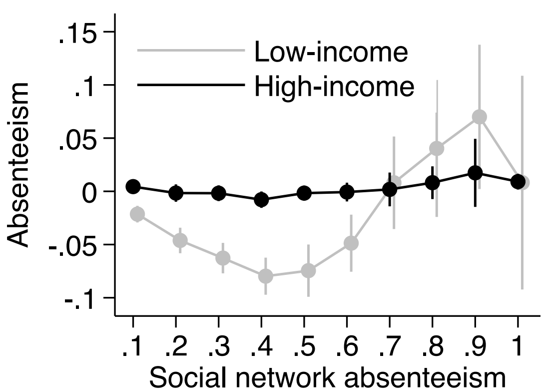
(b) Subsample of females



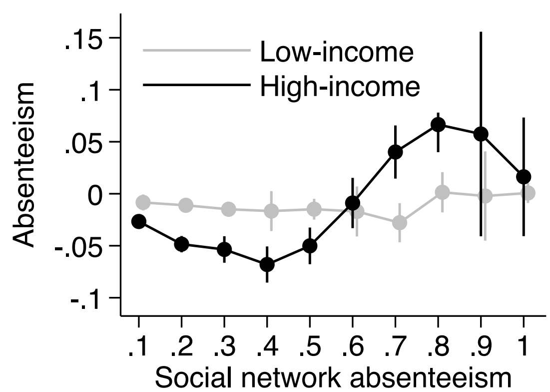
(c) Subsample with internet



(d) Subsample without internet



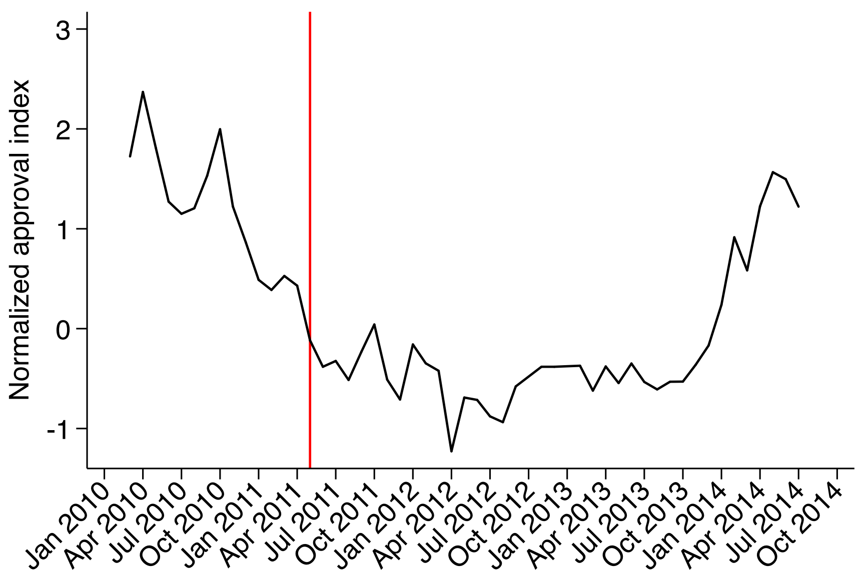
(e) Low-income households



(f) High-income households

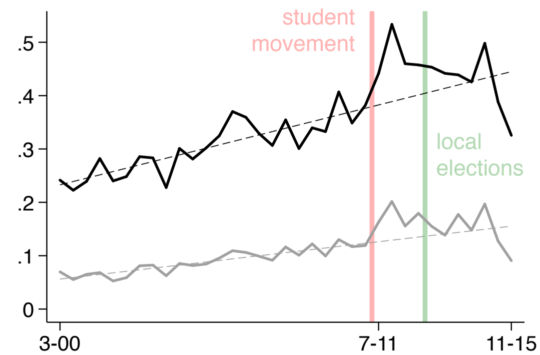
Notes: All panels plot 2SLS estimates from a regression of individual school absenteeism on 10 indicators of network absenteeism, controlling for student and network characteristics, and school fixed effects. Regressions are in sub-samples and split the network in groups.

Figure A.5: Citizens' evaluation of incumbent politicians

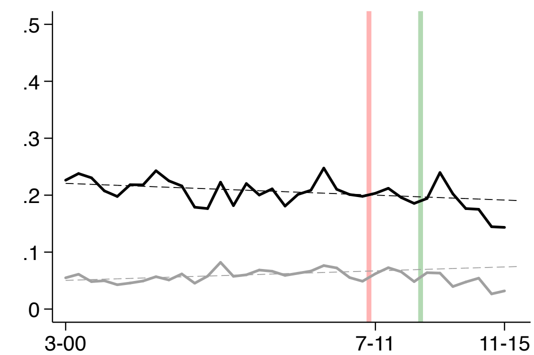


Notes: Normalized index (minus average and divide by standard deviation) for the approval of incumbent politicians. Data from the Centro de Estudios Públicos and Adimark.

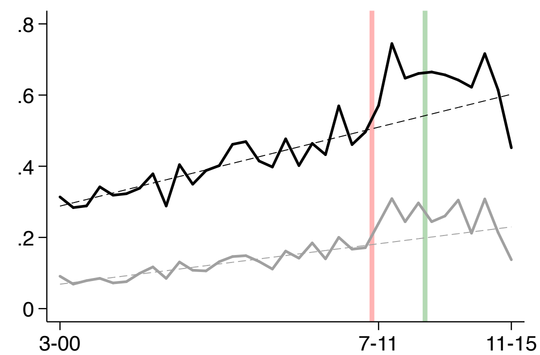
Figure A.6: Survey evidence for the impact of the student movement



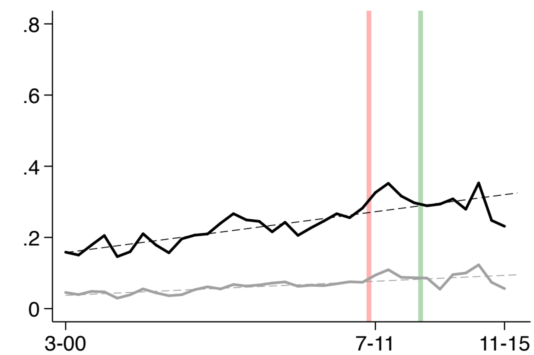
(a) Education should be priority



(b) Placebo

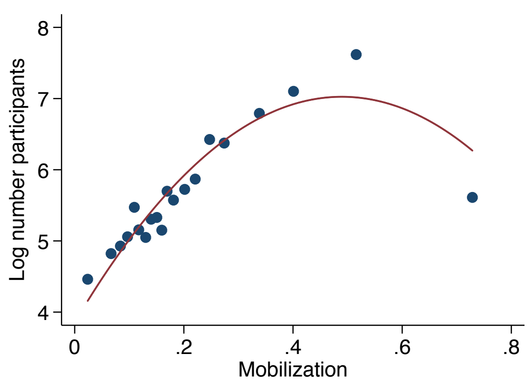


(c) Individuals 18–44 years old

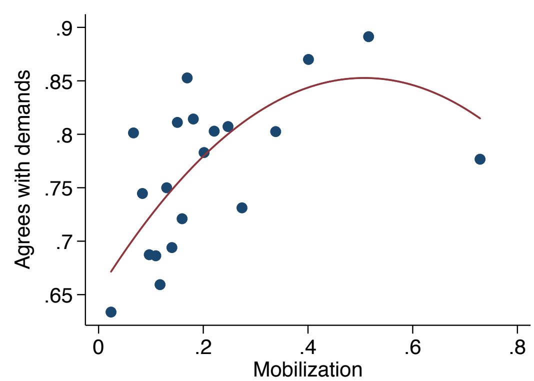


(d) Older than 44 years old

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.



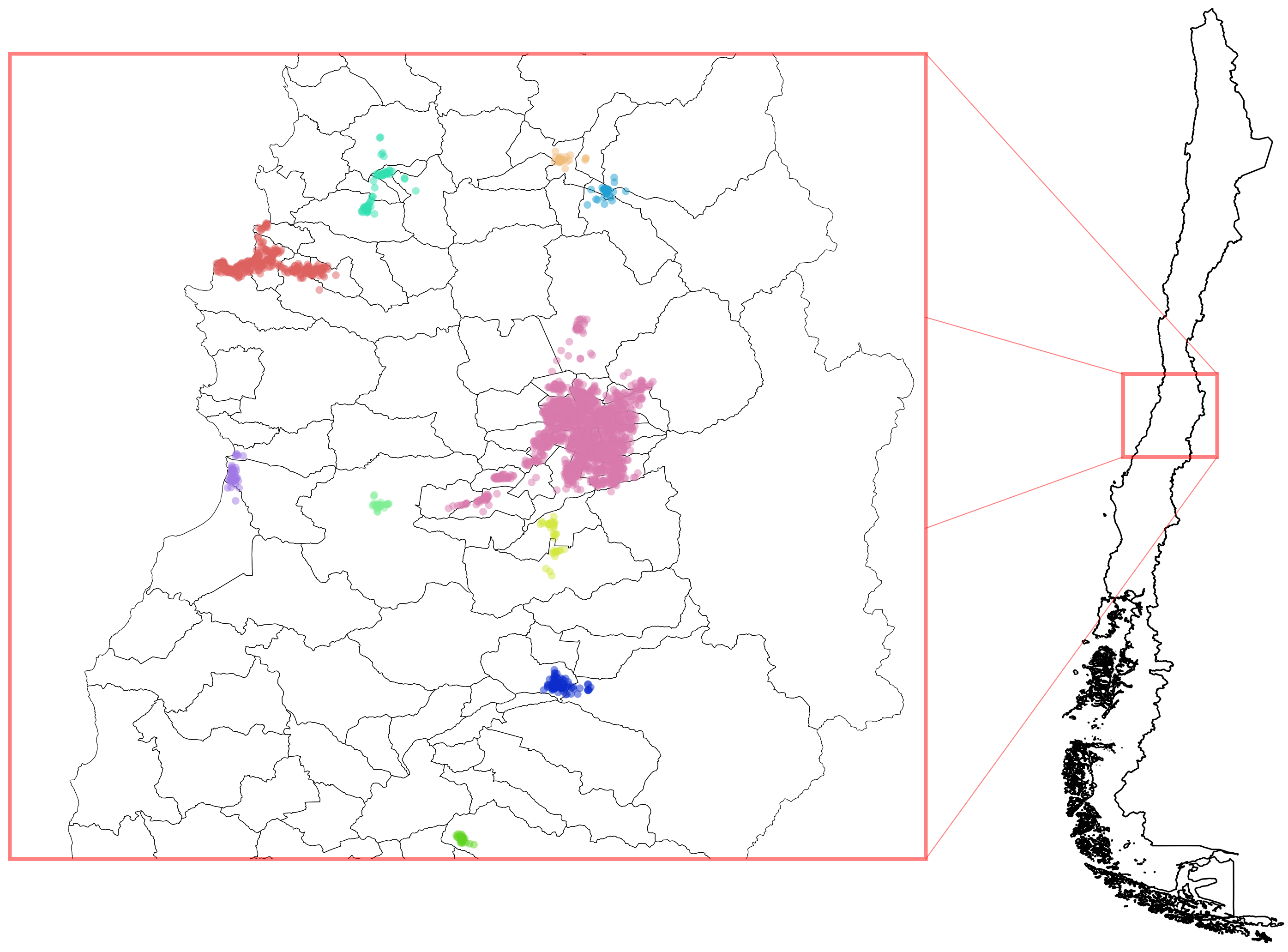
(e) Participation in plebiscite



(f) Agrees with demands

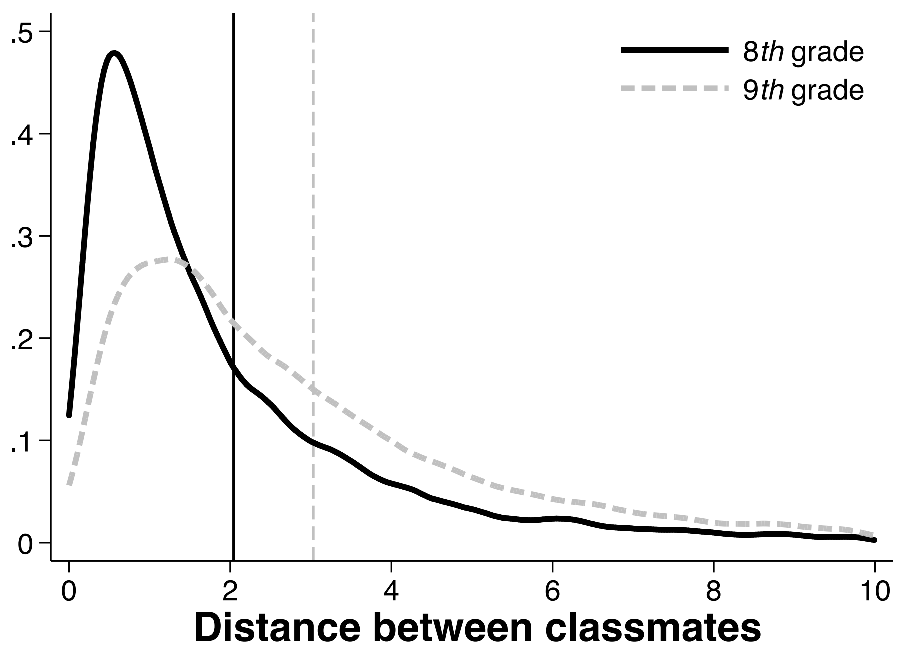
Notes: Panels (e) and (f) plot citizens' participation in the "National plebiscite for education" in October of 2011 at the county level and the percentage of people that agrees with the students' demands among those who participated.

Figure A.7: Cities



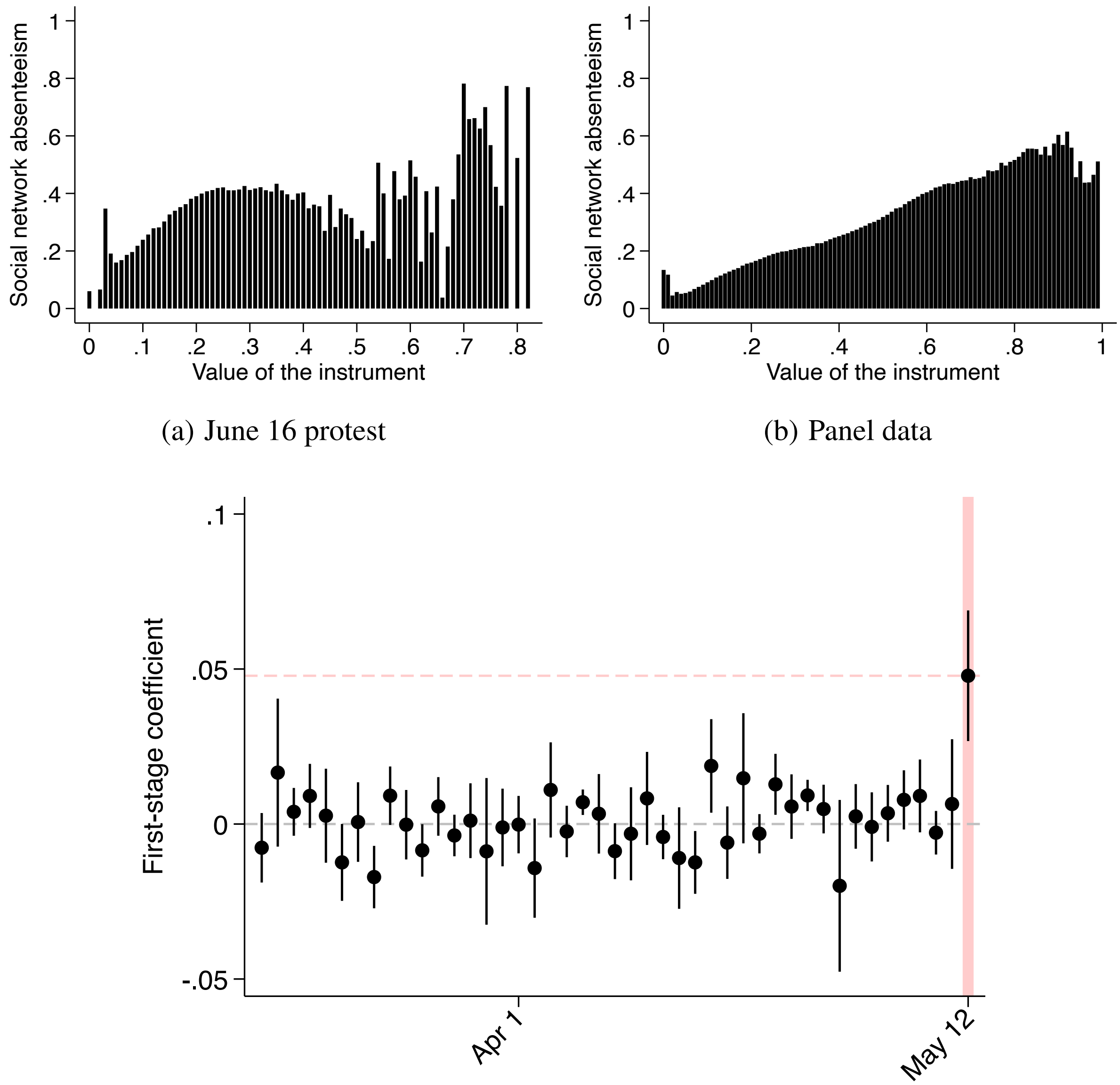
Notes: This map plots the ten largest cities in the most populated area of the country. Cities are defined as closed geographic polygons with schools closer than 5 kilometers.

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



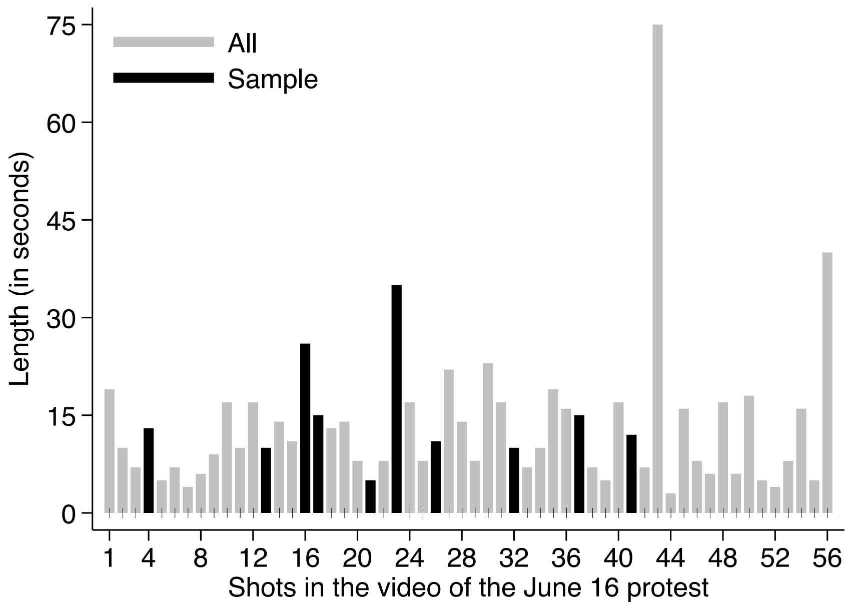
(a) June 16 protest

(b) Panel data

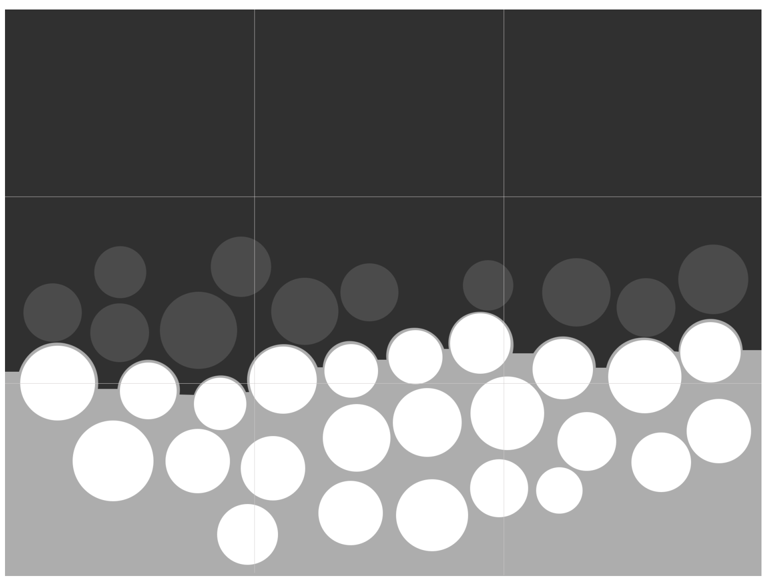
(c) Placebos

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



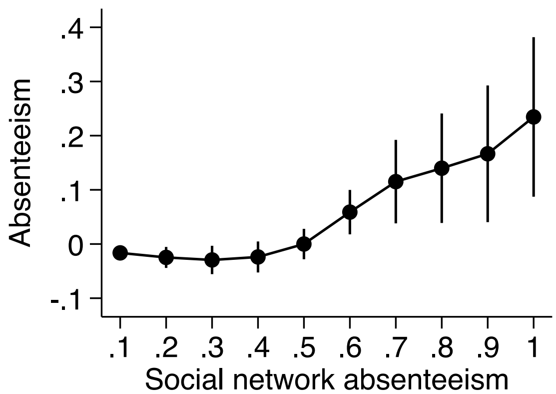
(a) Images in video of the protest



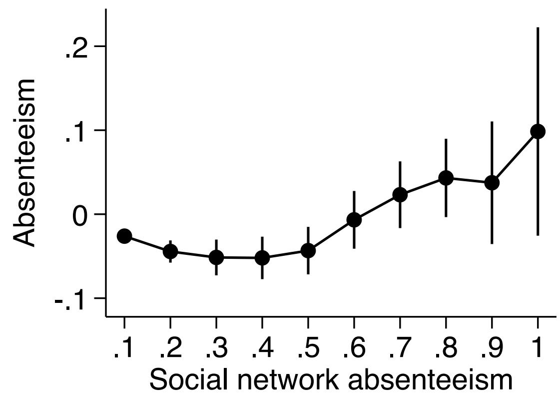
(b) Protocol for crowd-counting

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.

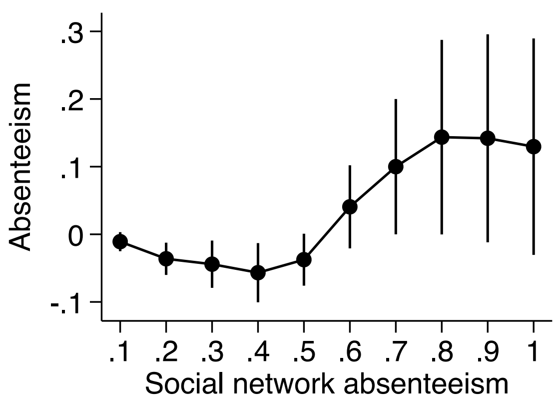
Figure A.11: Reduced-form results in sub-samples



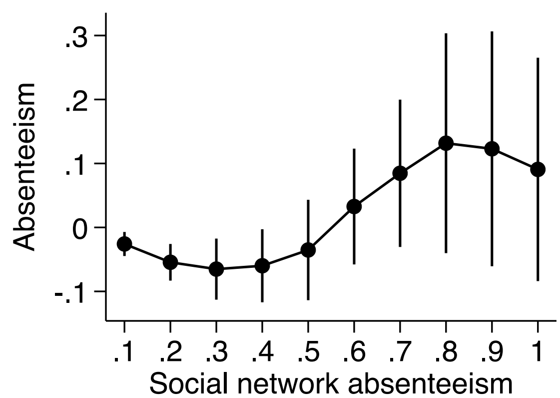
(a) Students in 1st year of high-school



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



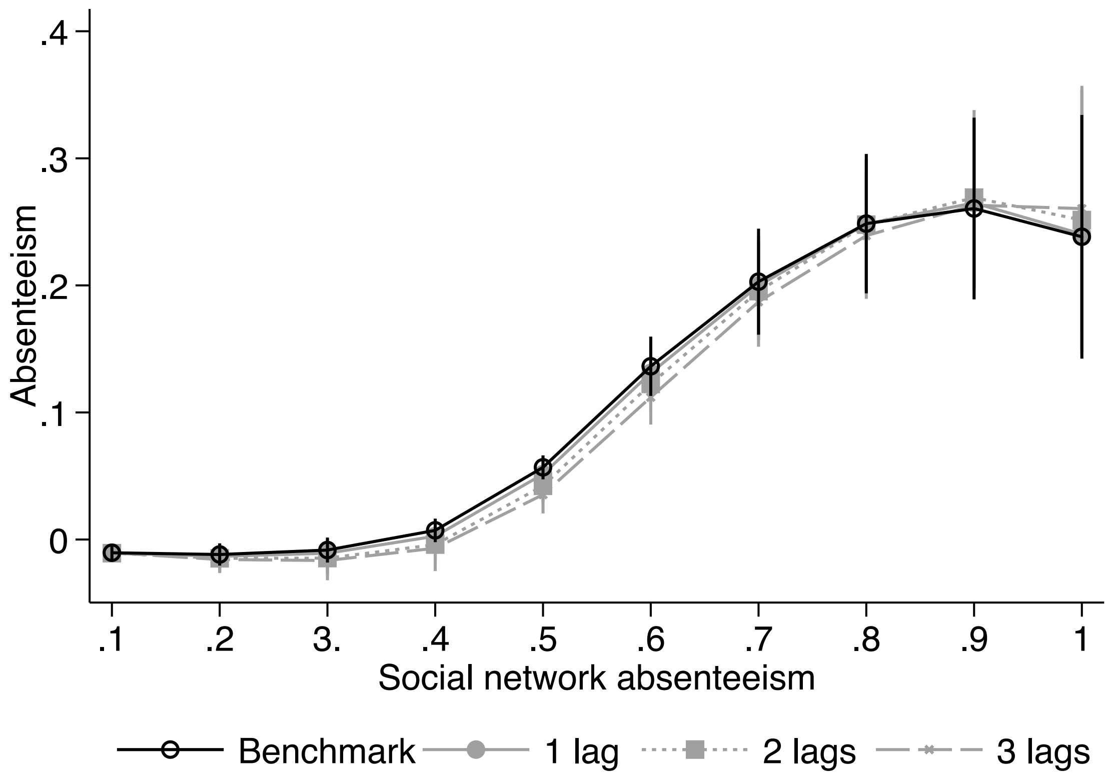
(c) Students in 3rd year of high-school



(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:

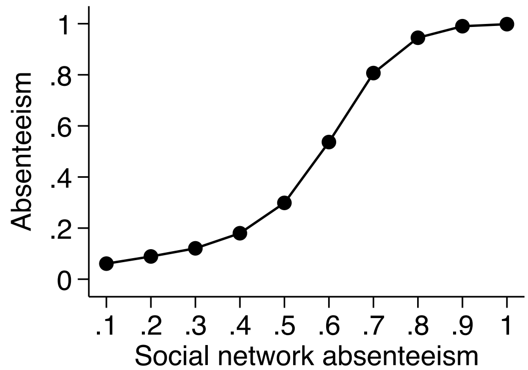
$$
\begin{aligned}
A_{isct} &= f(A_{j(i)t}) + \xi_i + \zeta_{ct} + \epsilon_{isct} \\
A_{isct} &= f(A_{j(i)t}) + A_{isc,t-1} + A_{j(i),t-1} + \xi_i + \zeta_{ct} + \epsilon_{isct} \\
A_{isct} &= f(A_{j(i)t}) + A_{isc,t-1} + A_{isc,t-2} + A_{j(i),t-1} + A_{j(i),t-2} + \xi_i + \zeta_{ct} + \epsilon_{isct}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

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 $f(\cdot)$:

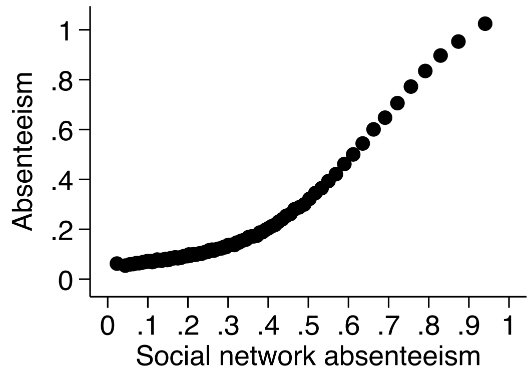
$$f(A_{j(i)}) = \beta_1 \cdot 1\left[\overline{A}_{j(i)} \in [0.1, 0.2)\right] + \cdots + \beta_9 \cdot 1\left[\overline{A}_{j(i)} \in [0.9, 1)\right] + \beta_{10} \cdot 1\left[\overline{A}_{j(i)} = 1\right]$$

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

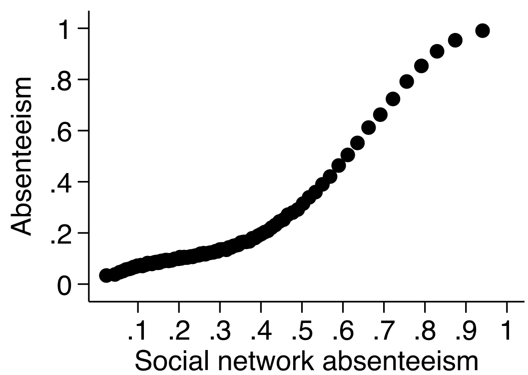
Figure A.13: Robustness of results to estimation method



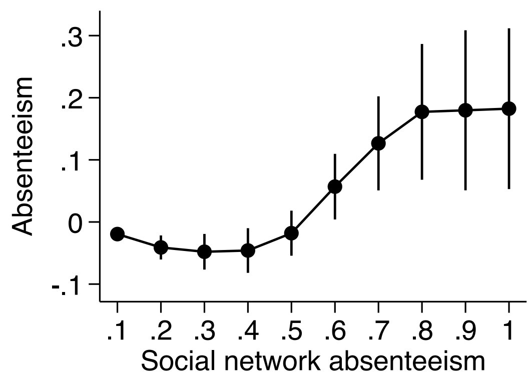
(a) Descriptive bivariate relationship



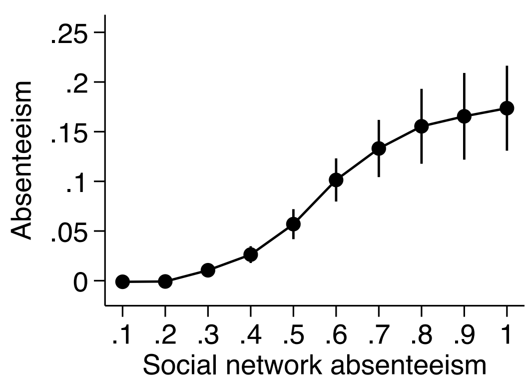
(b) Newey et al. (1999), 3rd degree pol.



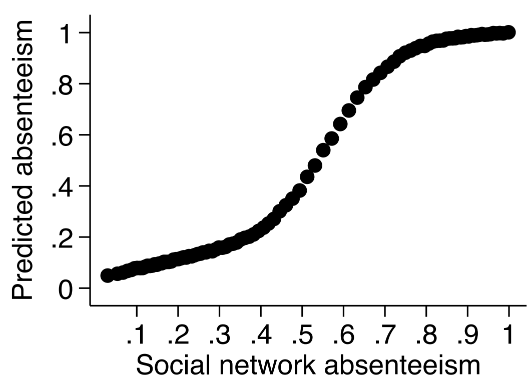
(c) Newey et al. (1999), 4th degree pol.



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



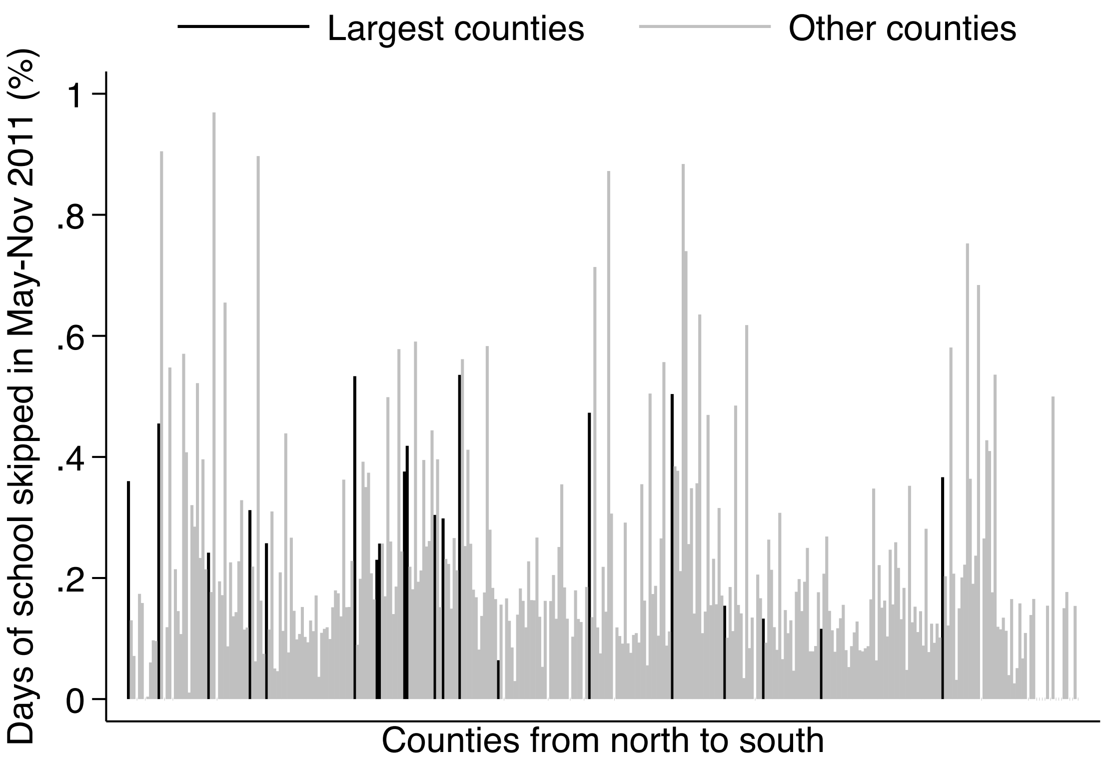
(e) Rau (2013)



(f) Newey and Powell (2003)

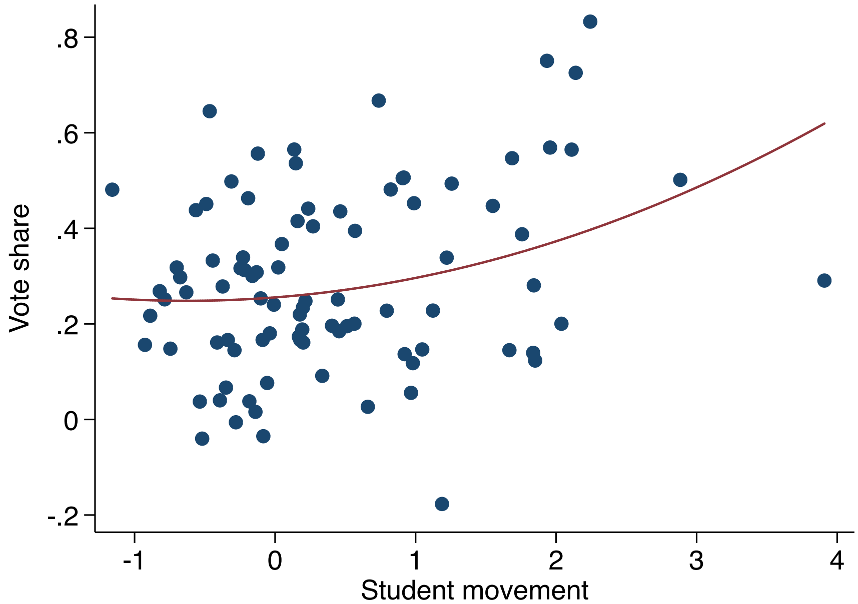
Notes: Each panel presents estimation results from an alternative nonparametric instrumental variables estimation. The exception is panel (a) in which the descriptive bivariate relationship between individual absenteeism and social network absenteeism is plotted. Panels (b), (c), and (f) present predicted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

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

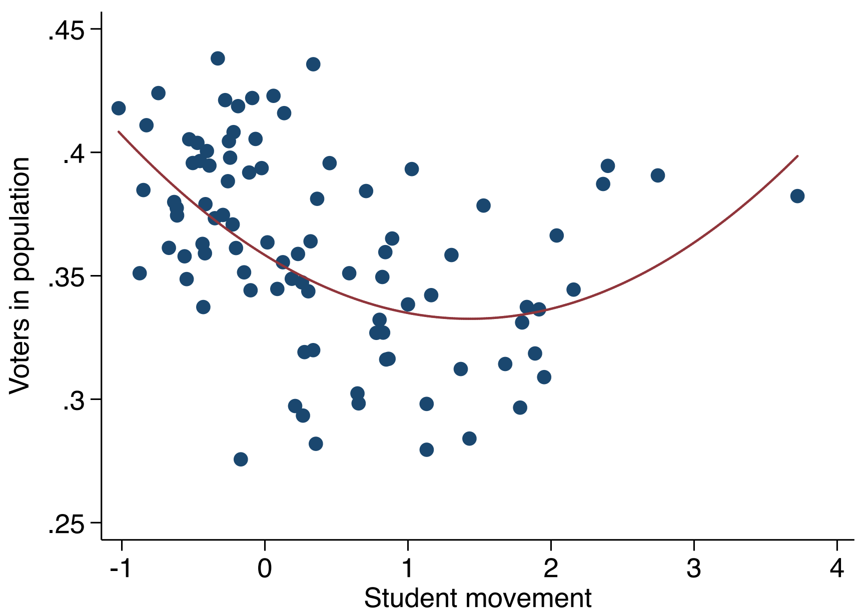


Notes: Own construction based on administrative data. Counties are ordered from north to south in the *x*-axis. The *y*-axis is defined as the percentage of additional days that high-school students skipped school between May and November 2011. There are 324 (out of 346) counties with non-zero intensity.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