Collective Action in Networks: Evidence from the Chilean Student Movement

Felipe González*†

Abstract. Organized groups of individuals challenging the status quo are critical for institutional change and economic development patterns. This paper studies the 2011 student movement in Chile, the largest protest mobilization in the country’s history, in which hundreds of thousands of students skipped school to protest with the goal of reforming the educational system. Using administrative data on millions of students’ daily school attendance decisions on protest and non-protest days, a large network composed by the lifetime history of classmates, and differential network exposure to the first national protest, I employ an instrumental variables approach to test how networks affect protest behavior. The main finding is that individual participation follows a threshold model of collective behavior: students were influenced by their networks to skip school on protest days only when more than 40 percent of the members of their networks also skipped school. Additional findings show that protest participation imposed significant educational costs on students and helped to shift votes towards non-traditional opposition parties. Taken together, results indicate that networks amplify the effect of protests in non-linear ways with potentially significant consequences for institutional change.

Keywords: collective action, networks, protests

*Job Market Paper. This version: October 31, 2016. Please click here for the most recent version.
†UC Berkeley, Department of Economics. Email: fgonzalez@econ.berkeley.edu. I would like to thank my advisors Fred Finan, Solomon Hsiang, Edward Miguel, and Suresh Naidu for their support and guidance. I also thank Ernesto Dal Bó, Supreet Kaur, Gerard Roland, Joachim Voth, Noam Yuchtman, and seminar participants at UC Berkeley for comments and suggestions.
1 Introduction

Throughout history, organized groups of individuals have challenged the status quo and achieved significant social, economic, and political transformations. From the French Revolution to the Arab Spring, examples of groups aiming to transform societies are abundant. These organized groups are critical for institutional change and therefore, economic development patterns.

Individual participation in collective action has puzzled social scientists due to the presence of common benefits and private costs. This “collective action problem” has given rise to a large amount of theoretical literature emphasizing that the actions of others are crucial in order to understand individual participation.\(^1\) However, an empirical investigation of how individual participation responds to the participation of others is still lacking. Given the enormous data requirements, the lack of evidence is not entirely surprising.

This paper studies the 2011 student movement in Chile, the largest protest mobilization in the country’s history, in which hundreds of thousands of students skipped school to protest with the goal of reforming the educational system. I employ an instrumental variables approach to test for the role of networks in protest behavior using administrative data on millions of students’ daily school attendance decisions on protest and non-protest days, a large network composed by the lifetime history of classmates, and differential network exposure to the first national protest.

The main finding is that participation in the student movement followed a pattern consistent with a threshold model of collective behavior. Students were influenced by their networks to skip school in protest days only when more than 40 percent of the members of their networks also skipped school, creating a bimodal distribution in participation across groups. Skipping school imposed significant costs on students but it also shifted vote shares towards non-traditional opposition parties in the 2012 local elections. Taken together, results indicate that networks amplify the effect of protests in non-linear ways with potentially significant consequences for institutional change.

To organize the empirical analysis, I begin by presenting a simple framework that focuses on the individual decision to participate in a social movement. The participation

---

\(^1\)See Olson (1965), Granovetter (1978), Tilly (1978), Kuran (1989), Lohmann (1993), and Marwell and Oliver (1993), among many others.
decision has an individual cost and depends on aggregate participation and the participation of others in an individual’s network. More participation in the network could decrease or increase individual participation in a linear or non-linear manner. The framework emphasizes a potential differential influence within a network and the possibility of multiple networks affecting the participation decision.

Two key features of the Chilean student movement allow me to empirically study participation in collective action. First, the central government assembles an exceptionally rich dataset of daily school attendance. Thus, I can measure participation in the movement for more than 800,000 high-school students using school absenteeism on days of protest. Second, students interact primarily with classmates (Araos et al., 2014), and information about their lifetime history of classmates is available. The latter data allows me to construct a country-wide network with more than 600 billion potential interactions and more than 60 million links between students who have shared a classroom.

The empirical analysis is divided in five parts. The first part uses an instrumental variables approach to estimate how network participation affects individual participation in the context of Manski (1993) “linear-in-means” model. The second part deviates from the linear model and tests for Granovetter (1978) “threshold model of collective behavior” using the non-parametric estimation proposed by Newey et al. (1999). The third part tests for the possibility of weak and strong ties embedded into the threshold model. The fourth part tests for additional influence from the network of neighbors. The fifth part estimates the effect of participation in the movement on students’ academic performance and also on electoral outcomes in the 2012 local elections.

A crucial element in the analysis is the use of an instrument that solves the simultaneity of decisions, the possibility of unobservable variables causing a spurious positive correlation between students and their networks, and potential measurement error. The instrument is the exposure of networks to the first protest in May 12, organized by college students. Exposure is measured as absenteeism in May 12 in the set of students that (1) are part of the network of networks, and (2) are attending a different school in 2011. The instrument is similar to the one proposed by Bramouillé et al. (2009) and De Giorgi et al. (2010), although it is a refinement in the sense that it uses variation across protest days and focuses on students in different schools.

Network exposure to the first protest in May 12 is highly predictive of participation in June 16 – the first large national protest – even after controlling for a large set of
observable variables for students, networks, schools, and city fixed effects. This is the
preferred specification, but all results are qualitatively robust to the inclusion of fixed
effects at the neighborhood, school, or school-grade level. Placebo checks using non-
protest days confirm the importance of the May 12 protest.

Using the linear-in-means model, the estimates suggest that a 10 percent increase in
network participation increases individual participation by 7 percent. This result is ro-
 bust, statistically precise, and smaller in magnitude than a naive estimation that does not
address endogeneity. However, a non-linear estimation reveals that a threshold model is
a better representation of individual participation decisions. If the share of students in
the network that participates in the movement is lower than 40 percent, the individual
decision to participate is not affected by the network. After this threshold, individual
participation increases rapidly with network participation. This result suggests that a
“critical mass” of individuals is needed to facilitate participation.²

The critical mass of 40 percent should be interpreted as an average threshold. Stu-
dents in larger schools, smaller networks, and smaller cities have lower thresholds. In
addition, augmenting the estimation to allow for differential non-linear effects within
networks, the estimates suggest that students are more influenced by others that are
similar to them, a result that I interpret as evidence of weak and strong ties. Taken to-
gether, the results reject a linear-in-means model and suggest a critical mass of others
that are similar is needed to foster participation. As a consequence, participation across
network groups in the country follows a bimodal distribution with low and high levels
of participation.

An exploration of additional sources of influence from students in schools close by
confirms the previous results in two different ways. First, results are similar using a
different identification assumption. These “multinetwork” results rely on network ex-
posure to the first protest but across (instead of within) networks: participation among
students in social networks is instrumented with exposure in spatial networks. To the
best of my knowledge this “cross-network” identification is novel and potentially use-
ful in other settings. Second, potential non-random measurement error in the network
of students could introduce bias (Laumann et al., 1983; Kossinets, 2006; Chandrasekhar
and Lewis, 2011). Reassuringly, results are robust to incorporate “neighbor” students

²This 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 (1971) dynamic model of segregation.
into the analysis.

The last part of the analysis studies the consequences of protests. In particular, the focus is on the private costs of participation and the effects of the movement on electoral outcomes. A differences-in-differences analysis among students aged 6–10 and high-school students in the period 2002–2015 reveals that grade repetition increased by 60 percent, from a base of 6 percent, among high-school students in 2011. Using within school variation in 2011, I estimate that participation in the June 16 national protest decreased GPA by 0.1 standard deviations and increased grade repetition by 33 percent. Remarkably, these private costs of participation resemble the “critical mass” patterns previously discussed.

In addition to the private costs that participation had for students, I provide suggestive evidence that the student movement was able to shift votes towards non-traditional opposition parties, which were relatively more aligned with the movement’s demands. A cross-sectional regression using county-level electoral data suggests that a one standard deviation increase in the intensity of the movement in local schools increased vote shares for non-traditional parties by 5–10 percentage points. Interestingly, although arguably speculative, the effect on vote shares seems to be non-linear, a result consistent with the previously described “critical mass” patterns of participation.

This paper contributes to the empirical understanding of participation in collective action. Only a few number of articles have studied protest participation. Using an annual panel dataset of geographic cells in Africa, Manacorda and Tesei (2016) show that mobile phone coverage facilitated protests when countries experienced economic downturns. Enikolopov et al. (2016) show that the penetration of an online social network in Russia increased the probability of a protest and the number of protesters across cities. Finally, Acemoglu et al. (2016) show that citizens’ discontentment on Twitter predicts daily protest participation in Tahir Square during Egypt’s Arab Spring. To the best of my knowledge this is the first paper to estimate how individual specific networks affect individual participation in collective action.

This paper also speaks to a related literature estimating the consequences of protests.

---

3There 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.

4There is, of course, a large theoretical literature studying social unrest and political transformation. See, for example, Acemoglu and Robinson (2000).
Madestam et al. (2013) uses rainfall shocks as exogenous variation affecting the number of protesters from the Tea Party Movement across U.S. counties to show how the movement affected electoral outcomes and the policies being implemented. Aidt and Franck (2015) show that the Swing riots in early 19th century Britain – credible signals of the threat of a revolution – facilitated democratic reforms. This paper contributes to this literature by providing novel evidence on the individual costs associated with participation and suggestive evidence on the effect of the student movement on electoral outcomes.

The next section presents a theoretical framework for the individual decision to participate in collective action. Section 3 provides details about the 2011 student movement in Chile. Section 4 presents the data and describes participants. Section 5 tests for different models of participation in collective action. Section 6 tests for additional complementarities in space. Section 7 estimates the costs of participation in the student movement and estimates its effect on the 2012 local elections. Section 8 concludes.

2 Theoretical Framework

This section presents a simple framework for the individual decision to participate in a social movement. The objective is to lay out testable features of participation as a function of the participation of others. In this framework, each individual interacts with a specific network and perfectly observe their participation in the movement.\(^5\)

Network participation could affect individual participation for multiple reasons, including pressure to conform, strategic complementarities, and information updating. Although some results in sections 5 and 6 arguably help to distinguish between these mechanisms, I remain agnostic about which one is relatively more important.

2.1 Environment

There are \( I \) individuals in a society. Let \( a_i \) be an indicator variable that takes the value of one if individual \( i \) participates in a movement that is revolting against the status quo and zero otherwise. Each individual \( i \) interacts with a group of individuals, \( i \)'s network. Individual \( i \) perfectly observes participation in her network, composed by \( n_i \) individuals.

\(^5\)For a thorough theoretical analysis, including equilibrium conditions, see Bramoullé et al. (2014) and Blume et al. (2015).
and denoted by the vector $\tilde{a}_{j(i)} = (a_1, \ldots, a_n)$. She also observes aggregate participation, denoted by the scalar $a_{-i} \equiv \sum_{k \neq i} a_k$. To simplify notation, let $n_i \equiv n$ and $\tilde{a}_{j(i)} \equiv \tilde{a}_j$.

The utility of individual $i$ from participating in the movement depends on her own action $a_i$, aggregate participation $a_{-i}$, and the participation of others in her network $\tilde{a}_j$:

$$u_i(a_i | \tilde{a}_j, a_{-i}) = p(a_{-i})B + (f(\tilde{a}_j) + b_i - c_i) a_i$$

(1)

where $p(a_{-i})$ is the probability of achieving change, $B$ is the benefit of changing the status quo, $f(\tilde{a}_j)$ is an unknown function of the participation of others in the network, $b_i$ is an individual benefit derived from participation, and $c_i$ is an individual cost of participation. Note that if $f(\tilde{a}_j) = 0$, then only individuals with $b_i > c_i$ participate.

Individual $i$ decides to participate in the movement if the utility from participating $u_i(a_i = 1 | \tilde{a}_j, a_{-i})$ is higher than the utility from not participating $u_i(a_i = 0 | \tilde{a}_j, a_{-i})$:

$$f(\tilde{a}_j) + b_i - c_i > 0$$

(2)

where it is easy to see that if $f(\tilde{a}_j) > 0$ (or $f(\tilde{a}_j) < 0$) some individuals with $b_i < c_i$ ($b_i > c_i$) will be pushed towards participation (non-participation) because of “network effects.” Section 7 calculates empirically this share of individuals.

### 2.2 Testable features of participation

This section describes four testable features of participation. First, is individual participation increasing or decreasing in the participation of others in the network? This question can be answered empirically using a simple “linear-in-means” model. This model was used by Manski (1993) to discuss econometric challenges and assumes that:

$$f(\tilde{a}_j) = \beta \cdot \frac{1}{n} (a_1 + \cdots + a_n) = \beta \cdot \bar{a}_j$$

(3)

This functional form is a modeling choice and may or may not be the best representation of the data. Nevertheless, if $\beta > 0$ then individual participation is increasing in the participation of others in the network and the reverse is true if $\beta < 0$.

---

6 Other predetermined observable variables of networks may also affect utility. These variables are omitted for simplicity but incorporated in the empirical analysis.
A slightly more general representation of the previous model allows for differential influence within i’s network. These “weak and strong ties” are emphasized by Granovetter (1973) and can be represented as:

\[ f(\bar{a}_j) = \beta \cdot (\omega_1 a_1 + \cdots + \omega_n a_n) \]  

(4)

This function allows individuals to take a weighted average of the participation of others. The weights \(\omega_k\), with \(k = 1, \ldots, n\) and \(\sum_k \omega_k = 1\), represent the differential influence that others have in i’s participation decision.

A model with differential influence can take many forms, depending on the modeling choice for the weights \(\omega_k\). Consider two examples. First, we could try to non-parametrically estimate these weights. After estimation, we could characterize influential individuals. Second, we could parameterize these weights using observable variables. The former approach is being explored by an ongoing research agenda (e.g. Manresa 2016). The latter approach has been, to the best of my knowledge, relatively unexplored empirically. Section 5 explores a “homophily model of influence” in which the difference in observable variables between the individual and others in her network determines the strength of the influence. In the absence of this type of influence, differences in observables should not determine the strength of influence.

The third testable feature of the model is a potential non-linearity in \(f\). In seminal studies of collective action, Tilly (1978) and Granovetter (1978) argue that the individual decision to participate in a movement may be influenced by the participation of others only when there is a “critical mass” participating. This is:

\[ f(\bar{a}_j) = g(\bar{a}_j) \cdot \mathbb{1}(\bar{a}_j > \alpha_i) \]  

(5)

where \(\bar{a}_j\) is the share of i’s network that participates, and \(\mathbb{1}(\bar{a}_j > \alpha_i)\) is an indicator function that takes the value of one if the share participating is larger than \(\alpha_i \in [0, 1]\). The threshold \(\alpha_i\) may be individual specific and could lead to multiple equilibria in participation.

The final testable feature of participation is that individuals may interact in K differ-

---

7In a study of participation in the 1964 Freedom Summer project, McAdam (1986, p. 88) shows that a strong tie to a participant is a better predictor of individual participation than a weak tie.
ent networks or, equivalently, there are $K$ different types of links:

$$f(\tilde{a}_j) = f_1(\tilde{a}_{j_1}) + \cdots + f_K(\tilde{a}_{j_K})$$

(6)

where $f_k(\tilde{a}_{j_k})$ represents the influence of others in network $j_k$. Two examples are social and spatial networks: individuals may be affected by the participation of friends that are not their neighbors or vice versa. If only social networks (say $k = 1$) affect individual participation, then $f(\tilde{a}_j) = f_1(\tilde{a}_{j_1})$ and $f_k(\tilde{a}_{j_k}) = 0 \forall k \neq 1$.

3 The Chilean Student Movement

From the Tunisian demonstrations sparking the Arab Spring to Occupy Wall Street triggering a movement against inequality, 2011 will be remembered as the year of the protester. The global wave of citizens demanding a “new democracy” also took place in Chile, where 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 the largest demonstrations in the country’s history, which were recognized worldwide as one of the most important social movements of the year.

The student movement began its protest activities in May of 2011, two months within the academic year and 14 months after a right-wing government took office for the first time in 50 years.\(^8\) The 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.\(^9\)

The first national protest was organized by the Confederation of Chilean Students, a national student organization, with the objective of exerting pressure on the annual presidential speech of May 21, day in which the government outlines next year policies. Students wrote a document outlining policies to decrease segregation in the educational system and increase government spending on education. After the presidential speech, the confederation of students sent a letter to the ministry of education expressing their

---

\(^8\)Chronicles written by leaders of the student movement are Figueroa (2012), Vallejo (2012), and Jackson (2013). A brief history of movement among high-school students can be found in Simonsen (2012).

\(^9\)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.
discontent with the presidential announcements (Confecch, 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 of June 1, and a failure to reach an agreement with the ministry of education in meetings held in May 30 and June 8, students intensified their protest activities. The movement was gradually supported by deans – including those in prestigious public universities – teachers, prominent worker 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 was low and continued to plummet after the rise of the movement (Figure A.3). Students called for another national protest day on June 16, at the time the largest mobilization in the country’s history. The government responded in June 25 with an offer and students rejected it and called for yet another national protest day on June 30.

Education was the main topic of conversation in the months of July and August. The leaders of the movement were regularly invited to television and radio shows, and diverse protest activities filled the country. Students ran and danced in front of the government palace, kissed in public spaces to gain citizens’ support, and exerted pressure on the government using different types of non-violent resistance. Rainy days and school holidays did not stop students’ activities, with large protests taking place under thousands of umbrellas. The president replaced the ministry of education in 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 school break of July, with the largest national protests taking place in August 24 and 25. These two days marked the peak of the student movement and protest activities declined in the following months.

Various reasons explain the decay of the student movement, including the beginning of formal negotiations, the focus of popular media on violent protesters, and students’ concerns about grade retention. After months of protests, what were the consequences?

[10]Jackson (2013, p. 22) states: “the constant emphasis on violence affected the strength of the movement”. The government threatened students with grade retention 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,
Contemporary surveys show that 80 percent of citizens supported the movement (Adi-
mark, 2011) and education became a national priority (Figure A.4). Candidates in the
subsequent local elections in 2012 and congress and presidential elections in 2013 were
constantly questioned about their ideological position regarding education. Some of the
older leaders of the movement founded political parties and four of them won seats at
the congress.

4 Data and Descriptive Statistics

The first part of this section describes the data used in the empirical analysis. All seven
administrative datasets were provided by the Ministry of Education in Chile. Six datasets
contain information about students and one dataset describes schools. The second part
of this section describes the participation of students in the student movement.

4.1 Administrative datasets and cities

The first dataset of students presents information about daily attendance to school in
2011. The second dataset contains students’ enrollment information (school, grade, class-
room). There are approximately 3,000,000 students every year, and 975,000 high-school
students enrolled in 2,700 high-schools in 2011. The third dataset contains information
on students’ annual academic performance (GPA, school attendance, repetition). These
three datasets are available for all students enrolled in the educational system.

Three additional datasets contain more information about high-school students in
2011. The fourth dataset corresponds to students’ performance in standardized tests,
taken some years by students in specific grades. Approximately 40 percent of high-
school students in 2011 took the test before that year. The fifth dataset corresponds to
household surveys, conducted in parallel to standardized tests. These surveys allow me
to measure household income and students’ internet connection at home, data available
for 57 percent and 36 percent of high-school students in 2011. The sixth dataset contains
self-reported home addresses and it is available for 35 percent of high-school students in
2011.

and summer holidays caused the movement to retreat until the next academic year.
The seventh and last dataset contains information about schools. Approximately 40 percent of students were enrolled in public schools in 2011, 60 percent in private schools, and 96 percent attended urban schools. School addresses are available and I use these to construct geographic clusters that I refer to as “cities.” These cities are isolated components in the network of schools, where schools are linked if these are closer than 5 kilometers from each other (see Figure A.5 for a map of cities). Table 1 presents descriptive statistics for students, schools, and cities.

4.2 Participation in the student movement

To measure students’ participation in the movement I use school absenteeism among high-school students during protest days in 2011. Several patterns in the data suggest this is indeed a useful way to measure participation. First, there are significant spikes in absenteeism during protest days. The upper panel in Figure 1 plots absenteeism from the beginning to the end of the 2011 school year. The first two national protest days (May 12 and June 1) are easy to observe in this figure. The sharp increase in school absenteeism between June 1 and June 16 maps 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 correct dates. As examples, the lower panels in Figure 1 present three time series of school-level absenteeism.

Given that the government collects information about students to measure performance and allocate public programs, I am able to describe participation in the movement in an unusually rich way. Let student $i$’s participation be defined as $\max\{0, A_{iT} - A_{iT}\}$, where $A_{iT}$ and $A_{iT}$ represent student $i$’s absenteeism before and after June 1. This measure of participation accounts for heterogeneity in absenteeism at the student level.

The average absenteeism of high-school students before and after June 1 was 13 percent and 40 percent respectively, which means that absenteeism increased by more than 200 percent. The average participation of a student is 0.30 (s.d. 0.32) and 4 percent of students did not go back to school in 2011. The participation of an average school is 0.20 (s.d. 0.22). Figures 2-A and 2-B plot the distribution of this participation measure for students and schools respectively.

$^{11}$The purpose of the maximum function is to truncate participation to be positive, although it is innocuous in the sense that few students decrease their absenteeism after June 1 in 2011.
Participation in the student movement was similar across students from different income groups and different performance measures. Figure 2-C plots the correlation between students’ participation and annual household income. Students from families earning US$10,000 annually skipped 30 days more than usual, while students from families earning more than US$30,000 missed 23 days more than usual. Figure 2-D shows that standardized test scores have little predictive power for participation. Both of these correlations are robust to the inclusion of student-level controls and city fixed effects.

Lower quality schools with more students connected to the internet participate more in the movement. Figure 2-E plots the correlation between participation and schools’ average test scores, a quality measure. A one standard deviation increase in quality signals is associated with a decrease of 4 percentage points in participation. Figure 2-F shows that, after controlling for household income, a one standard deviation increase in internet connection is associated with an increase of 8 percentage points in participation.

5 Collective Action in Social Networks

This section tests for potential complementarities in school absenteeism decisions between students and their networks during protest days. The focus is on June 16, at the time the largest mobilization in the country’s history and the first massive national protest day in a two and a half months period of intense protest activities (details in section 3). After describing the main regression equation of interest, I define social networks, discuss the main identification concerns, describe the identification strategy, and present results. In short, the identification strategy uses the exposure of students’ networks to the initial protest in other schools, an exposure that arises due to predetermined switching of students across schools.

5.1 Estimating equation and social networks

Consider the following regression relating a student’s decision to skip school in the first largest protest day as a function of school absenteeism in her social network:

\[ A_{isc} = f(A_{j(i)}) + g(b_{i}, c_{i}) + g(b_{j(i)}, c_{j(i)}) + \delta x_s + \zeta_c + \epsilon_{isc} \]  

(7)
where \( A_{isc} \) is an indicator that takes the value of one if student \( i \) in school \( s \) located in city \( c \) decides to skip school in June 16. In addition, \( f(A_{j(i)}) \) is a function of a vector containing the absenteeism decisions of \( i \)'s social network \( j(i) \), and \( g(b_i, c_i) \) and \( g(b_{j(i)}, c_{j(i)}) \) are functions of observable variables that account for the benefits and costs that may affect students’ decision to participate. Finally, \( x_s \) is a vector of control variables at the school level, \( \zeta_c \) is a city fixed effect, and \( \epsilon_{isc} \) is an error term clustered at the city level.

The analysis begins using a linear-in-means function \( f \). The vector of control variables \( b_i \) and \( c_i \) include average school attendance in 2010, GPA in 2010, an indicator for grade retention in 2010, an indicator for female, an indicator for students who switched school in 2010, and age. Averages of the same variables are included in \( b_{j(i)} \) and \( c_{j(i)} \), although results are robust to use more flexible functions. In addition, student controls also include school absenteeism during the May 12 and June 1 protest days. School-level controls include an indicator for public schools, reported quality signals (i.e. test score averages), the percentage of students who have repeated a grade in the past, and average household income.

Because students interact mostly with other students in their classroom, I define student \( i \)'s social network \( j(i) \) as the lifetime history of classmates. This definition of social networks gives rise to a large network of students linked within and across schools. Links across schools arise from switching of students across schools before 2011. Overall, this network contains more than 600 billion potential interactions between students across the country, and more than 60 million existing links in 2011. The average student has 80 other students in her social network, 60 percent attending the same school and 40 percent attending a different school in 2011.\(^{12}\)

### 5.2 Identification strategy

There are two concerns with a naive estimation of equation (7). First, the reflection problem emphasized by Manski (1993): students affect their social networks and social networks affect students. Second, there may be unobservable variables causing students and their networks to make similar decisions. Both concerns imply that an OLS estimat-

\(^{12}\)The term “social network” is coined for expositional purposes as other individuals may be part of a student’s network. The possibility of measurement error in social networks is addressed in the following sections. For computational reasons results use social networks defined as classmates in the previous four years. Results are robust to using more or less previous years.
tion will overestimate the effect of social networks on student’s decisions. To solve these issues, I exploit three sources of variation in an instrumental variables approach.\footnote{An additional source of bias 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 deal with this bias I follow Caeyers and Fafchamps (2016) and include the student’s value of the instrument (absenteeism in May 12) as an additional control variable.}

The first source of identifying variation is the exposure of social networks to protests in their social networks. The second source is a restriction to students attending a different school in 2011. The third source is the first national protest day in May 12, organized outside of the network of high-school students (see section 3). All in all, this strategy is similar to the one proposed by Bramoullé et al. (2009) and De Giorgi et al. (2010) with two important differences: the use of variation across days and, to minimize the threat of unobservables, a restriction on the set of second degree connections.

To gain intuition let student $i$’s network be denoted by the set $n_i$. The exposure of students in $n_i$ is measured by how much their networks $N_i$ participated in May 12, with $i \not\in N_i$. Students in the set $N_i$ may however still have similar unobservable variables than $i$. To deal with this concern, I restrict attention to a subset of students. Given the predetermined switching across schools, many students in $N_i$ are attending a different school in 2011. Let $M_i$ be the set of students that attend a different school than $i$ in 2011, with $M_i \subset N_i$ and $n_i \cap M_i = \emptyset$. The identification assumption is thus that school absenteeism in May 12 among students in the set $M_i$ only affects student $i$ absenteeism in June 16 through the absenteeism of $n_i$.

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

5.3 Linear estimates

Table 2 present estimation results of a linear-in-means model. Panel A presents OLS estimates of different specifications of equation (7). Although the focus is on the effect
network, school absenteeism during the initial protests of May 12 and June 1 are also interesting because these could potentially be measures of habit formation in protest activity.\textsuperscript{14} As the mean of the dependent variable and the main variable of interest are similar (0.49 and 0.50), point estimates can be interpreted directly as an elasticity.

Column 1 in Table 2-A presents estimates without control variables, a regression that explains almost two-thirds of the variation in June 16 absenteeism.\textsuperscript{15} The estimated coefficient implies that a one standard deviation increase in network absenteeism (0.31) is associated with an increase of 38 percentage points ($0.31 \times 1.23 = 0.38$) in the probability of skipping school. In terms of elasticities, a 10 percent increase in network absenteeism is associated with a 12 percent increase in student absenteeism. In addition, skipping school in the first two protest days increases the probability of skipping school in June 16 by 14 percentage points ($0.06 + 0.08 = 0.14$). Columns 2-5 progressively control for student, network, and school characteristics, and city fixed effects. As a result, the coefficient of networks remains stable. Although regressions control for a large set of observable variables at multiple levels, reflection and potential unobservable variables could cause a comovement of decisions between students and their networks.

A leading concern with estimates in Table 2-A are neighborhood unobservable variables causing a spurious positive correlation between students and their networks. To explore this possibility, I geo-coded 50,000 home addresses of students in Santiago, capital of Chile, and construct neighborhood fixed effects using latitude and longitude coordinates, creating areas of approximately $10 \times 10$ blocks (see Figure A.7 for a map). Column 7 includes these 714 indicators and estimated coefficients are unchanged, providing some evidence that neighborhood level variables are unlikely to be a concern. However, there may be additional complementarities in neighborhoods, i.e. absenteeism may be influenced by neighbors that were not classmates. I explore this possibility in section 6.

Table 2-B presents instrumental variables results. Table A.1 presents first-stages and reduced forms. As expected, estimated coefficients are positive and smaller in magnitude than their OLS counterparts. Columns 1-6 show the coefficient is robust with an elasticity of 0.6–0.8. Importantly, $F$-statistics in first-stages are always strong. Column

\begin{footnotesize}
\begin{enumerate}
\item The number of observations is presented in the bottom of Table 2. Differences in observations are due to missing values, which are more common in small schools located in rural areas.
\item In contrast, a regression on student, network, and school characteristics explains less than one-third of the June 16 variation in absenteeism, suggesting network effects and habit formation are important.
\end{enumerate}
\end{footnotesize}
in Table 2-B is the preferred specification. This result is robust to excluding schools closed by students in June 16: the 2SLS coefficient for networks is 0.53 (s.e. 0.14) with a first-stage $F$-stat of 30.2 (see Table A.2). When including school fixed effects the coefficient for networks decreases to 0.07 (s.e. 0.03, first-stage $F$-stat of 77.7).

The remainder of this section explores two deviations from the linear-in-means model. First, potential non-linear effects of the participation of others. Second, differential influence within students’ networks.

5.4 Critical mass

To test for non-linear networks effects, I use the nonparametric approach proposed by Newey et al. (1999). In this control function estimation, the coefficients of interest are associated to indicators for different values of absenteeism in social networks. The benchmark estimation uses eleven indicators: the first takes the value of one if absenteeism in social networks is between 0 and 10 percent, the second for 10-20 percent absenteeism, and so on until 100 percent absenteeism. Using equation (7):

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

where $(\beta_1, \ldots, \beta_{10})$ are the parameters of interest and the omitted category is absenteeism lower than 10 percent in social networks.

The upper-left panel of Figure 3 presents these ten estimated coefficients $(\hat{\beta}_1, \ldots, \hat{\beta}_{10})$ with their corresponding 95 percent confidence interval using the Newey et al. (1999) approach. The figure also plots the analog OLS estimates for comparison. The exact specification corresponds to column 5 in Table 2, which includes student, network, and school controls, and also city fixed effects. Similar to the linear estimates, the control function approach delivers estimates that are lower in magnitude than the OLS counterparts, as expected.

Consistent with the threshold model of collective behavior proposed by Granovetter (1978), the estimated coefficients show “critical mass” patterns. Student’s absenteeism is not affected by low values of absenteeism in social networks. In contrast, large values of social network absenteeism have strong effects on students’ decisions to skip school. The upper-right panel of Figure 3 plots the sequential difference between estimated coefficients to understand the marginal contribution of additional absenteeism in the
social network, suggesting that approximately 30-40 percent is the critical mass needed for networks to have an influence on students.

To estimate the average threshold in which networks begin to influence decisions, I repeat the previous estimation strategy but using 51 indicators for social network absenteeism, from 0 to 100 percent absenteeism in steps of 2 percentage points. The lower-left panel of Figure 3 presents the coefficients for these indicators. A vertical line marks the point in which the coefficient becomes statistically different from zero: 40 percent of absenteeism in the social network. The lower-right panel plots the distribution of absenteeism in social networks, a bimodal distribution consistent with a critical mass pattern of influence.

These critical mass results hold when omitting schools that were closed by students, and are qualitatively similar when including school or school-grade fixed effects (see panels A, B, and C in Figure 4). In addition, the critical mass of 40 percent should be interpreted as the average threshold. Panels D, E, and F in Figure 4 shows that students in larger schools, smaller networks, and smaller cities have lower thresholds.16

In sum, the estimated coefficients in Figures 3 and 4 provide evidence of “critical mass” type of complementarities in protest behavior: students’ absenteeism in protest days increases with network absenteeism only when more than 40 percent of other students in their networks skipped school.

5.5 Differential influence within networks

The empirical regularity of individuals forming links with others with similar characteristics is known as homophily (Jackson, 2010, chapter 6). Empirical work testing for differential influence following homophily patterns within networks is, however, more limited. Conditional on a network structure, i.e. links are already formed, does the strength of influence follows homophily patterns? This section tests this hypothesis focusing on three variables: gender, internet connection, and household income.

Table 3 presents results for the linear-in-means model. Columns 1 and 2 test for gender homophily patterns of influence by estimating a linear-in-means version of equation (7), restricting attention to males or females, and splitting the network in males and

---

16Figure A.8 shows additional patterns of heterogeneity. In addition, Figure A.9 shows that results are identical in schools with low and high levels of irregular spending of government transfers (CIPER, 2012).
females. For estimation I use the control function approach proposed by Newey et al. (1999). Under the null hypothesis of equal influence we should observe similar coefficients for the male and the female parts of the network. Results, however, indicate strong homophily patterns: same gender influence is more than ten times stronger than cross gender influence.

Columns 3 and 4 restrict attention to students with and without internet connection respectively, split the network in two, students with and without internet connection, and follow the same estimation strategy than before.\textsuperscript{17} Estimated coefficients indicate that the influence of students with internet on other students with internet is three times larger. The influence of students without internet on students without internet is two times larger.

Similar patterns of influence arise when restricting attention to the position of students in the income distribution. Columns 5–7 show that students from poor households are more influenced by students from poor households, and students from rich households are more influenced by students from rich households (and not by students from poor households).\textsuperscript{18}

Figure 5 present the nonparametric analogue of previous results. Patterns of critical mass are still clear when allowing for differential influence, and the null hypothesis of equal influence is easily rejected. Overall, results in Table 3 and Figure 5 provide evidence that supports the hypothesis of homophily patterns of influence within networks.

6 Multinetworks

This section tests for additional complementarities in other networks. Given the saliency of geographic location, I incorporate spatial networks into the analysis. The motivation is that individuals may be influenced by neighbors that are not in the previously defined social network. The findings in this section confirm previous results in two ways. First, these results rely on a different identification assumption. Second, these results are less

\textsuperscript{17}This is a partial test for the hypothesis of stronger coordination with internet because students (1) may have internet access at the school, and (2) may coordinate with other networks over the internet. Manacorda and Tesei (2016) and Enikolopov et al. (2016) provide city-level evidence of stronger network coordination with more access to cell phones and social media.

\textsuperscript{18}Rich households are defined as those with reported annual income higher than US$16,000, poor households with reported annual income lower than US$5,000, and the rest is defined as the middle class.
likely to be subject to measurement error.\textsuperscript{19}

The empirical challenge is to simultaneously test for complementarities in both social and spatial networks. Because home location is not available for all students and schools’ exact geographic location is, I now use schools as the unit of observation. First, I discuss the main estimating equation. Second, I propose an instrumental variables strategy that is based on “cross-network” exposure to the initial protest. Third, I present results.

6.1 Estimating equation

Consider an extended version of equation (7) that includes potential complementarities in social and spatial networks:

\[ A_{sc} = f(A_{n(s)}) + f(A_{m(s)}) + \gamma x_s + \theta_c + \varepsilon_{sc} \]  

(9)

where \( A_{sc} \) represents students’ average absenteeism in school \( s \) in city \( c \) in June 16, \( A_{n(s)} \) and \( A_{m(s)} \) represent students’ absenteeism in spatial \( n(s) \) and social \( m(s) \) networks in June 16, and \( x_s \) is a vector of control variables chosen using the method proposed by Belloni et al. (2013).\textsuperscript{20} The vector \( \theta_c \) controls for city fixed effects and the error term \( \varepsilon_{sc} \) is allowed to be spatially correlated within cities. The estimation of \( f \) corresponds to the statistical tests of interest, with \( f' > 0 \) providing evidence of complementarities in multinetworks. Given the fewer number of observations I use a linear-in-means function \( f \) for most of the analysis, but I also discuss and present estimates of flexible estimation of this function.

6.2 Identification strategy

The main concern with estimation of equation (9) is the potential existence of unobservable variables that affect both the absenteeism of schools and their networks. This source of bias will cause a spurious positive correlation between schools’ and networks’ absenteeism.

\textsuperscript{19}Laumann et al. (1983) show how “missing” links create bias in network statistics, Kossinets (2006) discusses different sources of measurement error, and Chandrasekhar and Lewis (2011) discusses the implications for regression analysis. Importantly, working with administrative data for the universe of students mitigates a significant number of concerns.

\textsuperscript{20}Control variables include: school absenteeism in May 12, school absenteeism in June 1, school absenteeism in 2010, and an indicator for public schools.
teeism. Crucially for the identification strategy I propose, the unobservables causing comovement can differ across networks. For example, the use of police force to decrease absenteeism in a geographic area will cause a comovement in spatial networks but not necessarily in social networks.

Before presenting the key equations of the econometric strategy, let me define two sets of schools. Let $m(s)$ represent the set of schools in the social network of $s$, $n(s)$ the set of schools in the spatial network of $s$, and $\ell(s) \equiv m(s) \cup n(s)$. The first set of interest corresponds to schools in the social network of a spatial network, $m(n(s)) \equiv m(n)$ under the previous notation, with $m(n) \cap \ell(s) = \emptyset$. The second set of interest corresponds to schools in the spatial network of a social network, $n(m(s)) \equiv n(m)$ with $n(m) \cap \ell(s) = \emptyset$.

Schools in the disjoint sets $n(s)$ and $m(n)$ are linked in the social network, while schools in the disjoint sets $m(s)$ and $n(m)$ are linked in space. With these two sets of schools in mind, consider the following first-stage regressions:

$$A_{m(s)} = \tau_1 Z_{m(s)} + \tau_2 Z_{m(n)} + \phi_1 A_{n(s)} + \gamma_1 x_s + \theta_c + \eta_{sc}$$

$$A_{n(s)} = \tau_3 Z_{m(s)} + \tau_4 Z_{m(n)} + \phi_2 A_{m(s)} + \gamma_2 x_s + \theta_c + \eta_{sc}$$

where $A_{m(s)}$ and $A_{n(s)}$ represent school absenteeism in the social and spatial networks of $s$ in June 16, $Z_{mn}$ represents school absenteeism in the spatial network of $m_s$ in May 12, and $Z_{nm}$ students’ average absenteeism in the social network of $n_s$ in May 12. Both $Z_{mn}$ and $Z_{nm}$ are the instruments I propose to isolate exogenous variation in network protest activity.

The relevance condition behind the proposed instruments follows a simple logic: if coefficients $\tau_2$ and $\tau_3$ are positive and statistically different from zero, then exposure to the initial protest increases protest participation in June 16. To gain intuition, Figure 6 presents this identification strategy graphically. In the upper panel the interest is on the effect of $B$ on $A$ through a link in the social network. The initial participation of $B$’s spatial links (marked in red) affects $B$’s participation, which in turn could affect $A$’s participation. The same logic applies in the lower panel, where the focus is now on the effect of $A$’s spatial links: the initial participation of $B$’s social links (marked in red) affects $B$’s participation, which again could affect $A$’s participation. Then, the exclusion restriction is that $B$’s links affect $A$ only through $B$.

Equations (10) and (11) also provide some evidence for this approach. Coefficients $\tau_1$
and $\tau_4$ should not be statistically different from zero. This should be the case because, after controlling for $A_{n(s)}$, we should not observe that social networks have an additional effect because all their influence is through $A_{n(s)}$, implying that $\tau_1 \approx 0$. The same argument applies in equation (11): after controlling for $A_{m(s)}$, there should be no effect of spatial networks, which implies that $\tau_4 \approx 0$.

### 6.3 Results

The multinetwork is composed by 2,070 high-schools (i.e. nodes) and two types of links (i.e. edges): spatial and social. Links are defined in the following way. Hypothetical schools $A$ and $B$ are linked in the social network if students transferred between these schools in previous years. In addition, schools $A$ and $B$ are linked in space if these are geographically close enough.\(^{21}\) Note that two schools can be theoretically linked in two networks and networks are imperfectly overlapped. Figure 7 presents a visualization of social and spatial networks. The average school has 3.4 spatial links, 1.7 social links, and a total of 4.6 links. The number of schools without any type of link is 177, with 709 schools having zero spatial links and 453 having zero social links.

Table 4 presents estimation results of three versions of equations (10) and (11). The first specification includes only the instruments $Z_{n(m)}$ and $Z_{m(n)}$, the second specification adds controls, and the third adds city fixed effects. The column pairs 1-4, 2-5, and 3-6 correspond to the first-stages in the three specifications. The bottom of the table presents the Angrist-Pischke $F$-statistic to test the statistical strength of the relevant instrument, and the Cragg-Donald $F$-stat for the combined strength of both first stages. In addition, columns 7-9 present the corresponding reduced form regressions. Overall, estimated coefficients and statistical tests suggest that the instruments are valid. First-stage coefficients are positive as hypothesized and statistically different from zero (i.e. $\tau_2 \gg 0$, $\tau_3 \gg 0$), I can reject the presence of weak instruments, there is evidence to support the approach (i.e. $\tau_1 \approx 0$, $\tau_4 \approx 0$), and the reduced form coefficients are always statistically significant and positive.

Table 5 presents two-stage least squares estimates (columns 4-6) and OLS coefficients for comparison (columns 1-3) for the three specifications previously discussed. Overall, estimated coefficients suggest the existence of complementarities in both spatial and

\(^{21}\)I use transfers of students in 2010 to define social links and one kilometer as the threshold for spatial links – roughly 10 blocks – although results are robust to different definitions.
social networks, and I cannot statistically reject that coefficients in both networks are different \((p\text{-value of } 0.63)\). These linear complementarities are robust to the exclusion of schools without students’ attendance in June 16: spatial and social network coefficients are 0.27 (s.e. 0.08) and 0.16 (s.e. 0.06) respectively. Estimated coefficients in my preferred specification (column 6) imply that a one standard deviation increase in absenteeism in the spatial (social) network causes an increase in school absenteeism of 9 (7) percentage points, an increase of 23 (18) percent. The corresponding elasticities with respect to absenteeism in spatial and social network are 0.3 and 0.2.

Given the number of observations, estimated coefficients are less precise than those in section 5 but still statistically different from zero. Although there is less statistical power to estimate potential non-linearities in both networks, for completeness Figure A.10 presents estimated coefficients similar to the ones presented in Figure 3. Not surprisingly, estimated coefficients have wide confidence intervals so I cannot reject the existence of non-linearities in spatial networks.

## 7 Consequences of Protests

This section estimates the costs of participating in the student movement and its effects on electoral outcomes, i.e. estimates of private costs and common benefits. A cohort analysis in a differences-in-differences framework reveals that grade repetition increased by 60 percent among high-school students in 2011. Within schools in 2011 more participation is associated with lower academic performance and a 33 percent increase in the probability of repeating the grade. An analysis of the 2012 local elections suggests that the movement shifted votes towards non-traditional opposition parties, relatively more aligned with the movement’s demands. The section ends with a counterfactual calculation for the contribution of networks to protest activities emphasizing the importance of allowing for non-linearities.

### 7.1 The cost of participation

*Cohort analysis.* Analysis of administrative data for the period 2002–2015 shows that participation in the movement lead to an increase in grade retention, an outcome causally associated with higher dropout rates, lower educational attainment, and more criminal
activities (Manacorda, 2012; Díaz et al., 2016).

To estimate the change in grade retention among high-school students in 2011, consider the following differences-in-differences regression:

\[ y_{hst} = \beta_t \times (G_{hs} \times T_t) + \zeta_{hs} + \lambda_t + \epsilon_{hst} \]  

(12)

where \( y_{hst} \) is retention of students in grade \( h \) of school \( s \) in year \( t \), with \( h \) either students in 1st-4th grade (non-participants) or students in 9-12th grade (participants). The indicator \( G_{sh} \) is equal to one for grades 9-12th and zero otherwise, \( T_t \) is a vector of indicator variables for years \( t = 2002, \ldots, 2015 \) (with 2010 as the omitted category), \( \zeta_{hs} \) and \( \lambda_t \) are school-grade and year fixed effects, and \( \epsilon_{hst} \) 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 8-A and 8-B present estimated coefficients \( \hat{\beta}_t \) using OLS. Figure 8-A uses absenteeism as dependent variable and Figure 8-B uses grade retention. Absenteeism among high-school students increased by 4.5 percentage points in annual official statistics, a 60 percent increase from a base of 0.08 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 0.06 in 2010, an unprecedented increase in the period 2002–2015.

\textit{Individual analysis.} To estimate the individual costs of participation consider a version of equation (7) that includes school fixed effects and uses academic performance at the end of 2011 as dependent variable. The coefficients of interest are again flexible estimates of social network absenteeism in June 16. Figures 8-C and 8-D present estimates using GPA and grade retention as dependent variables. Estimated coefficients imply that full absenteeism in social networks in June 16 is associated with (1) a decrease of 0.16 standard deviations in academic performance, and (2) an increase in grade retention of 38 percent (from a base retention of 0.06 in 2010).

Now consider a similar regression but using individual participation as independent variable. Estimated coefficients suggest that individual school absenteeism in June 16 leads to (1) a decrease of 0.1 standard deviations in GPA (coefficient of -0.07, \( p \)-value of 0.00), and (2) an increase in grade retention of 33 percent (coefficient of 0.02, \( p \)-value

\(^{22}\)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 schooling in 2011, mechanically decreasing an otherwise larger increase in absenteeism.
of 0.00). Results using annual participation, defined in section 4.2, imply that a one standard deviation increase in participation decreases GPA by 0.15 standard deviations and increases grade retention by 31 percent.

### 7.2 Political effects of the student movement

The first election after the rise of the student movement of 2011 was held in October of 2012.\(^{23}\) At these local elections citizens elected mayors in all of 345 counties in the country. Traditional parties, organized in left and right wing coalitions, competed against each other and against candidates from “non-traditional” parties. Although with new leaders and lower participation rates, the student movement was still active and many anticipated it would have an effect on electoral outcomes. The movement showed its discontent with traditional politics and publicly supported non-traditional parties.\(^{24}\) Despite its contemporary relevance, there is no systematic evidence of the impact the student movement had at these elections.

To estimate the effect of the student movement in the 2012 local elections, consider the regression:

\[
V_{c,2012} = \alpha + \beta \cdot \text{Student Movement}_{c,2011} + \gamma V_{c,2008} + \varepsilon_c
\]  

(13)

where \(V_{c,2012}\) and \(V_{c,2008}\) are electoral outcomes in the 2012 and 2008 local elections in county \(c\), \(\text{Student Movement}_{c,2011}\) is the county-level average participation of high-school students in the movement (see section 4.2), and \(\varepsilon_c\) is an robust error term. The dependent variable is either the vote share for non-traditional candidates or the percentage of voters in the county population.\(^{25}\)

The main concern with an OLS estimation of \(\beta\) is the potential existence of omitted variables correlated with the student movement and electoral outcomes. Three exercises

---

\(^{23}\)There was an informal plebiscite organized by citizens in October of 2011. Figure A.4 shows that participation was higher and people agreed more with students’ demands in counties with higher participation in the movement.

\(^{24}\)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 based on official data reported by the Electoral Service of Chile. Population data come from censuses. Figure A.11 plots the student movement variable for all counties.
suggest this is unlikely to be a major concern. First, regressions control for electoral outcomes in previous elections, which captures cross-sectional variation in political preferences. Second, placebo checks using school absenteeism in previous years support results. Third, I use the method proposed by Altonji et al. (2005) to construct bounds for estimates and qualitative conclusions remain.

Table 6 presents regression estimates. Column 2 indicates that a one standard deviation increase in the intensity of the student movement is associated with an increase of 5.3 percentage points in the vote share for non-traditional candidates, an increase of 16 percent on a base of 33 percent in 2008. This increase in vote shares is mostly explained from a decrease in vote shares for right-wing candidates, the coalition of the incumbent president. Column 5 suggest that the same increase in the movement intensity is associated with a decrease of 1.5 percentage points in voters. More speculatively, columns 3 and 6 provide some suggestive evidence of non-linear effects that are consistent with the previous “critical mass” patterns (see Figure A.12). As placebo checks, I create fake local movements using the differential increase in county-level school absenteeism between 2009 and 2010, i.e. before the rise of student movement. Reassuringly, this “fake movement” does not have an effect on electoral outcomes in the 2012 elections.

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 (2016) 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 is in the range $[0.053, 0.107]$ and the effect on voters is in the range $[-0.015, -0.011]$.26

7.3 Aggregate network effects

What is the aggregate contribution of networks to the observed daily protest activity? To answer this question, I estimate the difference in students’ choice probabilities of skipping school with and without network effects. More precisely, I estimate the following

\[ \hat{\beta} = \beta_c - (\beta_{nc} - \beta_c) \frac{R_{\text{max}} - R_c}{R_{nc} - R_c}, \]

where $\beta_c$ and $\beta_{nc}$ are coefficients from a regression with and without controls with corresponding $R$-squared of $R_c$ and $R_{nc}$, and $R_{\text{max}}$ is an unknown parameter in the interval $[R_c, 1]$. I use the conservative assumption of $R_{\text{max}} = 1$. See Oster (2016) for details.

---

26 Bounds use $\hat{\beta} = \beta_c - (\beta_{nc} - \beta_c) \frac{R_{\text{max}} - R_c}{R_{nc} - R_c}$, where $\beta_c$ and $\beta_{nc}$ are coefficients from a regression with and without controls with corresponding $R$-squared of $R_c$ and $R_{nc}$, and $R_{\text{max}}$ is an unknown parameter in the interval $[R_c, 1]$. I use the conservative assumption of $R_{\text{max}} = 1$. See Oster (2016) for details.
counterfactual difference:

$$\Pr \left( A_i = 1 \mid A_{j(i)} = a_{j(i)} \right) - \Pr \left( A_i = 1 \mid A_{j(i)} = 0 \right)$$

(14)

where $A_{j(i)}$ again represents absenteeism in students’ social networks. To estimate the choice probabilities, I use a control function approach and a probit model. The exact specification includes student, network, school controls, and city fixed effects. In addition, I compute the difference in equation (14) in two different ways: assuming linear network effects and allowing for non-linearities. The difference is informative about the importance of non-linearities. 27

The model with linear network effects predicts that 55 percent of students will skip school in a protest day. This percentage is equivalent to 440,000 students ($0.55 \times 800,000$) or 360,000 additional absent students ($440,000 - 80,000$) when compared to a regular school day. Now, in a counterfactual scenario without network effects, the model predicts than only 19 percent of students would skip school. These estimates imply that 78 percent of students that skipped school on a protest day would have attended school if there were no network effects ($1 - 80,000 / 360,000 = 0.78$). In contrast, the model with non-linearities predicts that 61 percent of students will skip school in a protest day. This number decreases to 45 percent when shutting down networks, implying that 31 percent of students skipping school on a protest day would have attended school in the absence of network effects ($1 - 280,000 / 408,000 = 0.31$).

Networks amplify protest activity significantly. However, the difference between linear and non-linear network effects is remarkable. Networks explain more than two-thirds of aggregate participation when effects are “constrained” to be linear (78 percent), and less than one-third when effects are non-linear (31 percent). A simple explanation for this pattern is found in Figure 4-B, where I plot the distribution of absenteeism in social networks. In the linear case, all students are influenced by their networks, regardless of network absenteeism. In the non-linear case, only students exposed to higher than 40 percent network absenteeism experience changes in their choice probabilities. The aggregate effect of networks captures this differential exposure.

27This calculation is similar to the one made in Yanagizawa-Drott (2014). Figure A.13 presents aggregate network effects for different numbers of initial protesters using estimated coefficients in section 5 and 7.2.
8 Conclusion

The individual decision to participate in a social movement is a crucial component behind the rise of groups demanding institutional change. Studying the Chilean student movement of 2011, this paper shows that students were influenced by their networks to participate in the movement only when a “critical mass” of 40 percent of their networks also participated. Overall, results support the popular idea of a tipping point in behavior (Gladwell, 2000) and the importance of strong ties to promote political activism (McAdam, 1986).

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

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

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

Future empirical studies of social movements may explore how protests create network links between participants and what are the consequences of this, how police violence in protests disrupt (or foster) participation, and how habit formation contributes to the escalation of a mobilization. For now, we have evidence that networks amplify the effect of protests in non-linear ways with potentially significant consequences for institutional change.

References


Figure 1: Absenteeism of high-school students in 2011

Notes: Own construction using administrative data. High-school students are students enrolled in grades 9-12. The y-axis is average school absenteeism among high-school students (in percentages) and the x-axis represents days in 2011. Vertical lines denote national protest days during week days. The gap in the center of the figures corresponds to winter break. More details about data in section 4.
**Figure 2:** Participation in the student movement

(a) Students ($N=836,988$)  
(b) Schools ($N=2,590$)  
(c) Students’ household income ($N=481,998$)  
(d) Students’ test scores ($N=326,820$)  
(e) Schools’ test scores ($N=2,428$)  
(f) Schools’ internet ($N=2,589$)

**Notes:** Participation is defined as additional school absenteeism after the beginning of the student movement. Red lines represent quadratic fits. School variables are averages of high-school students. Test scores are measured with standardized tests. Household income and internet at home are constructed from household surveys. More details about data in section 4.
Figure 3: Critical mass – main results

Notes: Panel A plots estimates (OLS and 2SLS) from a regression of individual school absenteeism on 10 indicators of network absenteeism in June 16, controlling for school, network, and school characteristics, and city fixed effects (see equations 7 and 8). Vertical lines denote 95 percent confidence intervals (s.e. clustered at the city level). Panel B plots the difference in the estimated coefficients in Panel A. Panel C plots the same coefficients that in Panel A, but using 51 (instead of 10) indicators of network absenteeism. Panel D plots the distribution of social network absenteeism. More details in section 5.4.
**Figure 4: Critical mass – additional results**

(a) Including school F.E.  
(b) Including school-grade F.E.  
(c) School closures  
(d) Network size  
(e) School size  
(f) City size

**Notes:** Panels A and B plot OLS and 2SLS estimates from a regression of individual school absenteeism on 10 indicators of network absenteeism, controlling for school, network, and school characteristics (see equations 7 and 8). Panels C-F present 2SLS estimates in sub-samples. More details in section 5.4.
Figure 5: Differential influence within networks

Notes: All panels plot 2SLS estimates from a regression of individual school absenteeism on 10 indicators of network absenteeism, controlling for school, network, and school characteristics, and city fixed effects. Regressions are in sub-samples and split the network in groups. More details in section 5.5.
Figure 6: Multinetwork instruments

Notes: This figure provides intuition for the multinetwork instruments to identify the causal effect of node B on node A in social networks (Panel A) and spatial networks (Panel B). Each circle represents a node (e.g., school) and each line represents a link in the social network (dark lines) or spatial network (gray lines). The dash circles define the area of spatial networks. For example, two nodes are linked if these are closer than 1 kilometer. The identifying variation is marked in red. Identification in this strategy relies on “cross network” exposure. The additional dimension of time is missing from this figure. More details in section 6.2.
**Figure 7:** Multinetworks in the data

Notes: This figure presents a visualization of high-schools in social networks (left-hand side), and spatial networks (right-hand side). Each node represents a high-school and lines represent social or spatial links. The legend with numbers in the top right of each network represents the number of isolated schools in those types of sub-networks, e.g. there are 709 schools without social links and 426 without spatial links.
Figure 8: The cost of participation

Notes: Panels A and B plot differences-in-differences estimates of absenteeism/retention rates between high-school students (participants in the movement) and students age 6–10 (non-participants) in the period 2002-2015. 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.

Notes: Panels C and D plot OLS estimates from a regression of academic performance on social 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

<table>
<thead>
<tr>
<th>Students</th>
<th>Mean</th>
<th>St. Dev.</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School absenteeism:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>May 12, 2011</td>
<td>0.15</td>
<td>0.36</td>
<td>760,801</td>
</tr>
<tr>
<td>June 1, 2011</td>
<td>0.19</td>
<td>0.39</td>
<td></td>
</tr>
<tr>
<td>June 16, 2011</td>
<td>0.49</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>Average in 2010</td>
<td>0.07</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>Repeated grade in 2010</td>
<td>0.06</td>
<td>0.23</td>
<td></td>
</tr>
<tr>
<td>GPA in 2010</td>
<td>5.40</td>
<td>0.59</td>
<td></td>
</tr>
<tr>
<td>Switched school after 2010</td>
<td>0.24</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0.51</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>15.8</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>Household income (annual US$)</td>
<td>7,891</td>
<td>7,892</td>
<td>481,998</td>
</tr>
<tr>
<td>Internet connection at home</td>
<td>0.55</td>
<td>0.50</td>
<td>304,448</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Schools</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Average test score in standardize test</td>
<td>250</td>
<td>40</td>
<td>2,224</td>
</tr>
<tr>
<td>Share of students who repeated grade</td>
<td>0.06</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>Average household income (annual US$)</td>
<td>8,877</td>
<td>5,227</td>
<td></td>
</tr>
<tr>
<td>Public</td>
<td>0.30</td>
<td>0.46</td>
<td></td>
</tr>
<tr>
<td>Students</td>
<td>342</td>
<td>325</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cities</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>High-schools in the city</td>
<td>7.7</td>
<td>44.3</td>
<td>290</td>
</tr>
<tr>
<td>High-school students in the city</td>
<td>2,623</td>
<td>16,134</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Own construction based on administrative data provided by the Ministry of Education. All variables are measured in 2011 unless otherwise stated. The number of observations is the same as in the first row of each panel unless otherwise stated. More details in section 4.
Table 2: Linear estimates

Dependent variable is absenteeism in June 16, 2011

<table>
<thead>
<tr>
<th>Panel A – OLS estimates</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network absenteeism in June 16</td>
<td>1.23***</td>
<td>1.22***</td>
<td>1.27***</td>
<td>1.21***</td>
<td>1.24***</td>
<td>1.47***</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.04)</td>
<td>(0.05)</td>
<td>(0.03)</td>
</tr>
<tr>
<td>Absenteeism in May 12</td>
<td>0.06***</td>
<td>0.05***</td>
<td>0.05***</td>
<td>0.03***</td>
<td>0.04***</td>
<td>0.04***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.00)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Absenteeism in June 1</td>
<td>0.08***</td>
<td>0.08***</td>
<td>0.07***</td>
<td>0.05***</td>
<td>0.06***</td>
<td>0.06***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B – 2SLS estimates</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Network absenteeism in June 16</td>
<td>0.80***</td>
<td>0.77***</td>
<td>0.69***</td>
<td>0.81***</td>
<td>0.69***</td>
<td>0.63***</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.05)</td>
<td>(0.08)</td>
<td>(0.07)</td>
<td>(0.10)</td>
<td>(0.21)</td>
</tr>
<tr>
<td>Absenteeism in May 12</td>
<td>0.12***</td>
<td>0.10***</td>
<td>0.09***</td>
<td>0.05***</td>
<td>0.05***</td>
<td>0.06***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
</tr>
<tr>
<td>Absenteeism in June 1</td>
<td>0.16***</td>
<td>0.15***</td>
<td>0.14***</td>
<td>0.08***</td>
<td>0.08***</td>
<td>0.12***</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Student controls</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Network controls</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>School controls</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>City F.E.</td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neighborhood F.E.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>F-stat 1st stage</td>
<td>53.3</td>
<td>50.5</td>
<td>30.6</td>
<td>36.0</td>
<td>24.1</td>
<td>14.0</td>
</tr>
<tr>
<td>R-squared (Panel A)</td>
<td>0.626</td>
<td>0.629</td>
<td>0.638</td>
<td>0.645</td>
<td>0.652</td>
<td>0.583</td>
</tr>
<tr>
<td>Observations</td>
<td>779,327</td>
<td>779,251</td>
<td>771,121</td>
<td>760,801</td>
<td>760,801</td>
<td>49,273</td>
</tr>
</tbody>
</table>

Notes: Student controls include academic performance, average school attendance in previous years and socioeconomic characteristics. Network controls include average student controls at the network level. School controls include indicators for publicly managed schools, average academic performance of students, and average socioeconomic characteristics of students. Neighborhoods are geographic areas where students live. More details in section 5. See Figure A.5 for a map of cities. In column 6, each neighborhood is of size 10×10 blocks. Neighborhood data is only available for some students. See Figure A.7 for a map of neighborhoods. Standard errors clustered at the city level are reported in parentheses. Significance level: *** p < 0.01.
Table 3: Linear estimates – differential influence within networks

*Dependent variable is absenteeism in June 16, 2011*

<table>
<thead>
<tr>
<th>Social network in June 16</th>
<th>(1) Male</th>
<th>(2) Female</th>
<th>(3) Yes</th>
<th>(4) No</th>
<th>(5) Poor</th>
<th>(6) Middle</th>
<th>(7) Rich</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>0.44***</td>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Females</td>
<td>0.03</td>
<td>0.42***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students with internet</td>
<td></td>
<td></td>
<td>0.27***</td>
<td>0.11***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students without internet</td>
<td></td>
<td></td>
<td>0.08***</td>
<td>0.22***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Students from poor households</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.15***</td>
<td>0.06***</td>
<td>-0.01</td>
</tr>
<tr>
<td>Students from middle income households</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.06***</td>
<td>0.14***</td>
<td>0.11***</td>
</tr>
<tr>
<td>Students from rich households</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.01***</td>
<td>0.01***</td>
<td>0.09***</td>
</tr>
<tr>
<td>Mean of dep. variable</td>
<td>0.48</td>
<td>0.49</td>
<td>0.42</td>
<td>0.49</td>
<td>0.51</td>
<td>0.45</td>
<td>0.37</td>
</tr>
<tr>
<td>Absenteeism previous protests</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student, network, and school controls</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>City F.E.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>375,737</td>
<td>385,057</td>
<td>229,066</td>
<td>187,034</td>
<td>234,615</td>
<td>257,555</td>
<td>78,760</td>
</tr>
</tbody>
</table>

Notes: All columns present 2SLS estimates. Regressions are estimated in sub-samples of students and the network is split in groups with similar observable variables. More details in section 5.5. Significance level: *** $p < 0.01$, ** $p < 0.05$. 
<table>
<thead>
<tr>
<th></th>
<th>Absenteeism in spatial network June 16</th>
<th>Absenteeism in social network June 16</th>
<th>School absenteeism June 16</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Absenteeism in spatial network of social network in May 12</td>
<td>0.15</td>
<td>0.16**</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.08)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>Absenteeism in social network of spatial network in May 12</td>
<td>0.82***</td>
<td>0.78***</td>
<td>0.60***</td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td>(0.09)</td>
<td>(0.05)</td>
</tr>
<tr>
<td>Controls</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>City F.E.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Angrist-Pischke F-test</td>
<td>137.0</td>
<td>70.4</td>
<td>148.0</td>
</tr>
<tr>
<td>Cragg-Donald (F-stat)</td>
<td>98.2</td>
<td>67.9</td>
<td>48.2</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.19</td>
<td>0.20</td>
<td>0.40</td>
</tr>
</tbody>
</table>

Notes: Total number of schools is 2,070 and regressions are weighted by the number of students. Standard errors clustered at the city level are reported in parentheses. More details in section 6. Significance level: *** $p < 0.01$, ** $p < 0.05$. 
### Table 5: Collective action in multinetworks

*Dependent variable is school absenteeism in June 16, 2011*

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th></th>
<th></th>
<th>2SLS</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td>Spatial network absenteeism in June 16</td>
<td>0.13**</td>
<td>0.18***</td>
<td>0.01</td>
<td>0.39***</td>
<td>0.31***</td>
<td>0.27***</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.03)</td>
<td>(0.06)</td>
<td>(0.09)</td>
<td>(0.06)</td>
<td>(0.07)</td>
</tr>
<tr>
<td>Social network absenteeism in June 16</td>
<td>0.33***</td>
<td>0.10***</td>
<td>0.04</td>
<td>0.75***</td>
<td>0.19*</td>
<td>0.20*</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.02)</td>
<td>(0.03)</td>
<td>(0.16)</td>
<td>(0.10)</td>
<td>(0.11)</td>
</tr>
<tr>
<td>School absenteeism in June 1</td>
<td>0.17**</td>
<td>0.16***</td>
<td></td>
<td>0.13*</td>
<td>0.13**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.07)</td>
<td>(0.06)</td>
<td></td>
<td>(0.07)</td>
<td>(0.06)</td>
<td></td>
</tr>
<tr>
<td>School absenteeism in May 12</td>
<td>0.08</td>
<td>0.09*</td>
<td></td>
<td>0.05</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.05)</td>
<td></td>
<td>(0.06)</td>
<td>(0.05)</td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>City F.E.</td>
<td>x</td>
<td></td>
<td></td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.106</td>
<td>0.446</td>
<td>0.553</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

**Notes:** Total number of schools is 2,070 and regressions are weighted by the number of students. Statistical tests and first-stages are presented in Table 4. More details in section 6. Standard errors clustered at the city level are reported in parentheses (320 clusters). Significance level: *** $p < 0.01$, ** $p < 0.05$. 
Table 6: The political effects of the student movement

*Dependent variables are electoral outcomes in the 2012 local elections*

<table>
<thead>
<tr>
<th></th>
<th>Vote share non-traditional parties</th>
<th>Voters in population</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td><strong>Main estimates</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student movement</td>
<td>0.025</td>
<td>0.053**</td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
<td>(0.022)</td>
</tr>
<tr>
<td>Student movement squared</td>
<td>0.016</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td></td>
</tr>
<tr>
<td><strong>Placebo</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Δ school absenteeism 2010-2009</td>
<td>-0.013</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.021)</td>
</tr>
<tr>
<td>Δ school absenteeism 2010-2009 squared</td>
<td>-0.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td></td>
</tr>
<tr>
<td>Dep. variable in 2008 election</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Student movement (p-value)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placebo (p-value)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean dep. variable</td>
<td>0.35</td>
<td>0.35</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.006</td>
<td>0.347</td>
</tr>
<tr>
<td>Counties</td>
<td>345</td>
<td>345</td>
</tr>
</tbody>
</table>

**Notes**: Regressions are weighted by the total number of voters in 2008. *Student movement* has been standardized to facilitate the interpretation of coefficients. Non-traditional parties correspond to parties that are different from the coalition “Concertación” and the right wing coalition, i.e. independent candidates. The coefficients for *Placebo* estimates come from separate regressions. The “Student movement (p-value)” and “Placebo (p-value)” in the bottom of the table correspond to p-values for the test that linear and quadratic terms are different from zero. Robust standard errors are reported in parentheses. Significance level: *** $p < 0.01$, ** $p < 0.05$. 

46
Online Appendix

Collective Action in Networks: Evidence from the Chilean Student Movement

Felipe González

List of Tables

A.1 Linear estimates – reduced form and first stage ........................................ 2
A.2 Linear estimates – robustness to school closures ........................................ 3

List of Figures

A.1 Protests in Chile 1979–2013 ........................................................................ 4
A.2 Economic indicators ..................................................................................... 5
A.3 Citizens’ evaluation of incumbent politicians ............................................. 5
A.4 Survey evidence for the impact of the student movement ......................... 6
A.5 Cities ........................................................................................................... 7
A.6 Placebos for first-stage ................................................................................ 8
A.7 Location of students in Santiago ................................................................. 9
A.8 Threshold model heterogeneity .................................................................. 10
A.9 Heterogeneity by irregular spending .......................................................... 11
A.10 Multinetworks non-linear estimates .......................................................... 12
A.11 Participation by county .............................................................................. 13
A.12 The political effects of the student movement .......................................... 14
A.13 Aggregate effects of networks ................................................................... 15
Table A.1: Linear estimates – reduced form and first stage

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

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

Notes: Student controls include academic performance, average school attendance in previous years and socioeconomic characteristics. Network controls include average student controls at the network level. School controls include indicators for publicly managed schools, average academic performance of students, and average socioeconomic characteristics of students. Neighborhoods are geographic areas where students live. More details in section 5. See Figure A.5 for a map of cities. In column 6, each neighborhood is of size 10×10 blocks. Neighborhood data is only available for some students. See Figure A.7 for a map of neighborhoods. Standard errors clustered at the city level are reported in parentheses. Significance level: *** p < 0.01.
Table A.2: Linear estimates – robustness to school closures

*Dependent variable is student absenteeism in June 16, 2011*

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network absenteeism in June 16</td>
<td>0.66***</td>
<td>0.57***</td>
<td>0.56***</td>
<td>0.63***</td>
<td>0.53***</td>
<td>0.68***</td>
<td>0.08**</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.10)</td>
<td>(0.13)</td>
<td>(0.11)</td>
<td>(0.14)</td>
<td>(0.18)</td>
<td>(0.04)</td>
</tr>
<tr>
<td>Absenteeism in May 12</td>
<td>0.14***</td>
<td>0.12***</td>
<td>0.10***</td>
<td>0.09***</td>
<td>0.09***</td>
<td>0.07***</td>
<td>0.09***</td>
</tr>
<tr>
<td></td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.00)</td>
</tr>
<tr>
<td>Absenteeism in June 1</td>
<td>0.18***</td>
<td>0.16***</td>
<td>0.14***</td>
<td>0.12***</td>
<td>0.12***</td>
<td>0.13***</td>
<td>0.12***</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.02)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.02)</td>
<td>(0.00)</td>
</tr>
</tbody>
</table>

3

Student controls x x x x x x x
Network controls x x x x x x
School controls x x x x
City F.E. x
Neighborhood F.E. x
School F.E. x
F-stat 1st stage 23.7 23.7 23.8 24.7 30.2 28.5 183.3
Observations 505,643 505,610 500,834 492,903 492,902 34,145 500,798

**Notes:** *Student controls* include academic performance, average school attendance in previous years and socioeconomic characteristics. *Network controls* include average student controls at the network level. *School controls* include indicators for publicly managed schools, average academic performance of students, and average socioeconomic characteristics of students. Estimation excludes schools with absenteeism larger than 0.99. *Neighborhoods* are geographic areas where students live. More details in section 5. See Figure A.5 for a map of cities. In column 6, each neighborhood is of size 10×10 blocks. Neighborhood data is only available for some students. See Figure A.7 for a map of neighborhoods. Standard errors clustered at the city level are reported in parentheses. Significance level: *** $p < 0.01.$
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**

![Economic indicators graph]

**Notes:** Data from the Central Bank of Chile. All variables have been normalized. The vertical red line denotes the beginning of the student movement.

**Figure A.3: Citizens’ evaluation of incumbent politicians**

![Citizens’ evaluation graph]

**Notes:** Normalized index for the approval of incumbent politicians. Data from the Centro de Estudios Públicos and Adimark.
Figure A.4: Survey evidence for the impact of the student movement

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).

Notes: Panels E and F plot the participation in the “National plebiscite for education” in October of 2011 at the county level (E) and the percentage of people that agrees with the students’ demands among those who participated (F).
Figure A.5: Cities

Notes: This map plots the ten largest cities in the most populated area of the country. Cities are defined as closed geographic polygons without any other close by school.
Figure A.6: Placebos for first-stage

Notes: This figure plots OLS estimates from a single cross-sectional regression. The dependent variable is June 16 school absenteeism in students’ social networks. The figure presents standardized coefficients for absenteeism in May 12 among out-of-school students in the social network of social networks. 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.
Figure A.7: Location of students in Santiago

Notes: Home addresses for approximately 50,000 students in Santiago in 2011. The road network is plotted in black for geographic reference.
Figure A.8: Threshold model heterogeneity

Notes: All panels plot 2SLS estimates from a regression of individual school absenteeism on 10 indicators of network absenteeism, controlling for school, network, and school characteristics (see equations 7 and 8) in sub-samples. More details in section 5.4.
Figure A.9: Heterogeneity by irregular spending

(a) Estimated parameters

(b) Marginal contribution of additional network absenteeism

Notes: Panel A plots 2SLS estimates from a regression of individual school absenteeism on 10 indicators of network absenteeism in June 16, controlling for school, network, and school characteristics, and city fixed effects (see equations 7 and 8) in sub-samples. Vertical lines denote 95 percent confidence intervals (s.e. clustered at the city level). Panel B plots the difference in the estimated coefficients in Panel A. Large and small refer to the sub-samples of counties with large or small percentage of government spending classified as “irregular.”
**Figure A.10:** Multinetworks non-linear estimates

![Graph showing the relationship between spatial and social network absenteeism and school absenteeism.](image)

(a) Coefficients for spatial network absenteeism

(b) Coefficients for social network absenteeism

**Notes:** This figure plots 2SLS estimates from a regression of school absenteeism on 10 indicators of social and spatial network absenteeism in June 16, controlling for school absenteeism before June 16, school characteristics, and city fixed effects. Vertical lines denote 95 percent confidence intervals with standard errors clustered at the city level.
Figure A.11: Participation by county

Notes: Own construction based on administrative data. Counties are ordered from north to south in the x-axis. There are 324 (out of 346) counties with positive participation in the student movement. Large counties are defined as the fifteen counties with the largest number of students.
Figure A.12: The political effects of the student movement

(a) Vote share for non-traditional candidates

(b) Voters in population

Notes: Quadratic fit of electoral outcomes in the 2012 elections on a measure of the intensity of the student movement in 2011. There are 345 counties in the country.
Notes: Simulation results of random initial protesters in the network of students. I proceed in four steps. In the first step, I choose the size of initial protesters, which varies in each simulation. The x-axis measures the percentage of the population of students that is initially protesting (“initial size of social movement”). In the second step, I calculate the choice probabilities of skipping school for all students in the country using the estimated parameters for the effect of networks (non-linear results). In the third step, I calculate how many additional students skip school because of network effects (y-axis in the left-hand side). In the fourth step, I take aggregate participation in the movement – measured as the percentage of students skipping school – and use the estimated coefficients to calculate additional votes for non-traditional parties (in percentage points in the right-hand side y-axis).