Friendship Networks and Political Opinions: A Natural Experiment among Future French Politicians

Yann Algan* Nicolò Dalvit† Quoc-Anh Do‡ Alexis Le Chapelain
Yves Zenou§

May 31, 2019

Abstract

We study how friendship shapes students' political opinions in a natural experiment. We use the indicator whether two students were exogenously assigned to a short-term "integration group", unrelated to scholar activities and dissolved before the school year, as instrumental variable for their friendship, to estimate the effect of friendship on pairwise political opinion outcomes in dyadic regressions. After six months, friendship causes a reduction of differences in opinions by one quarter of the mean difference. It likely works through a homophily-enforced mechanism, by which friendship causes politically-similar students to join political associations together, which reinforces their political similarity. The effect is strong among initially similar pairs, but absent in dissimilar pairs. Friendship affects opinion gaps by reducing divergence, therefore polarization and extremism, without forcing individuals' views to converge. Network characteristics also matter to the friendship effect.

Keywords: Political opinion, polarization, friendship effect, social networks, homophily, extremism, learning, natural experiment.

JEL classification codes: C93, D72, Z13.

---

*Sciences Po and CEPR. Email: yann.algan@sciencespo.fr.

†Sciences Po. Email: nicolo.dalvit@sciencespo.fr.

‡Sciences Po and CEPR. Email: quocanh.do@sciencespo.fr (corresponding author).

§Monash University, Australia, IFN, and CEPR. Email: yves.zenou@monash.edu.

[1] We thank valuable comments from Alberto Alesina, Ghazala Azmat, Roland Bénabou, Yann Bramoullé, Emily Breza, Jing Cai, Arun Chandrasekhar, Aureo De Paula, Ruben Durante, Ruben Enikolopov, Marcel Fafchamps, Gigi Foster, Jeff Frieden, Matt Gentzkow, Ben Golub, Sanjeev Goyal, Nicolaj Harmon, Emeric Henry, Yannis Ioannides, Emir Kamenica, Ethan Kaplan, Rachel Kranton, Eliana La Ferrara, Horacio Larreguy, Alexandre Mas, Friederike Mengel, Markus M. Möbius, Gautam Rao, Imran Rasul, Andrei Shleifer, Enrico Spolaore, Andreas Steinmayr, Uwe Sunde, Adam Szeidl, Fabian Waldinger, Noam Yuchtman, and from seminar and conference participants at Adansonia conference at U Bocconi, HSE Moscow, NES, UBC, CREST Political Economy workshop, UNSW, Monash U, NUS, Queen Mary U, U of Copenhagen, Royal Holloway, Brazilian Econometric Society meeting, LMU Munich, Harvard U, U of Chicago, U Autònoma de Barcelona, Monash U Symposium on Social and Economic Networks, Barcelona Summer Forum, the Social Impacts of Economic Policies and Change conference in Heidelberg, the AEA Meeting, the Information Transmission in Networks conference at Harvard, and Aix-Marseille School of Economics. We are thankful for Sri Srikandan's assistance with the survey website. Algan acknowledges support from the European Research Council Grant 647870. Do is grateful for the generous hospitality of Harvard University's Department of Economics and Weatherhead Center For International Affairs, and financial support from the French National Research Agency's (ANR) "Investissements d'Avenir" grants ANR-11-LABX-0091 (LIEPP) and ANR-11-IDEX-0005-02 (USPC), and the Banque de France - Sciences Po partnership. The remaining errors are ours.

1 Introduction

The recent rise of populism and political polarization are attracting a burgeoning research area on the role of social network in the formation of political beliefs. Many authors attribute political polarization to the rise of social media (e.g., Sunstein, 2009, 2018; Pariser, 2011), by which social networks' echo chambers and filter bubble reinforce prejudices among like-minded group members, while others debate the quantitative importance of such mechanism (e.g., Boxell et al., 2017; Allcott and Gentzkow, 2017, for the 2016 U.S. presidential election). A key missing input in this heated debate remains the causal impact of social networks on belief formation.

This paper seeks to provide a set of estimates of such impact that are immune to the concern of bias due to endogenous network formation, by exploiting a natural experiment at the elite French Institute of Political Studies, Sciences Po, that quasi-randomly allocates first-year students into groups at the beginning of their studies.

While recent research has flourished on the question how political opinion and participation, especially voting, are influenced by leaders and groups (e.g., DellaVigna and Gentzkow, 2010; Gabel and Scheve, 2007; Carlsson et al., 2015), and by the media (e.g., DellaVigna and Kaplan, 2007; Gentzkow, 2006; Gentzkow et al., 2011; Kendall et al., 2015; Gerber et al., 2009), in the spirit of the seminal, descriptive study by Lazarsfeld et al. (1944) on friends' influence on US voters, this paper focuses on friendship interactions between individuals in the same group. We investigate how a friendship link between two individuals may raise or lower the chance that their political opinions converge or diverge. We explore how friendship affects individual choices of shared activities, and how it may reinforce or reduce friendship's effect on opinions.

We consider the network of first-year students at Sciences Po, for its central role in the formation of most top French politicians since World War II, and for its students' enthusiasm in politics.[1] Compared with other French higher education institutions, Sciences Po students are much more interested and proactive in political movements (one out of ten first-year students is already registered with a political party), and have more exposure to politically-oriented events and activities, organized by either student associations or the Institute. We survey all first-year students in March 2014 with incentive-compatible questions to elicit their social networks (the method proposed by Leider et al., 2009, 2010), as well as questions on political opinions and views, and specify dyadic regressions of pairwise differences in opinions on friendship links between pairs.

The major concern in such regressions is the homophily bias, i.e., omitted variable bias due to endogenous network formation. In presence of homophily (the proclivity to befriend similar individuals) based on unobserved characteristics,[2] the OLS estimate will likely bias the effect of

[1] Sciences Po's alumni include notably six of the seven French presidents after Charles de Gaulle, namely Emmanuel Macron, François Hollande, Nicolas Sarkozy, Jacques Chirac, François Mitterrand, and Georges Pompidou; and the majority of Prime Ministers.

[2] The concept of homophily was first highlighted by seminal studies in sociology since Lazarsfeld and Merton

friendship on opinion differences away from zero.

We address this concern with an instrumental variable for friendship that arises from the 'integration week" before the first year starts. During this week, students are assigned by alphabetical order to separate groups of around 16, to conduct social activities to facilitate students' socialization and integration into the new environment. Consequently, common membership in the same group increases the chance of friendship, estimated at 16 percentage points, while it is arguably excludable from the formation of political opinions at the moment of our survey six months later. The same-integration-group dyadic variable can thus serve as instrument for pairwise friendship in the specification of interest, which estimates the Local Average Treatment Effect (LATE) of friendship among complier pairs (those who become friends only due to being in the same group).

Our methodology's use of an exogenous source of variation in network formation is distinctively novel in the recent empirical and econometric literature on social networks. Traditionally, the endogeneity of network formation received rather limited attention and treatment in studies that rely mostly on restrictions on the structure of interactions and uses of control variables, including fixed effects, with an identification underlined by Bramoullé et al.'s (2009) results, such as Bifulco et al. (2011); Calvó-Armengol et al. (2009); Patacchini and Zenou (2016); DeGiorgi et al. (2010). A different strand of the literature takes a structural approach that explicitly models the formation of network based on assumptions on individuals' interactions and expectations, and derives identification conditions from the model, including recent developments such as Mele (2017); Badev (2018); Goldsmith-Pinkham and Imbens (2013), as reviewed by De Paula (2017) and Graham (2015).[3] Different from those approaches, ours relies on a source of variation that draws its exogeneity and validity from design, not modeling assumptions, and then uses a relatively simple and transparent econometric technique, namely an IV strategy to identify the LATE.

While our use of an exogenous group assignment echoes the vast literature on peer effects under randomized assignment (as surveyed by Sacerdote, 2011, 2014; Epple and Romano, 2011), our focus on friendship links, instead of peer-group relationships, is fundamentally different. We consider that friendship is chosen by individuals, not assigned by design, thus it naturally is influenced by, and interacts with, individual characteristics and behaviors, as discovered in the case of Carrell et al. (2013).[4] It is thus important to understand the effect of friendship beyond that of peer group

(1954), as surveyed by McPherson et al. (2001). Soon highlighted as a barrier to empirical identification by Manski (1993), it has been further studied in economics by, e.g., Currarini et al. (2009) and Golub and Jackson (2012).

[3]For a review of the literature on empirical methods in social networks, also see Advani and Malde (2018); Blume et al. (2011); Bramoullé et al. (2016); Graham and De Paula (2018); Jackson (2011); Jackson et al. (2017); Ioannides (2013); Topa and Zenou (2015).

[4]Similar recent work using exogenous group exposure, such as Boisjoly et al. (2006); Burns et al. (2016); Cai and Szeidl (2018); Harmon et al. (forthcoming); List et al. (2019); Rao (2019), explores changes in views and behaviors in response to peer-group assignments, without the consideration of friendship and network connections. In this literature, peer-effect studies using instrumental variables, as surveyed in Epple and Romano (2011), are mostly concerned with endogenous group formation and measurement errors. Notably, Foster (2006) uses a related monadic specification, constructing a monadic instrument for same-dorm sophomore peers based on freshman dorm assignments, and finds no peer effect.

assignment.[5]

Our method yields precise and powerful effects of friendship. Connecting two students with a friendship link reduces their differences in political opinions by half a point (on a scale from 1 to 10) after 6 months. The effect is equivalent to a quarter of the mean difference, and a third of its standard deviation. It is considerably larger than the OLS estimate, suggesting that complying pairs, namely those that make friends precisely because of the same integration group, experience a stronger friendship effect than others. It is also much larger than the peer effect of the tutorial groups in which students take all their classes, which stresses the importance of discerning friendship effects from non-friend peers effects using friendship data.[6]

We further find that the causal friendship effect is strongest among students with similar pre-Sciences Po political views. The evidence is consistent with what we call the "homophily-enforced mechanism," by which homophily along a dimension, such as political views, is complementary to the friendship effect on that dimension. Accordingly, between a pair of individuals with strong similarity on a dimension, friendship could make them interact much more on that dimension, consequently strengthen such similarity. In contrast, friendship might not matter much to that dimension between initially dissimilar pairs. Empirically, among politically-similar pairs, friendship strongly induces them to join the same politically-related associations, thereby likely pushes them to interact more on politics. Those pairs end up with a friendship effect on political opinions that is 50% larger than the benchmark effect. Yet, among pairs with far-apart pre-Sciences Po opinions, friendship does not push them towards the same political associations, and consequently does not produce a significant friendship effect on the subsequent political opinion gap. In short, similarity breeds friendship, which breeds similarity on the same dimension.[7]

We also discover a markedly asymmetric pattern of the friendship effect on polarization and extremism. Friendship contributes to a narrower opinion gap mostly by reducing the incidence of divergence (when two opinions drift apart), and especially among politically similar students. In contrast, friendship does not encourage two opinions to converge towards each other. Consequently, friendship lowers polarization and reduces the prevalence of extremist political views, while maintaining sufficient diversity of opinions.

Friendship effect heterogeneity also manifests by network characteristics, as we find that the ef-

[5]There is also an econometric advantage in considering generic social networks, rather than the special case of peer groups. That is, the generic nature of networks (e.g., based on friendship links) introduces identifying restrictions that avoid Manski's (1993) reflection problem in linear-in-means models with peer groups, as mentioned in Jackson (2008), and formally proven in Bramoullé et al. (2009) (also see Lee and Liu, 2010; Lin, 2010; Liu et al., 2014).

[6]This point echoes Carrell et al.'s (2013) emphasis on friendship within peer groups, and Leider et al.'s (2009) finding that, even in a peer group, directed altruism dies out after second-degree friends.

[7]This mechanism can explain the difference between this paper's sizable friendship effect and the insignificant peer effect in the tutorial group, as well as other small, sometimes insignificant peer effects on academic performance occasionally found in the literature (e.g., Angrist and Lang, 2004). When friendship is built and consolidated voluntarily on a dimension, it matters to the gap in that dimension. In contrast, in a peer effect study, peer groups may or may not be formed to reinforce interactions on the same dimension as what is measured as outcome, so peer effects are not guaranteed.

fect is stronger among close friendships and more direct social distance. It extends to second-degree friends (friends of friends), but is not present between network stars (top quartile in eigenvector centrality). Taking into account the effect on second-degree friends, the friendship effect on network can explain 20% of the reduction in overall opinion gaps in the cohort. Those findings connect directly to the recent literature on non-Bayesian learning in social networks.[8] They further stress the importance of friendship next to other major determinants of the persistence and change in beliefs, preferences, and norms.[9]

The rest of the paper unfolds as follows. Section 2 describes the study's context. Section 3 details our empirical strategy, the timing and design of our surveys, and discusses the collected data. Section 4 presents the main friendship effect on opinions and behaviors. Section 5 investigates the main drivers and mechanisms at work, section 6 shows how the friendship effect varies with network characteristics, and section 7 concludes.

2 Sciences Po background and organization

This section provides a description of the context of the natural experiment at Sciences Po, including its role in French politics and the organization of the integration week that we exploit as an exogenous source of variation in the formation of social networks. It also relates to our sample of the cohort that enters Sciences Po in 2013. Sciences Po, or the Institute of Political Studies, has always had a major role in the training of French politicians and high level civil servants, as it was explicitly conceived to provide a modern training for the French elite since its foundation in 1872 following France's defeat in the Franco-Prussia War of 1871. Between 12 to 15% of deputies of the French National Assembly elected in the last decades graduated from Sciences Po (Rouban, 2011), as well as more than fifteen percent of the mayors of cities above 30,000 inhabitants (Rouban, 2014). Sciences Po alumni are also highly present in the government, as well as at the top of the French bureaucracy.

While not all Sciences Po students want to become politician or civil servant, politics is much more important for them than for students from other universities or business schools. One tenth of the students are member of a political party, a very large proportion compared to their age

[8]In the typical non-Bayesian model of learning in networks à la DeGroot (1974), effects of connected nodes are usually modeled as homogenous and linear. The literature on social learning, as reviewed by Goyal (2011), Möbius and Rosenblat (2014), and Golub and Sadler (2016), includes both Bayesian learning (e.g., Bala and Goyal, 1998, 2001; Acemoglu et al., 2011) and non-Bayesian learning (e.g., DeGroot, 1974; DeMarzo et al., 2003; Golub and Jackson, 2010, 2012). Recent designed experiments on the sources and mechanisms of information diffusion (e.g., Chandrasekhar et al., 2018; Möbius et al., 2015; Grimm and Mengel, forthcoming) have shown an important role of non-Bayesian learning.

[9]Previously investigated mechanisms that shape values and norms include notably influence by family (Giuliano, 2007; Bisin and Verdier, 2001, 2000, 2011), exposure to socio-economic contexts (Giuliano and Spilimbergo, 2014) and peers (Rao, 2019), mass media (see review by DellaVigna and La Ferrara, 2015), and deep-root factors (e.g., Alesina et al., 2013, Giuliano and Nunn, 2017).

group. Sciences Po students are very different from the students enrolled in public universities. On average they are academically stronger, and come from a much wealthier background.

Most of Sciences Po students have not known each other before their first year starts, largely because they are competitively selected from high schools from all over France (only 5% of the students coming from abroad).[10] As in other education contexts, friendships are quickly formed within a short span of time, especially through activities that boost exposure and contact among students. The following three types of activities are the most catalytic for building friendships.

Integration groups (IGs). In the integration week just before the scholar year, the incoming cohort of undergraduates are formally introduced to Sciences Po, and assigned to IGs of around 16 each based on alphabetical order. Our sample's integration week takes place in the last week of August 2013. Students experience a variety of extra-curricular activities, such as games and guided visits of Paris, separately in those groups, in purpose of creating and solidifying links among students.[11] No activity during this week is related to academic or political matters, or students' political opinions. Individual conversations with students reveal that they remember the integration week primarily for its social activities, including bonding between new friends, and not for any other content.

Tutorial groups. Throughout the first and second years at Sciences Po, students are divided into tutorial groups of around 20 each, in which they take all tutorial classes together. The tutorials are mandatory classes that meet for two hours each week, in a total of three over each semester, each one supporting a core first-year course at Sciences Po.[12] The tutorials involve a lot of collective work on assignments and presentations, and are thus key to much of students' social interactions, and conducive to friendship formation. We will control for tutorial group membership throughout this study.[13] In our sample, students work in tutorial groups starting from September 2013.

[10] While we do not observe their high school, the incidence of having been friends from before Sciences Po, as reported in our survey, is extremely rare. In the sample of dyads of students who were assigned to the same integration group (IGs), namely the treatment group in our empirical design, there are only two pairs that were friends before Sciences Po, or 0.2% of friendship pairs who were in the same IG, and 0.02% of all friendship pairs.

[11] While the integration week has been criticized as unrealistic in fostering friendship after just one week, our first stage results in Table 2 lend credit to its designers as a surprisingly effective factor in friendship formation.

[12] The first-year compulsory courses include microeconomics, macroeconomics, history, sociology, political science, and constitutional law (three in each semester). They are taught in large weekly lectures at the same time for all 800 first-year students.

[13] The assignment into tutorial groups is based on students' choices of tutorial schedules during a very short opened window at the beginning of the year, with no information on each group's instructors, nor other relevant information except scheduled hours. As most slots run out quickly in matter of minutes, students have little control over their tutorial group assignment, and it was almost impossible to coordinate on the same group. Thus, in practice, the tutorial group assignment can be considered as arbitrary as randomized. In this paper, we use the tutorial group membership as a control variable with a meaningful coefficient, but not as an instrumental variable of friendship, because common membership in a tutorial group throughout the year also correlates with other continual factors such as the common instructors' influences, and invalidates the exclusion restriction.

Student associations. The third type of activities take place within about one hundred student associations, including notably those with close links to political parties and movements. Many meet frequently in practices (such as in sports and art associations), events, and social gatherings. Association participation is entirely voluntary, and open to all Sciences Po students of any background. In our sample, associative participation starts in September 2013.

Among many dimensions of heterogeneity that may foster homophily and hinder friendship formation between students of different backgrounds, one stands out in this context: whether a student has been admitted through an affirmative-action process called "Convention Education Prioritaire" (CEP), representing around 20% of each cohort. This admission procedure is reserved for many high schools in disadvantaged areas in France under an agreement with Sciences Po, by which their best students can apply and get admitted through dossier and oral evaluation, instead of the standard, highly competitive written contest. Compared with the rest, CEP students come from poorer families, lower socio-economic backgrounds, and many may struggle academically, at least in their first year (Tiberj, 2011).

3 Empirical design, methodology, and measurement

3.1 Empirical strategy

Our empirical design focuses on the dyadic relationship among all pairs of students (i, j), between a measure of pairwise difference DY_{ij} and a measure of friendship $Link_{ij}$,

$$DY_{ij} = g(Link_{ij}, \mathbf{X_{ij}}, \eta_{ij}).$$

Undirected friendship $Link_{ij}$ is defined as an indicator of whether i or j names the other as friend in their answers. We focus on the undirected network of friendship and use all symmetric dyadic variables.[14] DY_{ij} is the absolute difference between i and j of the variable Y, such as political opinion on a scale from 1 to 10. $\mathbf{X_{ij}}$ and η_{ij} represent respectively observable dyadic covariates and the unobservable idiosyncratic residual. $\mathbf{X_{ij}}$ gathers all observable pairwise variables representing commonness and differences across predetermined dimensions, notably the pre-Sciences Po difference in political opinions DY^0_{ij} (surveyed from a retrospective question), and others.[15] As will be

[14] We use the OR network, similar to Leider et al. (2009) and many other papers that survey friendships. The results remain robust to using the AND network.

[15] The covariates include common gender, common nationality, common academic program, common admission type (essentially regular admission versus priority admission through the affirmative action channel), common graduation with honor from high school, common district (French département) of high school, common professions of parents, common current residence's ZIP code, dummies for being both female, for being both French with double nationality, for both having to pay no tuition, and the difference in tuition fees that proxies for the difference in parents' income. We create those dyadic control variables based on the list of predetermined variables that we could collect from administrative data, and on our assessment of a priori importance of those conditions in the formation of friendship among Sciences Po students. Unfortunately, there is no better information on the precise household income, as the

shown, this list of observables has very limited explanatory power on friendship.[16]

We are generally interested in the average causal effect of friendship on pairwise differences of opinion, as $\beta_L = \mathbb{E}[DY_{ij}|Link_{ij} = 1, \mathbf{X_{ij}}] - \mathbb{E}[DY_{ij}|Link_{ij} = 0, \mathbf{X_{ij}}]$. A negative (positive) β_L means that friendship makes people's opinions closer (further apart) over the observed period of 6 months from August 2013 to March 2014, and vice versa. Under the assumption that conditional on observables, friendship is assigned exogenously, a simple OLS regression would provide an unbiased estimate of β_L. However, this conditional independence assumption is likely violated in presence of homophily in friendship formation.

The homophily bias occurs when there is a certain unobserved dimension U such that (i) individuals' similarity U_{ij} correlates with the formation of friendship links $Link_{ij}$ (homophily), and (ii) it also influences the outcome $DY_{ij} = g(Link_{ij}, \mathbf{X_{ij}}, U_{ij}, \eta_{ij})$ (outcome-relevance). When friendship formation is empirically related to U_{ij} as $U_{ij} = f(Link_{ij}, \mathbf{X_{ij}}, \varepsilon_{ij})$, then the homophily bias due to the omission of U is $\frac{\partial g}{\partial U} \times \frac{\partial f}{\partial Link}$.[17] It is larger when U is more important to the outcome, and when it is more associated with link formation. In our context, the bias likely pushes the OLS estimate away from zero. This remains a thorny issue in existing estimations of effects of network links, and one that, to our best knowledge, has not been addressed in the empirical literature using an exogenous source of variation.[18]

Our key methodological innovation consists of treating the homophily bias by instrumenting $Link_{ij}$ by the indicator IG_{ij} of whether i and j participated in the same integration group (IG) at the beginning of the year (August 2013). We argue that this instrument satisfies all the LATE conditions (Imbens and Angrist, 1994).

Instrument validity. First, in section 4.1, we will test the instrument's relevance, namely $\beta_{IG} = \mathbb{E}[Link_{ij}|IG_{ij} = 1, \mathbf{X_{ij}}] - \mathbb{E}[Link_{ij}|IGij = 0, \mathbf{X_{ij}}] \neq 0$. This first stage condition is satisfied if the integration week is a strong enough catalyst to form lasting friendships among students.

Second, this instrument's exogeneity is based on the mechanism of assignment into IGs by *alphabetical order* of the family name, arguably independent from individual characteristics that matter to the formation of links. We will further test the claim of exogeneity in a balance test in

administrative data have a very high rate of missing observations for this question. The inclusion or exclusion of any dyadic control variable does not make any noticeable qualitative or quantitative difference to our results.

[16] In addition, we also include TG_{ij}, a dyadic indicator whether two students are members of the same tutorial group or not. Since membership in a specific tutorial group is essentially arbitrary, a linear regression can estimate the causal average effect of being in the same tutorial group on the outcome, namely $\beta_{TG} = \mathbb{E}[DY_{ij}|TG_{ij} = 1, Link_{ij}, \mathbf{X_{ij}}] - \mathbb{E}[DY_{ij}|TG_{ij} = 0, Link_{ij}, \mathbf{X_{ij}}]$ (the partial effect of TG_{ij} on $f(\cdot)$). It is then possible to compare the different effects of tutorial group membership β_{TG} and friendship β_L on opinion differences.

[17] More precisely, the partial derivatives denote corresponding regression coefficients, controlling for covariates $\mathbf{X_{ij}}$. The direction of causality in those regressions does not matter to the homophily bias.

[18] In the spirit of Altonji et al. (2005), we may gauge the size of this bias due to unobservables by estimating the bias when the observables $\mathbf{X_{ij}}$ are deliberately omitted, and then argue that the homophily bias due to U is of the same order of magnitude. However, this method is weak and unreliable in case $\mathbf{X_{ij}}$ only explains a small fraction of the variation in $Link_{ij}$.

section 3.4.

Third, the instrument IG_{ij} arguably satisfies the exclusion restriction. The integration week was exclusively meant to facilitate students' familiarization and socialization with their new peers and new environment in Paris, without any academic- or political-related activities. The IGs are dissolved after that week, and does not relate to any other academic or extra-curricular activities afterwards.[19] Hence, it should have no meaningful channel to affect the formation and adjustment of individual opinions six months later, which guarantees the exclusion restriction of the instrument.[20]

Fourth, it is natural to make the monotonicity assumption that being in the same IG always (weakly) increases the incidence of friendship formation for any pair of potential friends, such that $f(IG_{ij} = 1, \mathbf{X_{ij}}, \varepsilon_{ij}) \geq f(IG_{ij} = 0, \mathbf{X_{ij}}, \varepsilon_{ij})\ \forall(i, j)$.[21]

Taken together, those four assumptions guarantee a *causal* LATE interpretation of our estimate (Imbens and Angrist, 1994). That is, our IV estimate can be interpreted as the average causal effect of friendship on the "compliers" pairs of students, namely those who would have become friends thanks to being in the same IG group in the integration week. Since this is a condition that characterizes a rather strongly-complying group of student pairs (for instance, pairs that only become friends after weeks or months of acquaintance are not included), we remain cautious in generalizing our estimates to all possible pairs of Sciences Po students. However, in Imbens's (2010) spirit of "better LATE than nothing", we argue that the correct estimation of the LATE in our context already lays strong ground for further research on transmission of beliefs among students.

Analysis of compliers. Since the LATE is defined over the compliers, it is useful to characterize those pairs in order to better understand the potential difference between the IV and OLS estimates. Appendix A describes the calculation of the sample share of each group, as well as any distributional statistics within each group. We then compare the difference between treated and untreated compliers, from which the LATE estimate obtains, and the OLS estimate that likely draws more from never-takers and always-takers.

[19] No subsequent academic or extra-curricular activities among Sciences Po students are organized based on alphabetical order.

[20] Kitagawa (2015) shows that the exclusion restriction can be jointly tested with the other LATE assumptions. In that spirit, we examine the inferred distributions of outcomes in the two subsamples of treated and untreated compliers, by drawing respectively the distribution of unassigned never-takers with and without unassigned compliers in Appendix Figure A1, and the distribution of assigned always-takers with and without assigned compliers in Appendix Figure A2 (in each case, the difference between the two plotted distributions is the distribution of corresponding compliers). We verify that the density of compliers in each case is nonnegative, thus the LATE assumptions cannot be rejected.

[21] Its violation would mean the rather improbable event of a "defier" pair that would have become friends had they not met in the same IG, but would not have become friends because they met early in the same IG. Even without the monotonicity assumption, de Chaisemartin (2017) shows that, under a much weaker condition, one could still interpret the IV estimator as the Average Treatment Effect among a subgroup of compliers.

Robustness tests. Two types of identification concerns merit further robustness tests. First, compliance in the IGs might be imperfect, as students may refuse to follow their assigned group. We address this issue in a test using an instrument based on the alphabetical distance between names that approximates the designed IG structure. To construct it, we first rank all last names in alphabetical order, assign the rank distance $AlphRank_i$ to each student i, and compute the alphabetical rank distance $AlphDist_{ij} = |AlphRank_i - AlphaRank_j|$. We then use $\min(AlphDist_{ij}, 16)$ as instrument for friendship L_{ij}. This instrument's logic is that by initial assignment (and independent of students' choice to comply with this assignment), two names with a shorter alphabetical distance between them are more likely to fall into the same IG, and then are more likely to become friends. However, this first stage should disappear beyond 16, the standard size of the IG – thus the truncation of $AlphDist_{ij}$ at 16.

Second, there is a potential concern that the alphabetical rank of certain family names may be correlated with confounding characteristics such as ethnic origin.[22] To address this concern, we first run simple "jackknife" tests, in each of which we drop all names starting with a specific letter, or all students with a specific non-French nationality.

Next, we further strengthen our IV approach by restricting the sample to only pairs of students whose alphabetical distance is sufficiently close. Intuitively, we consider same-group and different-group pairs of students within a "bandwidth" of the cutoff between two consecutive groups. Analogous to the logic of a Regression Discontinuity Design, around the threshold between two groups, same-group and different-group pairs are almost identical in both observable and unobservable characteristics (Lee and Lemieux, 2010), which reinforces the identification assumption of exogeneity of IG assignment.[23]

Statistical inference. In a dyadic setting, each individual is repeated in her pairs with all other students, resulting in natural correlations between the residual terms of pairs sharing an individual. Furthermore, there can naturally occur common shocks within the same group, such as teacher's biases, that could drive all group members' opinions. While those shocks are uncorrelated to our instrument, and cannot bias our IV estimates, they produce clustered standard errors, and must be taken care of in order to obtain correct standard errors and confidence intervals.

Throughout the paper, we choose to correct for potential clustered standard errors by a two-way group clustering strategy. That is, we allow for arbitrary correlations in the idiosyncratic

[22] For example, the large share of Zhang, Zhao, and other names starting with Z among Chinese, Nguyen among Vietnamese, and Kim and Park among Koreans, may over-populate certain IGs with same-ethnic students. In reality, there are almost none of those ethnicities in our sample, and we do not observe this phenomenon of ethnic clustering. The remaining concern is that French family names starting with "de" might correspond to an aristocratic background.

[23] While similar, this is not a proper Regression Discontinuity Design, since the exact cutoff is unknown due to partial compliance. It is thus not possible to implement standard RDD methods, or choose an optimal bandwidth. We pick the restriction that $AlphDist_{ij} \leq 24$, at 3% of the range of 800.

component η_{ij} between any pair of observations that overlap in a group.[24] We make sure results are robust to different types of clustering correction.

3.2 Survey design and data sources

We conduct our major internet-based survey in March 2014 on the cohort of Sciences Po first-year students who start in September 2013. We offer strong material incentives in the first survey in the form of a lottery for fifty mini iPads at approximately 300 Euros each (each student has an average probability of about 9% to win one). We seek a high rate of participation to avoid the problem of complex biases in network measures due to missing information on network structure (Chandrasekhar and Lewis, 2011). Eventually, 68.4% (547 out of 800) of the students answer to at least some question in the survey, and 65.6% (526 out of 800) complete the whole survey. This is about the same level of participation as the best-participated studies of social networks of students, such as Leider et al. (2009) or Goeree et al. (2010). It is well above the standard participation rate of around 20% found in studies using online surveys (Cantoni et al., 2017).[25]

In order to incentivize truthful answers, we design the elicitation of friendships as a coordination game, similarly to Leider et al. (2009). Not only do we ask students to name a list of friends of up to 10 names, but we also ask how they meet each of them, how much time they spend together, and in which activities, and how strong do they evaluate their relationships. We announce in the survey that their answers would be cross-checked with those of the other students, and that if both answers match, they would gain points , later converted into an additional probability of winning the iPad. We do not disclose the exact mechanism, in order to avoid that some students engaged into strategic behavior and try to actively coordinate with other people. The survey is carried out during a vacation week, which limits the possibility for the students to interact with each other and to complete the survey together. To further avoid the possibility of collusion on the friendship questions, we censor the top 5% of the sample by the amount of time spent on the friendship question, in order to avoid individuals who have spent too much time pondering this question.[26]

[24] Cameron and Miller (2014) discusses Fafchamps and Gubert (2007) method to fully account for all possible correlations between all dyads that overlap with a group or share an individual. Unfortunately in this case, Cameron et al. (2011) decomposition of the the sandwich formula for standard errors (used for a fast, economical calculation of the two-way clustering correction) becomes intractable. The only possible implementation is to undertake the full calculation of Fafchamps and Gubert's formula, which requires an excessive amount of computing memory and time, given our large sample size. Therefore, throughout our paper, we choose to implement a simplified version of this method, in which we allow for non-zero correlations between any residual terms η_{ij} and $\eta_{i'j'}$ such that either i and i' belong to the same group, or j and j' belong to the same group, or both (thus ignoring the possible same-group memberships of i and j', or of i' and j. We also have fully implemented Fafchamps and Gubert (2007) formula in a few benchmark regressions, and found similar and better levels of standard errors and p-value.

[25] A second survey conducted in June 2015 on the same cohort is unfortunately much less well-funded, and only attracts 300 participants. Overall, there are 235 students who have completed in both surveys. The paper makes most use of the first survey, while the second only serves in robustness checks.

[26] This is equivalent to dropping individuals who spend more than 81.625 seconds per friend on that question. The results remain practically the same over a broad range of possibilities of right tail censorship. Right tail censorship looks necessary, given that at the top of the distribution certain students spend up to half an hour per friend. Results

We also require that they complete the whole questionnaire in order to be included in the lottery.

The second part of the survey is devoted to questions about political opinion and values. We ask students' current political opinion, and that before their arrival at Sciences Po in August 2013. These questions use a common scale from 1 to 10 (1 being extreme left and 10 extreme right). The survey also provides information on their political participation, and any participation in associations at Sciences Po.

We further obtain Sciences Po's administrative data including student characteristics used as covariates (see list in section 3.1), and their IG and tutorial groups.

3.3 Data description

We consider the (symmetric) OR network in which two students are linked if at least one nominates the other. Table 1 Panel A describes the quality of the network survey. About half of the nominated friends reciprocate, a considerably larger rate than in the literature since Leider et al. (2009). The probabilities of a well-matched answer in terms of the context of the first meeting between the two friends, of the amount of time spent every week, of the type of activities mostly spent together, and of the self-evaluated strength of friendship are respectively 76%, 52%, 46%, and 52%, quite larger than in Leider et al. (2009). If answers are completely made up and randomized, the probability of matching on any of those dimensions would be rather low, given that respondents have many choices for each answer (especially in the question on the context of their first meeting). Taken together, those statistics imply that the survey answers are indeed very reliable, especially for the purpose of picking up friendships.

Panel B reports the major statistics on the number of friends and the social network structure. The average and maximum number of nominated friends per student is 8.8 and 21, respectively, with a very high variance.[27] Moreover, there seems to be some small world properties with a very small average path length (3.7) and a relatively small diameter (9). The clustering coefficient is also relatively high, which means that roughly 25 percent of students have friends of friends who are friends. In terms of network position, the mean eigenvector centrality is relatively low (0.0361).

Panel C shows the descriptive statistics of the friendship dyadic measures. We distinguish between the *full sample* (column 1) of all students who have participated and the *benchmark sample* (column 2) that corresponds to the benchmark regression (the two samples differ slightly because of certain missing values). By nature, the share of measured friendship links is relatively small at 1.6%, and that of second and third order indirect links are larger at 9.3% and 38%, respectively. The dyadic same group variables are of similar magnitudes, at an average of 1.6% for same IG,

are also robust to left tail censorship, although the case for censorship is much less clear, as the fastest answers still took an acceptable amount of time (more than 10 seconds on average).

[27] Even if the maximum number of friends that someone can nominate is 10, a student can have 21 friends since we use an undirected network approach so that a friend is assigned to a person if either her or her friend has nominated the other.

Table 1: Descriptive Statistics

Panel A: Quality of the Survey

	(1) Full Sample	(2) Benchmark Sample
Number of reported friends	8.234	8.613
	(2.522)	(1.984)
Probability of reciprocal friend	0.461	0.479
	(0.499)	(0.500)
Correct answer: meeting	0.800	0.815
	(0.400)	(0.389)
Correct answer : time spent	0.483	0.497
	(0.500)	(0.501)
Correct answer : activity	0.568	0.587
	(0.496)	(0.493)
Correct answer : strength of the relationship	0.532	0.532
	(0.499)	(0.500)

Notes: Summary statistics (1) refer to the full sample, where full sample is defined as the set of all pairs for which both members named at least one friend or stated that they have no friends in Sciences Po. Summary statistics (2) refer to the benchmark sample as detailed in Table A1.

Panel B: "OR" Network statistics

Mean of degree per individual	8.8625
Variance of degree per individual	18.4842
Median of degree per individual	10
Maximum of degree per individual	21
Minimum of degree per individual	0
Diameter of the network	9
Average path length	3.7008
Overall clustering coefficient	0.241
Average clustering coefficient	0.271
Mean eigenvector centrality	0.0361
Standard deviation of eigenvector centrality	0.0200

Notes: Summary statistics are computed on the full sample.

Panel C: Dyadic Links and Groups

Variable	(1) Full Sample			(2) Benchmark Sample		
	Mean	Standard deviation	Obs.	Mean	Standard deviation	Obs.
Friendship	0.0160	(0.1240)	147,153	0.0170	(0.1300)	54,615
2nd Order Links	0.0930	(0.2900)	147,153	0.0990	(0.2990)	54,615
3rd Order Links	0.3800	(0.4850)	147,153	0.4020	(0.4900)	54,615
Mere relationship (strength 1)	0.0014	(0.0382)	147,153	0.0017	(0.0421)	54,615
Friendship link (strength 2)	0.0063	(0.0791)	147,153	0.0068	(0.082)	54,615
Close friendship (strength 3)	0.0041	(0.0642)	147,153	0.0045	(0.067)	54,615
Very close friendship (strength 4)	0.0035	(0.0593)	147,153	0.0040	(0.0632)	54,615
Same Integration Group	0.0160	(0.1280)	147,153	0.0180	(0.1330)	54,615
Same Tutorial Group	0.0230	(0.1490)	147,153	0.0230	(0.1500)	54,615

Notes: Summary statistics (1) refer to the full dyadic sample, where full sample is defined as the set of all pairs for which both members named at least one friend or stated that they have no friends in Sciences Po. Summary statistics (2) refer to the benchmark dyadic sample as detailed in Table A1.

Panel D: Monadic Dependent Variables

Variable	(1) Full Sample			(2) Benchmark Sample		
	Mean	Standard deviation	Obs.	Mean	Standard deviation	Obs.
Pre-Sciences Po Political Opinion (in 2013) (1-10)	5.108	(1.958)	463	5.148	(1.934)	331
Political Opinion in 2014 (1-10)	5.044	(1.755)	472	5.091	(1.712)	331
Political Opinion in 2014 as recalled in 2015	4.913	(1.650)	287	4.941	(1.642)	331
Political Opinion in 2015	4.853	(1.807)	285	4.818	(1.746)	331
Enrollment in a Political Party in 2014 (yes / no)	0.104	(0.303)	521	0.121	(0.326)	331
Enrollment in a Political Party in 2013 (yes / no)	0.067	(0.249)	519	0.076	(0.265)	331
Enrollment in an Association in 2014	0.597	(0.491)	499	0.642	(0.480)	330

Notes: Summary statistics (1) refer to the full individual sample, where the full sample is made of all the individual observations for which the variable described is not missing. Summary statistics (2) refer to the benchmark sample, where the benchmark sample is defined as the individual sample containing all the individuals that are present in our benchmark dyadic sample as detailed in Table A1.

Panel E: Dyadic Dependent Variables

Variable	(1) Full Sample			(2) Benchmark Sample		
	Mean	Standard deviation	Obs.	Mean	Standard deviation	Obs.
Difference in Political Opinion in 2014	1.932	(1.467)	105,111	1.926	(1.468)	54,615
Initial difference in Political Opinion (2013)	2.211	(1.631)	101,025	2.200	(1.623)	54,615
Difference in Political Opinion in 2015	2.014	(1.538)	27,027	1.940	(1.496)	15,920
Difference in Political Opinion in 2014 (as recalled in 2015)	1.835	(1.424)	126,756	1.798	(1.412)	15,920
Membership in Some Political Party in 2014	0.815	(0.388)	127,260	0.785	(0.411)	53,628
Membership in Some Political Party in 2013	0.874	(0.332)	126,756	0.858	(0.349)	52,975
Membership in the Same Political Party in 2014	0.827	(0.379)	1,326	0.800	(0.400)	780
Membership in Some Association	0.479	(0.500)	114,960	0.461	(0.498)	47,895
Membership in the Same Association	0.915	(0.279)	52,003	0.903	(0.296)	24,310

Notes: Summary statistics (1) refer to the full dyadic sample, where full sample is defined as the set of all pairs for which both members named at least one friend or stated that they have no friends in Sciences Po. Summary statistics (2) refer to the benchmark dyadic sample as detailed in Table A1.

and 2.3% for same tutorial groups. The friendships are partitioned rather evenly across different levels of friendship strength, especially from 2 (ordinary friends) to 4 (very close friends). We also observe that there is little difference between the full sample and the benchmark sample.

Panel D lists the descriptive statistics of students' political opinion and behavior. While political opinion slightly shifts to center-left over time (i.e., to lower value, as 5.5 represents the center), participation in political parties has increased substantially. Meanwhile, the variance of political opinion decreases by 24 percent, as the measured standard deviation of opinions in March 2014 is only 1.76 on a scale of 1 to 10.[28]

Figure 1 shows the distributions of political opinions in March 2014 (orange) and in August 2013 (green). The bimodal distribution in 2013, with two modes at 4 and 7 corresponding to rather mainstream left-right politics, becomes unimodal in 2014 with strongly dominant center in 5-6. That fact, and a strong reduction in right to extreme right positions (8-9-10), altogether explains the net decrease in variance of opinion.

3.4 Exogenous assignment mechanisms and balance test

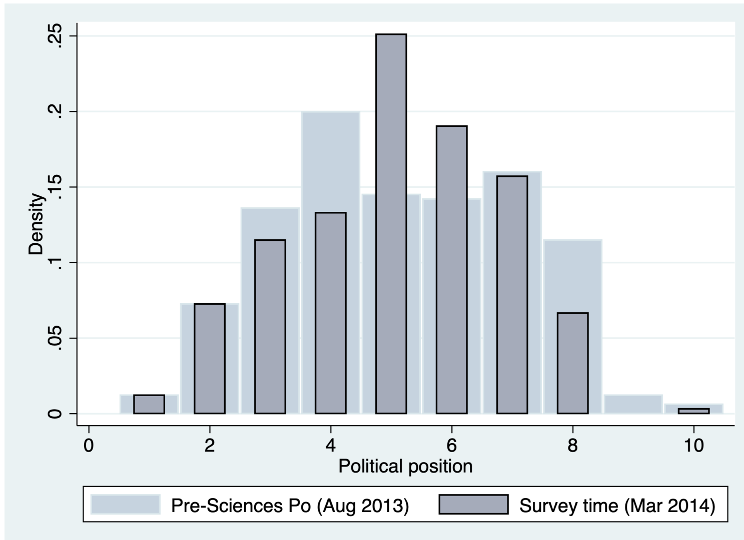
Our instrumental variable strategy depends fundamentally on the claim that the assignment into IGs by alphabetical order of the students' family names is exogenous. We check that alphabetically close family names do not carry other information that could stack up students with similar backgrounds in the same group. First, we show in Table 2 the results of a balance test of exogeneity in a linear regression of IG_{ij} on all observable pairwise covariates.[29] In comparison with the mean and standard deviation of the dyadic variable $Link_{ij}$ (respectively 0.017 and 0.13, as shown in Table 1.B), all coefficients in Table 2 are very small and mostly insignificant. Among the few significant coefficients, it is natural to find the variable common program, which specifies mostly dual-degree programs joint with other universities, in which students simultaneously take courses at both Sciences Po and the other institution, therefore are grouped together since the integration week. We make sure to control for this covariate throughout the paper.[30] (Removing all students from double-degree programs leaves all estimates practically unaltered.) Finally, another significant coefficient is that of the same high school major. As its magnitude is also small, and its sign is opposite to that predicted by homophily, we interpret it as significant by chance.

[28] Appendix Table A1 describes in details all variable definitions, and Appendix Table A2 completes the descriptive statistics of other variables used in the empirical analysis.

[29] Standard errors are clustered two-way among observations sharing either student i or student j. This clustering approach is less conservative than the standard two-way clustering by i's and j's groups carried out throughout the paper, so we expect to have higher power in detecting observables that significantly correlate with IG.

[30] This is the standard method to deal with imbalances in experimental samples. See Bruhn and McKenzie (2009) for a discussion on common practices of balance tests in field experiments.

Figure 1: Distributions of Political Opinions



Notes: Distributions of Individual Political Opinions just before joining Sciences Po (August 2013) and at the time of survey (March 2014).

Table 2: Balance Test of Integration Group

Dependent Variable	Same Integration Group	Dependent Variable	Same Integration Group
Same Gender	0.000673	Same Region of High School	0.00170
	(0.00148)		(0.00156)
Both Female	-0.00194	Same High School Major	-0.00203*
	(0.00135)		(0.00105)
Same Nationality	0.00622	Diff. in Tuition Fees	-3.18e-07
	(0.00449)		(2.35e-07)
Both French with Double Nationality	-0.00206	Both Free Tuition	-0.00138
	(0.00421)		(0.00115)
Same Admission Type	0.00137	Same Parents Profession	0.00103
	(0.00116)		(0.00130)
Both Priority Admission	0.00182	Same ZIP Code	-0.000761
	(0.00697)		(0.00337)
Same Département of High School	0.00299	Same Program	0.00270**
	(0.00420)		(0.00121)
Observations		54,615	
R-squared		0.001	
F-stat		2.957	

Notes: OLS regression of *Same Integration Group* on covariates. The benchmark sample is detailed in Appendix Table A1. F-stats are for the joint significance of the covariates. Standard errors in brackets are two-way clustered by individual 1's group and by individual 2's group.

Since same-group membership is not a simple, i.i.d. near-randomized treatment, we also propose permutation tests of the exogeneity of IGs with respect to observable characteristics. The major concern is that there is selection into groups by certain individual characteristics. We frame this concern as the alternative hypothesis, against the null hypothesis of randomized IGs, and consider two statistics that measure within-group variation in each characteristics: the average within-group standard deviation for continuous variables such as political opinions and tuition fees, and the average Herfindahl-Hirschman Index (HHI) of concentration for categorical variables.[31] Under the alternative hypothesis, there is too little within-group variation compared to the permuted distribution of the statistics. We thus run a left-tail test for the within-group standard deviation, and a right-tail test for the within-group HHI. Results shown in Appendix Table A3 largely confirm that we cannot reject the randomization of IGs.

4 Friendship effect on opinions and behaviors

4.1 Same-group exposure and friendship formation

We first establish the relevance of the instrumental variable IG_{ij}, i.e., the causal effect of participating in the same IG in August 2013 on forming and maintaining a lasting friendship 6 months later. Columns (1) and (2) of Table 3 present the regression of $Link_{ij}$ on IG_{ij}, with and without observable covariates $\mathbf{X_{ij}}$, yielding an estimate of $\beta_{IG} = \mathbb{E}[Link_{ij}|IG_{ij} = 1, \mathbf{X_{ij}}] - \mathbb{E}[Link_{ij}|IGij = 0, \mathbf{X_{ij}}]$ in the coefficient of IG_{ij} of around 16%.[32] In comparison, column (3) shows the corresponding effect of the tutorial group (TG_{ij}) on friendship formation at around 36%.

It is remarkable that this coefficient is more than 10 times larger than any coefficient on observable predetermined characteristics (the next largest coefficients are on students' ZIP code and high school département.), and about half the size of the same tutorial group coefficient.[33] It shows that "exposure by chance" to other students during the first week of a student's college life has an effect on friendship formation several orders of magnitude larger than that of most predetermined characteristics that are typically observed or obtained from administrative records.

The result can be further interpreted as evidence of the first week's special role as a "window of opportunity" for friendship formation. In perspective, the one-week-long integration exposure has about half the power of constant weekly exposure in the tutorial groups during 6 months (which also includes off class interactions due to academic assignments in tutorial groups). It is plausible that friendships tend to form at the beginning of college, in activities meant to facilitate socialization

[31] A group's HHI is calculated as the sum $\sum_k s_k^2$ over the shares of all categories k. Each IG's HHI is weighted by group size.

[32] This table's F-statistics are taken from OLS regressions without the required adjustment for clustered standard errors. Table 4 shows the corrected Kleibergen-Paap cluster-robust F-statistics that account for this issue.

[33] The interpretation of the coefficient on differences in tuition fees, themselves a function of family income bracket, relies on the tuition fee scale from zero (i.e., a full scholarship that a fifth of each cohort receives) to full tuition of 10,000 euros.

Table 3: Same Group Membership and Friendship Formation (First Stage)

Dependent Variable:	Friendship					
	(1)	(2)	(3)	(4)	(5)	(6)
Same Integration Group	0.166***	0.165***	-	-	0.168***	0.169***
	(0.0154)	(0.0153)			(0.0140)	(0.0140)
Same Tutorial Group	-	-	0.355***	-	-	-
			(0.0208)			
Alphabetical Distance	-	-	-	-0.00970***	-	-
				(0.000984)		
Initial Diff. in Political Opinion (August 2013)	-	-0.000861**	-0.000760*	-0.000790*	-0.00469**	-0.00375*
		(0.000431)	(0.000392)	(0.000429)	(0.00228)	(0.00208)
Same Gender	-	0.0133***	0.0116***	0.0135***	0.0457***	0.0298***
		(0.000968)	(0.00165)	(0.00111)	(0.0111)	(0.00925)
Both Female	-	-0.0107***	-0.00825***	-0.0109***	-0.0327**	-0.0167
		(0.000171)	(0.00179)	(0.000329)	(0.0133)	(0.0118)
Same Nationality	-	0.00393	-0.0120	0.00509	0.0286	-0.00916
		(0.00406)	(0.0150)	(0.00366)	(0.0364)	(0.0324)
Both Mixed French Nationality	-	0.00127	-0.0139	0.00162	0.0332	-0.0100
		(0.00528)	(0.0143)	(0.00495)	(0.0394)	(0.0350)
Same Admission Type	-	0.00531***	0.00499***	0.00556***	0.00739	0.00297
		(0.00155)	(0.00128)	(0.00164)	(0.00675)	(0.00677)
Both Priority Admission	-	-0.00431	0.00215	-0.00468	0.00634	-0.0108
		(0.00750)	(0.00622)	(0.00746)	(0.0402)	(0.0247)
Same Département of High School	-	0.0112***	0.0117***	0.0115***	0.0153	0.000197
		(0.00340)	(0.00371)	(0.00345)	(0.0185)	(0.0168)
Same Region of High School	-	0.00130	0.00249*	0.00155	0.00482	0.00583
		(0.00168)	(0.00147)	(0.00173)	(0.0113)	(0.00902)
Same High School Major	-	0.00558***	0.00348***	0.00532***	0.0292***	0.0275***
		(0.00199)	(0.000969)	(0.00199)	(0.00944)	(0.00776)
Diff. in Tuition Fees	-	-6.03e-07*	-4.95e-07**	-6.16e-07**	-4.34e-06***	-3.66e-06***
		(3.10e-07)	(2.19e-07)	(3.10e-07)	(1.68e-06)	(1.39e-06)
Both Free Tuition	-	0.00117	0.00132	0.00120	-0.00929	-0.00369
		(0.00137)	(0.00128)	(0.00139)	(0.0105)	(0.00927)
Same Parents Profession	-	0.00123	0.00126	0.00135	0.00575	0.00343
		(0.00124)	(0.00101)	(0.00130)	(0.00877)	(0.00658)
Same ZIP Code	-	0.0150***	0.0127***	0.0148***	0.0215	0.0191
		(0.00419)	(0.00489)	(0.00422)	(0.0240)	(0.0203)
Same Program	-	0.0240***	0.00825***	0.0244***	0.0344***	0.0380***
		(0.00336)	(0.00168)	(0.00333)	(0.00457)	(0.00779)
Observations	54,615	54,615	54,615	54,615	3,309	3,902
R-squared	0.029	0.041	0.175	0.030	0.124	0.120
Double Group Clustering	Yes	Yes	Yes	Yes	Yes	Yes
F-stat	114.2	15.84	87.79	10.51	34.30	26.02

Notes: Standard errors are two-way clustered by individual 1's group and by individual 2's group. F-stats are for the joint significance of the variables included in the model. The benchmark sample is detailed in Appendix Table A1. Controls include the following variables: Same Tutorial Group, Same Gender, BothBoth Female Female, Same Nationality, Both Mixed French Nationality, Same Admission Type, Both Priority Admission, Same District of High School, Same Region of High School, Same High School Major, Diff. in Tuition Fees, Both Free Tuition, Same Parents Profession, Same ZIP Code, Same Program. Column (5) restricts the sample to pairs within a short alphabetical distance (of 24, or 3% of the range of 800). Column (6) restricts the sample to pairs with the same first letter of the last names.

with same-cohort peers, and familiarization with a completely new environment, when everyone's stock of college friends is almost zero. What is perhaps more striking is that those friendships tend to last much longer beyond the window of opportunity, despite the subsequent full exposure to the whole cohort of first year students.

Regarding alternative IV strategies (see section 3.1 for details), column (4) shows the effect of a more primitive instrument, the alphabetical distance between two last names (censored at 16, the average group size) on their potential friendship. Column (5) restricts on pairs at alphabetical distance smaller than 25, and column (6) focuses on pairs whose last names share the same first letter. Those restrictions consider subsamples of pairs that are very close by alphabetical order, following an RDD-similar logic. The common instrument of same IG remains strongly predictive of friendship in those subsamples, with very similar coefficients to those from columns (1) and (2).

Table 3 also shows that homophily plays a statistically significant role in friendship formation, notably along the dimensions of political opinions, gender, background and origin (such as département and region of high school, or admission category), interest (the type of high school major), and family income (proxied by the level of tuition fees). However, its role is rather limited, as shown by the very small size of all coefficients on predetermined characteristics. The two variables IG_{ij} and TG_{ij} combined explain about 20% of the variation in friendship formation, while all the variables based on predetermined characteristics only add 1% to the R-squared. The inclusion of predetermined covariates does not alter the coefficient of IG_{ij}.

Overall, the results in Table 3 clearly confirms the relevance of the instrumental variable IG_{ij}. They also highlight the importance of "encounters by chance" among similar individuals during a window of opportunity in formation of friendship links, and a much weaker role of homophily based on predetermined observable characteristics.

4.2 Friendship effect on opinion differences

In columns (1) and (2) of Table 4, OLS regressions of differences in political opinions on friendship, controlling for pre-Sciences Po differences and eventually the other covariates, show that on average pairs of friends have lower differences in political opinion. The magnitude of the coefficient of friendship is rather stable, ranging from 0.10 to 0.13, or 5-6% of the mean difference 1.93, and 7-8% of the standard deviation 1.47 (Table 1 Panel E), and statistically significant at 5%. The reduction of the coefficient size from 0.13 to 0.10 after column (2)'s inclusion of control variables suggests that there is some homophily bias due to the omission of observable covariates. It is thus important to address the potential homophily bias due to unobservables.

Columns (3) and (4) present the estimated LATE of friendship on differences in political opinions (controlling for initial differences), using the IV strategy described in section 3.1. The estimate varies from -0.49 to -0.54, both statistically significant at 5%. They imply that the causal effect of

Table 4: Friendship and Differences in Political Opinion

Dependent Variable:	Difference in Political Opinion (March 2014)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	OLS	IV	IV	OLS	IV	IV	IV
Instrumental Variable:	No		Same Integration Group		No	Alphabetical Distance	Same Integration Group	
Sample of Pairs:	Full Sample						Alphabetical Distance ≤ 24	Same First Letter
Friendship	-0.127***	-0.098**	-0.489**	-0.538**	-	-0.561**	-0.523**	-0.763**
	(0.041)	(0.049)	(0.217)	(0.229)	-	(0.285)	(0.263)	(0.331)
Same Tutorial Group	-	-	-	-	-0.014	-	-	-
					(0.034)			
Initial Diff. in Political Opinion (August 2013)	0.527***	0.526***	0.526***	0.526***	0.526***	0.526***	0.557***	0.555***
	(0.026)	(0.025)	(0.026)	(0.025)	(0.025)	(0.025)	(0.031)	(0.032)
Observations	54,615	54,615	54,615	54,615	54,615	54,615	3,309	3,902
Controls	No	Yes	No	Yes	Yes	Yes	Yes	Yes
Double Group Clustering	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
WeakIV test stat.	-	-	114.334	113.229	-	94.962	139.583	141.281

Notes: Standard errors are two-way clustered by individual 1's group and by individual 2's group. The benchmark sample is detailed in Appendix Table A1. Controls include the following variables: Same Tutorial Group, Same Gender, Both Female, Same Nationality, Both French with Double Nationality, Same Admission Type, Both Priority Admission, Same District of High School, Same Region of High School, Same High School Major, Diff. in Tuition Fees, Both Free Tuition, Same Parents Profession, Same ZIP Code, Same Program. Alphabetical distance is computed as the pairwise difference in the ranks of all last names in the cohort. Column (6) censors it at 16 (average integration group size). Column (7) restricts the sample to pairs within a short distance (of 24, or 3% of the range of 800). Column (8) restricts the sample to pairs with the same first letter of the last names. Weak IV stat reports the Kleibergen-Paap cluster-robust statistic, distributed as a Chi-squared under the null hypothesis of weak identification.

friendship on opinion differences among compliers, over the course of the first 6 months at Sciences Po, is about half a point on a 1-to-10 scale, which approximates a quarter of the mean difference, and a third of the standard deviation of the differences. Given the usual caveats of extrapolation, a linear extrapolation to 24 months spent at Sciences Po[34] implies an effect equivalent to the average pairwise difference.[35]

A few remarks follow from Table 4. First, it is useful to consider the case of an econometrician who does not observe friendship data, and could only regress differences in political opinions to peer-group membership, such as that in the tutorial group. In that empirical exercise, shown in column (5), the coefficient of same-tutorial-group membership TG_{ij} is only -0.01 and not statistically significant at conventional levels.[36] It represents the "peer effect" often discussed in the literature (Sacerdote, 2011, 2014), which is a weighted average of effects among pairs of friends and pairs of non-friends. The comparison between peer effect and friendship effect resonates Carrell et al.'s (2013) finding that most of its peer-group effect can be attributed to an effect among friendships endogenously chosen by individuals.

Second, a "back of the envelope" calculation shows how much friendship links have contributed to the reduction of the total pairwise differences in political opinions in the sample from before Sciences Po until the survey. Per dyad, there is on average 0.0170 friendships, so an effect of -0.538

[34] The undergraduate program at Sciences Po includes two years at its campus and one exchange year abroad.

[35] Appendix Table A7 shows results from an alternative specification that uses the same IV to estimate friendship's LATE on the change in pairwise differences in opinions, and not controlling for initial differences. The effect is larger, at around 0.9 points on the scale from 1 to 10.

[36] Without covariates, the coefficient is -0.06, with a p-value of 0.007.

can explain $\frac{0.0538 \times 0.0170}{2.2 - 1.926} = 3.34\%$ of the change in total pairwise differences (see Table 1 Panels C and D). This modest proportion is due to the very low frequency of direct friendships in the dyadic sample. In section 6 we will re-examine this accounting exercise with the effect of second-degree friendships.

4.3 Analysis of compliers

Table 4 also shows that the IV estimates are four times larger in absolute value than the OLS estimates. This difference is unlikely due to a correction of the homophily bias (which alone would have predicted a smaller IV estimate), but rather to the LATE interpretation of the average treatment effect among compliers. Appendix Table A4 analyzes and compares some statistics of compliers with those of never-takers and always-takers. The average opinion gap among untreated compliers (after controlling for covariates) is 0.31, while if they are treated the average opinion gap is reduced to -0.20 (the difference between the two constitutes the LATE estimate). In the OLS regression of opinion gap on friendship link, the estimate will be dominated by the comparison between pairs with $L = 0, IG = 0$ (unassigned to integration group, no friendship linnk) and pairs with $L = 1, IG = 0$ (unassigned always-takers), given the sizes of those subpopulations. Consequently, the OLS estimate should be much closer to zero, as seen in Table 4, because it essentially compares between groups that cannot be comparable (never-takers plus compliers versus always-takers). The strong effect among compliers thus dominates the homophily bias.[37]

Why are average opinions among treated and untreated compliers very different from those of never-takers and alwayss-takers? In our context, compliers are the pairs who befriend easily only because of the exposure during the integration week, i.e., complier pairs may have strong similarities along some unobserved dimension, but also face natural barriers based on differences in background or appearance. We can see such differences by the division between affirmative action students (coming from very disadvantaged background) and the rest. Appendix Table A4 shows that pairs between an affirmative action student and a non-affirmative action student are heavily represented among compliers (29% in $L = 0, IG = 0$, 23% in $L = 1, IG = 1$), compared to the overall share (20% among never-takers, or around 22% overall) and to that among always-takers (only 11%). Thus friendship can be seen as particularly strong in reducing the opinion gap between those pairs.[38]

[37] Appendix Figure A3 further compares the distributions of the opinion gap (after controlling for covariates) of untreated and treated compliers. The difference clearly shows that treated compliers have smaller opinion gaps.

[38] Using the benchmark specification, we also obtain a strong estimate of the friendship effect among pairs between an affirmative-action and a non-affirmative action student.

4.4 Robustness tests using alternative empirical strategies

Columns (6) to (8) in Table 4 show similar results using alternative IV strategies as explained in section 3.1. In column (6), the instrument is the alphabetical distance between last names (based on the alphabetical ordering of all students' last names, and censored at 16), a predictor of membership in the same IG, which itself predicts friendship. This instrument addresses the concern that students may choose to not participate in their assigned IG, or selectively avoid the integration week. While the numerical nature of the instrument is largely different, the IV estimate remains highly similar to the benchmark.

In columns (7) and (8), we use the standard instrument, namely membership in the same IG, but restrict the sample to pairs whose last names are alphabetical neighbors, thus likely to have very similar observed and unobserved characteristics. This strategy addresses the concern that certain patterns of last name ordering may correlate with common unobserved characteristics that can influence differences in political opinions. Again, the IV estimate remains quite stable in those columns, and statistically significant at 5%, even when the sample has shrunk by about 15 times.[39]

4.5 Retrospective answers and validity of estimates

We use a retrospective question in the survey in March 2014 on students' political opinions just before they join Sciences Po (see the timing in section 2), which raises a potential concern that retrospective answers may incorporate a bias in the direction of the respondent's opinion today. While such a measurement error regarding retrospective survey questions on events and answers may be rather small after only 6 months,[40] the bias on opinions may also relate to the rationalization of new information that results in a hindsight bias, according to which individuals reconstruct their past opinion in light of their newly updated opinion (Fischhoff and Beyth, 1975). It is thus useful to investigate our method's robustness to this issue.

To evaluate the magnitude of the retrospective answer measurement error, we use the second survey in June 2015 to compare the answers to its retrospective question on recalled opinion back in March 2014 with the actual answers in 2014. First, Appendix Table A5 shows the joint distribution of both surveyed and recalled opinions for 2014. The mass is clearly concentrated on the diagonal, with 90% of the observations not differing more than 1 point between the two measures, implying a very strong correspondence between recalled and actual answers. This lends confidence to the accuracy of the recalled opinion expressed in March 2014 over the political opinion in August

[39] We further run robustness checks by dropping all last names starting with each letter, dropping all students with a specific non-French nationality, or dropping all French family names starting with "de", which might correspond to an aristocratic family background. The results shown in Appendix Tables A10 and A11 are strongly robust to all those concerns.

[40] Wagenaar (1986) finds that 20% of subjects forget key personal events after one year. See review by Bradburn et al. (1987).

2013.[41]

Appendix Table A6 presents further results on students' recall error, measured as recalled opinion for 2014 minus actual opinion surveyed in 2014. The absolute magnitude of the recall error has practically zero partial correlations with past and present actual political opinions, as shown in column (1). However, in column (2) we do find evidence that the signed recall error is strongly correlated with the change in opinions from 2014 to 2015, signifying that recalled opinions are biased towards present opinions (as surveyed in 2015) by the same magnitude as estimated, e.g., by Fischhoff and Beyth (1975); Biais and Weber (2009); Camerer et al. (1989).

How much can the recall error affect our results? First, if less than 10% of answers suffer a recall error, the resulting bias on our benchmark result would probably be minimal. Second, since we mostly control for pre-Sciences Po political opinions, if this variable is biased towards actual opinions of March 2014, it would create an attenuation bias of our coefficient of interest towards zero. Indeed, the recalled opinion with error will tend to absorb more variation in the outcome variable than the actual initial opinion, and by doing so it reduces the coefficient of the other variables. Third, the control variables are not needed for the IV strategy's validity, and only included to improve estimates' precision. Indeed, the results remain very similar, albeit less precise, if we do not control for pre-Sciences Po political opinions.

4.6 Friendship effect on association activities

It is important to link students' beliefs to their behaviors. In this context, the most natural behavior following changes in political opinions is participation in political organizations, including students' associations with certain political inclinations, and political parties. Table 5 shows results on the effect of friendship on the indicator whether a pair of students enroll in the same organization.[42]

In Panel A, while we do not see a significant effect of friendship in the indicator of joining any association (column (1)), students do follow their friends in joining exactly the same association. The effect is about 0.18 among students who do join some organization (column (2)), and 0.08 in the full sample (column (3)). On the other hand, columns (4) and (5) show no significant effect on enrollment in political parties. Indeed, almost roughly half of the small number of students who are registered with a political party is explained by their pre-Sciences Po registration (column (4)), leaving little variation of this outcome that can be explained by friendship.

Table 5 Panel B focuses on indicators of students joining associations of a certain type, in order to understand friends' motivation in joining the same association. The effect of friendship is

[41]Unfortunately, due to budget constraints in 2015, the participation rate in 2015 is much lower than in 2014, resulting in a small sample that overlaps between the two waves that we cannot use as a panel to study friendship effect.

[42]Unfortunately, we do not observe the intensity of participation, and could only consider the extensive margin. Another shortcoming is that formal political party enrollment is still very rare among first-year students. Also, most of them have not reached voting age in previous political elections.

Table 5: Friendship Effect on Political and Associative Activities

Panel A: Friendship Effect on Participation in Political Parties and Associations

Dependent Variables:	Membership in Some Associations	Conditional Membership in Same Association*	Membership in Same Association	Membership in Some Political Party	Conditional Membership in Same Political Party*
	(1)	(2)	(3)	(4)	(5)
Friendship	-0.041	0.178*	0.076*	0.035	-0.015
	(0.1157)	(0.0931)	(0.0423)	(0.0651)	(0.3197)
Both Pre-Sciences Po Party Membership	-	-	-	0.469***	-
				(0.0735)	
Observations	47,895	24,310	54,615	52,650	780
IV			Same Integration Group		
Controls	Yes	Yes	Yes	Yes	Yes
Double Group Clust.	Yes	Yes	Yes	Yes	Yes
WeakIV test stat.	110.7	69.8	113.2	104.9	5.714

* Conditional on both individuals being enrolled in some association/political party.
Notes: The dyadic dummy variables are equal to 1 if the characteristic is the same for the two individuals in the pair and 0 if the characteristic is not the same. Standard errors are two-way clustered by individual 1's group and by individual 2's group. The benchmark sample is detailed in Appendix Table A1. Weak IV stat reports the Kleibergen-Paap cluster-robust statistic, distributed as a Chi-squared under the null hypothesis of weak identification.

Panel B: Friendship Effect on Different Association Types

	Membership in						
Dependent Variable:	Same Association					Associations of Same Type	Different Associations of Same Type
Type of Association:	Strictly Political or Student Union	Political	Related to One's Identity	Related to Policy Issues	Sports	Political	Political
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Friendship	0.0488*	0.133*	0.138	0.00437	0.0418	0.00371	-0.0111
	(0.0272)	(0.0715)	(0.0899)	(0.0808)	(0.1720)	(0.0170)	(0.0210)
Observations	24,310	24,310	24,310	24,310	24,310	24,310	24,310
IV				Same Integration Group			
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Double Group Clust.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
WeakIV test stat.	69.84	69.84	69.84	69.84	69.84	69.84	69.84

Notes: The dependent dummy variables are equal to 0 if the characteristic is the same for the two individuals in the pair and 1 if the characteristic is not the same. Standard errors are two-way clustered by individual 1's group and by individual 2's group. The benchmark sample is detailed in Appendix Table A1. Weak IV stat reports the Kleibergen-Paap cluster-robust statistic, distributed as a Chi-squared under the null hypothesis of weak identification.

concentrated on associations that have a political inclination and those related to one's identity. Column (2) groups associations that are strongly linked to political parties and/or motivated by regular political activities.[43] Column (3) considers all associations with some political motivation, including the ones considered in column (2), as well as others with focus on a political issue, but without a particular political leaning. The effect is statistically significant at 10% in both columns (2) and (3), that is, a friendship link significantly increases the possibility that both students join a politically-leaning association. The difference in the effect size between the two columns may be due in part to the fact that there are considerably fewer associations considered in column (2). Column (4) gathers all associations pertaining to certain aspect of a student's identity, such as nationality, region of origin, religion, and ethnicity. The estimate is strongly positive, but not significant, which could reflect great heterogeneity in the activities and the appeal of those associations. Other types of associations such as related to sports (column (1)) or policy issues (column (5)) do not exhibit a friendship effect.

The similar participation in political associations among friends may be further decomposed into two mechanisms: (i) friends influence each other's intensity of interest in politics, which brings both to politically-leaning associations, and (ii) friends prefer to have more interactions in (the same) political associations. In columns (6) and (7) we attempt to explore the first mechanism by considering the effect of friendship on both students' joining some political associations (column (6)), and especially on their joining some *different* political associations (column (7)). Both estimates are close to zero, suggesting that friendship does not drive students' interest in politics, and students are no more interested in any political associations other than their friends'. Friendship thus has a very targeted effect on participation in associations.

Taken together, these results are important in showing an effect of friendship on actual behaviors, beyond the one on self-reported political opinion. Interestingly, while friendship leads friends to further exposure within an association, in this context it only does so for politically-motivated associations.

5 Main drivers of the friendship effect

5.1 Direction of opinion changes

In this section, we investigate the mechanisms that drive our main result, first by considering the different types of dyadic changes in opinions from before students' entrance into Sciences Po (August 2013) to the moment of the survey (March 2014). First, we classify the changes in a pair's opinions into three major categories: Opinions that move towards each other (convergence),

[43] They include student unions, notably the famous UNEF, known for its rather radical agenda and organization of most of the students' loudest manifestations on campus, even though it is officially declared as apolitical. While a student at Sciences Po, the former French president François Hollande was also president of UNEF.

Table 6: Friendship Effect on Convergence, Divergence, and Co-movement

Dependent Variables:	Staying in Same Position	Weak Convergence	Single Convergence	Strong Convergence	Weak Divergence	Single Divergence	Strong Divergence	Moving in Same Direction
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Friendship	0.0793	0.0341	0.0802	-0.0466	-0.164*	-0.0520	-0.112***	0.100*
	(0.0676)	(0.0957)	(0.0970)	(0.0848)	(0.0843)	(0.0760)	(0.0314)	(0.0540)
Initial Diff. in Political Opinion (August 2013)	-0.00528	0.110***	0.0598***	0.0598***	-0.0833***	-0.0646***	-0.0187***	-0.0277***
	(0.00378)	(0.00672)	(0.00323)	(0.00656)	(0.00256)	(0.00288)	(0.00200)	(0.00600)
Mean of Dependent Variable	0.156	0.456	0.289	0.15	0.227	0.188	0.038	0.183
Standard Deviation of Dependent Variable	0.3628	0.4981	0.4535	0.3571	0.4186	0.391	0.1919	0.3868
Observations	54,615	46,901	49,636	46,901	54,615	54,615	54,615	54,615
IV				Same Integration Group				
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Double Group Clust.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
WeakIV test stat.	113.2	112.7	107.3	112.7	113.2	113.2	113.2	113.2

Notes: Standard errors are two-way clustered by individual 1's group and by individual 2's group. The benchmark sample is detailed in Appendix Table A1. Weak IV stat reports the Kleibergen-Paap cluster-robust statistic, distributed as a Chi-squared under the null hypothesis of weak identification.

opinions that move away from each other (divergence), and opinions that move in the same direction (co-movement). In each category, we further define the possibilities of a "strong" change in which both opinions move, a "single" change in which only one opinion changes while the other remains fixed, and a "weak" case that groups the strong change, the single change, and the possibility that both remain fixed.[44]The categories are not mutually exclusive, e.g., the case of one fixed opinion and the other moving away would be branded as both divergence and co-movement. Table 6 shows estimates of the effect of friendship on the incidence of each category, using the IV strategy set out in section 3.1.

The estimates of the effect of friendship on the incidences of fixed opinions, weak convergence, single convergence, and strong convergence (columns (1) to (4)) are not statistically significant, and of rather modest magnitude. On the other hand, columns (5) to (7) show strong evidence that friendship reduces the incidences of weak and strong divergence. Friendship reduces the possibility of divergence by 12 to 17 percentage points, a large magnitude compared with the mean probability of divergence. We also find evidence of friendship inducing students to co-move in the same direction, and doing so while reducing their differences (columns (8) and (9)).[45]

We draw two major conclusions from this subsection. First, the lack of result on convergence shows that friendship does not necessarily create echo chambers or filter bubbles, in terms of compressing the diversity of opinions. Yet the strong effect of discouragement on divergence suggests

[44]Denoting ΔY_i as i's signed change in opinion from $t = 0$ (before Sciences Po) to $t = 1$ (survey time), those concepts are formally defined as follows. The pair (i, j) maintain fixed opinions iff $\Delta Y_i = \Delta Y_j = 0$. They experience a strong convergence iff $\Delta Y_i(Y_{j0} - Y_{i0}) > 0$ & $\Delta Y_j(Y_{i0} - Y_{j0}) > 0$, a single convergence iff ($\Delta Y_j = 0$ & $\Delta Y_i(Y_{j0} - Y_{i0}) > 0$) or ($\Delta Y_i = 0$ & $\Delta Y_j(Y_{i0} - Y_{j0}) > 0$), and a weak convergence iff $\Delta Y_i(Y_{j0} - Y_{i0}) \geq 0$ & $\Delta Y_j(Y_{i0} - Y_{j0}) \geq 0$. Similarly, strong divergence means $\Delta Y_i(Y_{j0} - Y_{i0}) < 0$ & $\Delta Y_j(Y_{i0} - Y_{j0}) < 0$, single divergence ($\Delta Y_j = 0$ & $\Delta Y_i(Y_{j0} - Y_{i0}) < 0$) or ($\Delta Y_i = 0$ & $\Delta Y_j(Y_{i0} - Y_{j0}) < 0$), and weak divergence $\Delta Y_i(Y_{j0} - Y_{i0}) \leq 0$ & $\Delta Y_j(Y_{i0} - Y_{j0}) \leq 0$. Co-movement is defined as $\Delta Y_i \Delta Y_j \geq 0$.

[45]Table 6's approach focuses on the extensive margin of each category of movement, thus alleviates the concern of recall bias due to the retrospective nature of students' answers on political opinion before joining Sciences Po. Indeed, even if those answers can be biased towards students' opinions, this bias does not affect the results on the weak categories, such as friendship's reduction effect on divergence. We will return to this issue in section 4.4.

that friends do keep each other from deviating away to the extremes.

Second, it is important to consider the nonlinearity of the effect of friendship on each other's opinion, as the effect can be dependent on the direction of opinion change. This finding questions the typical assumption of homogenous, linear effects of direct links on one's beliefs, as modeled and estimated in the theoretical and empirical literatures on non-Bayesian learning in networks (Möbius and Rosenblat, 2014). Examples include theories using average-based belief updating processes (the term coined by Golub and Jackson 2012, for a generalized definition of DeGroot 1974's belief updating), or other types of updating (Campbell et al. 2019) or empirical estimations of peer effect in networks such as Calvó-Armengol et al. (2009), or Bramoullé et al. (2009).

5.2 Friendship effect among similar students

While our estimate of the friendship effect is free of the homophily bias, the homophily property itself is useful to understand the mechanism behind the friendship effect, one that we term the "homophily-enforced channel". To illustrate the role of homophily, consider two pairs of students: the first pair, François (F) and Ségolène (S), started Sciences Po with very similar characteristics on the key dimension in our analysis, namely political opinions, whereas the second pair, Michel (M) and Dominique (D), started Sciences Po with very different political views. Given homophily, the unconditional probability for F and S to become friends is higher than that of M and D. However, conditional on both pairs becoming friends, homophily implies that it is likely that M and D have other, non-political characteristics in common, e.g., their interest in tennis. Thence, throughout their time at Sciences Po, each pair's friendship conducts to more interaction on the dimension that helps them become friends, thus reinforces the corresponding similarity, namely, politics between F and S, and tennis between M and D.

Friendship thus has very different consequences on friends' similarity depending on the friends' initial dimension of homophily. Friendship should have a strong effect on the political similarity among pairs that started out with similar political views, but a weak or inexistent effect on political views of pairs that started out with dissimilar political views. This implication is tested on different subsamples based on initial differences in political opinions, and reported in Table 7.

Columns (2) to (6) report the benchmark IV regression coefficient of friendship on subsamples that condition pre-Sciences Po differences in political opinions to be within the indicated ranges, namely equal to 0, 1, 2, and greater and less than 2 (recall that the average of dyadic differences is about 1.9). The friendship effect on friends' difference in political opinion is mostly powerful among pairs with an initial difference in political views of less than 2: the coefficient in column (6) in Table 7 is about 50% larger than the benchmark coefficient in Table 4. As initial differences increase beyond 2, the effect fades out with smaller and statistically insignificant coefficients. This pattern is confirmed in column (1)'s full-sample specification that includes a saturated model in

Table 7: Friendship Effect and Initial Differences in Political Opinion

Dependent Variables:	Difference in Political Opinion (March 2014)					
Conditions		Difference in Initial Political Opinion: =0	=1	=2	> 2	< 2
	(1)	(2)	(3)	(4)	(5)	(6)
Friendship	-0.978**	-0.792***	-0.690**	-0.404	-0.265	-0.732***
	(0.422)	(0.302)	(0.292)	(0.471)	(0.407)	(0.231)
Friendship*Initial Diff. in Political Opinion	0.205	-	-	-	-	-
	(0.174)					
Initial Diff. in Political Opinion (August 2013)	0.522***	-	-	-	0.715***	0.140***
	(0.0252)				(0.0416)	(0.0180)
Observations	54,615	7,714	14,235	11,644	21,022	21,949
IV	Same Integration Group					
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Double Group Clust.	Yes	Yes	Yes	Yes	Yes	Yes
WeakIV test stat.	30.97	46.36	41.59	18.14	68.69	59.77

Notes: Standard errors are two-way clustered by individual 1's group and by individual 2's group. The benchmark sample is detailed in Table A1. Weak IV stat reports the Kleibergen-Paap cluster-robust statistic, distributed as a Chi-squared under the null hypothesis of weak identification.

friendship and initial differences in political opinions, using both the same-IG variable IG_{ij} as well as its interaction with initial differences as instruments. The coefficient on the interaction between friendship and initial differences is large and positive, but noisily estimated.[46]

As those results highlight the role of pairs with a small difference in pre-Sciences Po political opinions, we further explore how these pairs reinforce their relationship if they become friends. The second commonest type of environment that fosters interactions among Sciences Po students, as explained in section 2, is association activities. We thus revisit the analysis by association types in Table 5 Panel B, now restricted to the subsample of pairs with a pre-Sciences Po difference in political opinion less than 2 (around the mean of political opinion gap).[47] The results in columns (2) to (4) of Table 8 show a very strong effect of friendship on a pair's participation in the same type of associations, but only among those related to politics and personal identity (such as based on national and regional origins). Other types of associations, be it sports or practical policy issues, enjoy no such concentration of friends (columns (1) and (5)). Similar to Table 5 Panel B, friendship only increases a pair's participation in the same association, but not their interest in participating in different associations of the same type (columns (6) and (7)).

Taken together, the findings from Tables 7 and 8 are consistent with the homophily-enforced channel. By homophily, students who form new friendships because of their chance encounter (i.e., in an IG, in this paper's context) may do so according to one or a few areas in which they share common interest. One important area can be politics: certain pairs of students form friendships because of their similarity in political opinions. Others, however, may form friendships along other

[46]Its lack of statistical significance can also reflect the nonlinear nature of the relationship, as well as the large amount of noise among pairs with large initial differences.

[47]For completeness, the analysis similar to Table 5 Panel A on this subsample, as well as all of Table 5's analysis on the complementary subsample of all pairs with pre-Sciences Po difference in political opinion of at least 2, are presented in Appendix Table A8.

Table 8: Convergence on Associative Activities - Conditional on Similar Initial Opinion

	d. Membership in						
Dependent Variable:	Same Association					Associations of Same Type	Different Associations of Same Type
Type of Association:	Strictly Political or Student Union	Political	Related to One's Identity	Related to Policy Issues	Sports	Political	Political
Condition:	Difference in Initial Political Opinion < 2						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Friendship	0.128**	0.244**	0.426**	0.180	0.041	0.0406	-0.0357***
	(0.0567)	(0.103)	(0.180)	(0.169)	(0.231)	(0.0375)	(0.0101)
Observations	9,727	9,727	9,727	9,727	9,727	9,727	9,727
IV	Same Integration Group						
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Double Group Clust.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
WeakIV test stat.	36.26	36.26	36.26	36.26	36.26	36.26	36.26

Notes: The dependent dummy variables are equal to 0 if the characteristic is the same for the two individuals in the pair and 1 if the characteristic is not the same. Standard errors are two-way clustered by individual 1's group and by individual 2's group. The benchmark sample is detailed in Appendix Table A1. Weak IV stat reports the Kleibergen-Paap cluster-robust statistic, distributed as a Chi-squared under the null hypothesis of weak identification.

dimensions, and can become friends despite large differences in political opinion. Over time, interactions between pairs are shaped along the lines of common interests, so friends with similar political opinions tend to often discuss politics and join the same politically-inclined associations, while friends with different political opinions may strengthen their relationship through other dimensions on which they are more similar. In consequence, friends who started out with similar political opinions continue to influence each other's political opinions, while friends who started out with large differences in opinions do not exert much influence on each other's political opinions.

This mechanism echoes Golub and Jackson's (2012) analysis on homophily and the speed of convergence in beliefs, but with the introduction of an endogenous selection of the dimension of interaction based on homophilous preferences. Since our newly discovered empirical facts imply a rather nonlinear mechanism of diffusion of beliefs, notably in the asymmetry between converging and diverging, it would be interesting to reconsider their results in light of those facts.

5.3 Polarization and extremism among similar students

We further investigate the homophily-enforced mechanism of friendship on the directions of opinion changes. Table 9 repeats the regressions in Table 8, with the outcome variable indicating whether the pair diverges strongly.[48] Friendship between of pair of students with similar initial political opinions has a very strong and statistically significant effect in reducing the possibility of both diverging in opposite directions. The overall effect of friendship on divergence seems to be dominated by similar pairs, while the pairs with large differences in pre-Sciences Po political opinions remain as likely to diverge even if they become friends.

One natural corollary of the friendship effect on divergence is that friendship reduces the inci-

[48] That is, the changes in opinions $\Delta Y_i, \Delta Y_j$ satisfy $\Delta Y_i(Y_{j0} - Y_{i0}) < 0$ & $\Delta Y_j(Y_{i0} - Y_{j0}) < 0$. Similar cases in which the outcome variable is the indicator of single or weak divergence, or strong, single, or weak convergence, or comovements, are presented in Appendix Table A9.

Table 9: Strong Divergence and Initial Differences in Political Opinion

Dependent Variables:	Strong Divergence in Political Opinion					
Conditions		Difference in Initial Political Opinion: =0	=1	=2	> 2	< 2
	(1)	(2)	(3)	(4)	(5)	(6)
Friendship	-0.248***	-0.422***	-0.0737	-0.0790	-0.0214	-0.193***
	(0.0480)	(0.115)	(0.0455)	(0.0555)	(0.0341)	(0.0462)
Friendship*Initial Diff.	0.0631***	-	-	-	-	-
in Political Opinion	(0.0118)					
Initial Diff. in Political	-0.0197***	-	-	-	-0.00516***	-0.0808***
Opinion (August 2013)	(0.00210)				(0.00136)	(0.00716)
Observations	54,615	7,714	14,235	11,644	21,022	21,949
IV	Same Integration Group					
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Double Group Clust.	Yes	Yes	Yes	Yes	Yes	Yes
WeakIV test stat.	30.97	46.36	41.59	18.14	68.69	59.77

Notes: Standard errors are two-way clustered by individual 1's group and by individual 2's group. The benchmark sample is detailed in Appendix Table A1. Weak IV stat reports the Kleibergen-Paap cluster-robust statistic, distributed as a Chi-squared under the null hypothesis of weak identification.

Table 10: Friendship Effect on Extremism

Dependent Variables:	At Least One Extremist in the Couple in March 2014			
	Initial Difference in Initial Political Opinion < 2			
Initial Conditions:	All	Same Gender	Same Admission Type	Both Moderate
	(1)	(2)	(3)	(4)
Friendship	-0.146**	-0.204**	-0.193**	-0.123*
	(0.0658)	(0.102)	(0.0904)	(0.0652)
Observations	21,949	11,171	14,929	20,324
IV	Same Integration Group			
Controls	Yes	Yes	Yes	Yes
Double Group Clust.	Yes	Yes	Yes	Yes
WeakIV test stat.	59.74	42.63	73.30	62.51

Notes: Standard errors are two-way clustered by individual 1's group and by individual 2's group. The benchmark sample is detailed in Appendix Table A1. Weak IV stat reports the Kleibergen-Paap cluster-robust statistic, distributed as a Chi-squared under the null hypothesis of weak identification.

dence of students with extremist views. Table 10 tests this prediction by applying our IV strategy to study the effect of friendship on the indicator of at least one student with an extreme political opinion, defined as an opinion score either in $\{1, 2\}$ (extreme left) or $\{9, 10\}$ (extreme right).[49] As in the previous table, we further condition the sample to those whose pre-Sciences Po political opinions are rather similar. Column (1) shows that a friendship link reduces the incidence of extremism after 6 months at Sciences Po by 15 percentage points, an estimate statistically significant at 5%. When we condition the sample on other dimensions of homophily that can be more easily measured, including same gender (column (2)), same type of admission procedure (which mostly separates out the affirmative action students from the rest, shown in column (3)), and both being moderate (defined as an opinion score in $\{3, 8\}$, shown in column (4)), the results remain important and statistically significant throughout all subsamples.

Given the estimated friendship effect and the frequency of friendship of 0.017, friendship among similar students has thus contributed 9% of the reduction of 2.7 percentage points, from 19.5% to

[49] As usually found in French universities, there exist few extreme right views. The number of extreme left views is much larger, as can be seen in Figure 1.

16.8%, in the proportion of pairs with at least one extreme view.[50]

6 Friendship effect and network characteristics

After having focused on the channels of the friendship effect in the previous section, we now turn to its heterogeneous magnitudes based on the network structure. Subsection 5.1 has already shown different friendship effects on divergence versus convergence, which disputes the standard assumption of linear homogenous friendship effects across all types of links in DeGroot-like averaged-based belief updating processes (Golub and Jackson, 2012) typically used to model non-Bayesian learning (Möbius and Rosenblat, 2014). In this section we examine this assumption across network characteristics.

We first consider the intensity of friendship, surveyed on a scale of 1 (mere friendship) to 4 (very close friends). To highlight the difference between each level x and its lower level $x-1$ ($x \in \{1, \ldots, 4\}$, with level 0 meaning the pair are not friends), we create a binary variable D^x equal to 1 if the friendship intensity is x or stronger, and 0 otherwise (thus any variation of intensity above x or below $x-1$ is collapsed to a single level), and use our IV strategy with D^x_{ij} as the treatment variable instead of friendship $Link_{ij}$.[51]

Columns (2) to (5) of Table 11 report the results of those specifications, for different levels of intensity from $x = 4$ down to $x = 1$. At $x = 1$, we are back to the benchmark specification in Table 4, in which all intensity levels are considered the same. The friendship effect is not homogenous when one passes from one level of intensity to the next, but ranges from -0.558 for the lowest level of friendship to up to -3.101 for very close friends (all estimates are statistically significant at 5%, although estimates at higher intensity levels are less precise). This effect could be inherently due to the strength of friendship, or to the possibility that friends with similar political opinions tend to be close friends, and based on Table 8 they tend to experience a stronger friendship effect.

In Table 12, we further differentiate pairwise relationships in categories based on social distance, i.e., the shortest-path distance between any pair of students in their social network. Column (1) reproduces the benchmark result from Table 4's column (4) to highlight the causal comparison between direct friends (social distance 1) and non-direct friends (social distance greater than 1). In column (2), we consider the effect of increasing a pair's social distance on their difference in political opinion. Using IG membership as instrument for social distance, we can interpret the estimated effect as an Average Causal Response, a weighted average over causal effects among all pairs that

[50] As in the rest of our analysis, this simple calculation is based on the dyadic sample of student pairs, not the sample of individuals. Unfortunately, there is no simple way to directly map dyadic results to a monadic interpretation.

[51] That is, we use the same IG membership IG_{ij} as instrument for each new treatment variable D^x_{ij} in a separate specification, to estimate the LATE of improving one degree of friendship intensity at each level. Similarly to the arguments in section 3.1, in this context the instrument still satisfies all conditions for a causal LATE interpretation. The only potential concern is whether the variable IG_{ij} is also a strong instrument for D^x_{ij} for any level of friendship intensity x, a condition that Table 11 duly tests and confirms.

Table 11: Friendship Effect by Friendship Intensity

Dependent Variable:	Difference in Political Opinion (March 2014)				
Grouped Intensity Levels:	I	-	-	-	-
	(1)	(2)	(3)	(4)	(5)
Friendship	-0.549**	-	-	-	-
	(0.224)				
Very Close Friendship	-	-3.101**	-	-	-
		(1.312)			
Close Friendship	-	-	-1.324**	-	-
or Stronger			(0.563)		
Friendship Link	-	-	-	-0.717**	-
or Stronger				(0.296)	
Mere Relationship	-	-	-	-	-0.558**
or Stronger					(0.238)
Initial Diff. in Political	0.526***	0.525***	0.525***	0.525***	0.526***
Opinion (August 2013)	(0.0253)	(0.0253)	(0.0253)	(0.0253)	(0.0253)
Observations	54,556	54,615	54,615	54,615	54,615
IV	Same Integration Group				
Controls	Yes	Yes	Yes	Yes	Yes
Double Group Clust.	Yes	Yes	Yes	Yes	Yes
WeakIV test stat.	116.1	29.83	54.89	75.33	102.9

I - Only friends that did not know each other from before entering Sciences Po.
Notes: Standard errors are two-way clustered by individual 1's group and by individual 2's group. The benchmark sample is detailed in Appendix Table A1. Weak IV stat reports the Kleibergen-Paap cluster-robust statistic, distributed as a Chi-squared under the null hypothesis of weak identification.

comply to the IV to move closer in social distance (Angrist and Imbens, 1995; Angrist and Pischke, 2008). The magnitude of the effect in column (2) is a lot smaller than the corresponding benchmark coefficient, suggesting that the effect is not always strong for all values of social distance (now a positive coefficient means a shorter social distance leads to closer opinions.)

In this direction, we cut the sample into subsamples to compare between pairs of consecutive social distances: social distance 1 versus 2 in column (3), social distance 2 versus 3 in column (4), and social distance 3 versus farther distances in column (5). In each column, we can always use the same IG membership as instrument, subject to a test of instrument strength. We find a strong and precisely estimation effect on political opinion when social distance shrinks from 2 to 1, i.e., when indirect friends become direct friends. The magnitude of the convergence effect of this treatment is not too far from the benchmark result: 0.38 versus 0.54 points on a scale from 1 to 10. Beyond direct friends, we find some evidence of a similar-sized effect of reducing social distance from 3 to 2, although the estimate remains rather imprecise, but not any statistically significant result above social distance 3.[52]

Similar to section 4.2, a simple calculation shows how much of the total change in pairwise differences in political opinions can be attributed to direct friendships and second-degree friendships (friends of friends). Using estimates from columns (3) and (4) in Table 12, and the frequencies of direct and second-degree friendships at 0.0170 and 0.0990, the explained proportion is $\frac{(0.375+0.436)\times 0.0170+0.436\times 0.0990}{2.2-1.926} = 20.8\%$. This high figure should be taken with caution, however, given the low precision of the estimate 0.436.

[52]This is similar to Leider et al.'s (2009) description that from a social preference perspective, indirect friends of distance 3 are not distinguishable from strangers.

Table 12: Effect of Reduction of Social Distance on Difference in Political Opinion

Dependent Variable:	Difference in Political Opinion				
Condition:	-	-	Degree< 3	Degree $\in \{2, 3\}$	Degree > 2
	(1)	(2)	(3)	(4)	(5)
Friendship	-0.538**	-	-	-	-
	(0.228)				
Shortest Path	-	0.151**	-	-	-
		(0.0624)			
1st vs 2nd order only	-	-	0.375***	-	-
			(0.120)		
2nd vs 3rd order only	-	-	-	0.436*	-
				(0.240)	
3rd vs more order only	-	-	-	-	3.945
					(6.916)
Observations	54,615	54,615	6,369	27,359	48,246
IV	Same Integration Group				
Controls	Yes	Yes	Yes	Yes	Yes
Double Group Clust.	Yes	Yes	Yes	Yes	Yes
WeakIV test stat.	113.2	151.3	212	171.3	0.357
A.R. Wald test F-stat.	5.96	5.96	9.00	3.41	1.33

Notes: specification (1) - benchmark sample; specification (2) - benchmark sample; specification (3) - only pairs of individuals that are at most within two degrees of distance; specification (4) - only pairs of individuals that are between two and three degrees of distance; specification (5) - only pairs of individuals that are at three or more degrees of distance. Standard errors are two-way clustered by individual 1's group and by individual 2's group. The benchmark sample is detailed in Appendix Table A1. Weak IV stat reports the Kleibergen-Paap cluster-robust statistic, distributed as a Chi-squared under the null hypothesis of weak identification.

Next, we study whether students' positions in the social network, notably in terms of *network centrality*, may matter to the friendship effect on differences in political opinions. This question is related to the line of thought that centrality in social networks also conveys information about a person's influence and charisma. If that is the case, the friendship effect should be much larger for pairs of a highly-central student (a "star") and a non-central student (a "non-star") than pairs of two non-central students or pairs of two highly-central students. Table 13 explores this idea using the measure of *eigenvector centrality* (see Table 1, Panel A), according to which a student's centrality measure is a linear aggregate of his friends' centrality. The three columns consider respectively the subsamples of pairs of two stars (column (1)), of one star and one non-star (column (2)), and of two non-stars (column (3)), for which a student is considered highly-central (a star) if his centrality measure is in the highest quartile of its distribution. The friendship effect is not present among pairs of stars, with a positive, statistically insignificant estimate. Among the other two subsamples, the effect is rather strong and statistically significant at 10%.[53]Overall, the evidence is consistent with the view that network stars are hardly influenceable, whereas non-stars tend to be more influenceable. However, network stars do not necessarily wield stronger influence than non-stars.

7 Concluding remarks

In this paper, we investigate empirically how newly-formed friendships through a one-week exposure among new students at the entry of college may shape friends' political beliefs, in the context of

[53]Those results are explored and confirmed in details in Appendix Table A12.

Table 13: Friendship Effect and Network Centrality

Dependent Variable:	Difference in Political Opinion (March 2014)		
Condition:	Two Central	One Central and One Non-Central	Two Non-Central
	(1)	(2)	(3)
Friendship	0.141	-0.834*	-0.764*
	(0.229)	(0.467)	(0.412)
Observations	3,240	20,250	31,125
IV	Same Integration Group		
Controls	Yes	Yes	Yes
Double Group Clust.	Yes	Yes	Yes
WeakIV test stat.	161	58.83	57.02

Notes: specification (1) - two individuals in the top quartile of eigenvector centrality; specification (2) - one individual in the top quartile of eigenvector centrality; specification (3) - no individual in the top quartile of eigenvector centrality. Standard errors are two-way clustered by individual 1's group and by individual 2's group. The benchmark sample is detailed in Appendix Table A1. Weak IV stat reports the Kleibergen-Paap cluster-robust statistic, distributed as a Chi-squared under the null hypothesis of weak identification.

students at a French elite school well-known as the most prominent alma mater of French politicians. We find that within the first 6 months of college, friendship causes a substantial reduction in the gap between friends' political opinions by around 0.5 points (over a scale from 1 to 10 in political opinions), equivalent to a quarter of the average gap and a third of its standard deviation. Friendship also causes friends' participation in the same politically-inclined associations. The effect of friendship works through a discouragement of divergence and an encouragement of co-movement of opinions, but not an encouragement of convergence. Consequently, friendship tends to reduce extremist views among students, without forcing the same views among friends.

The results are consistent with what we term as a "homophily-enforced" mechanism of friendship effect. When students who are similar on the dimension of political opinions become friends, they continue to interact on related topics, as shown by their decisions to participate in the same politically-inclined associations (and not other types of associations). Such continual interactions, as enforced by homophily, are the factor that produces the strong effect of friendship among students with high similarity in political opinions before joining Sciences Po. In consequence, those pairs of students are strongly discouraged from diverging, and, therefore, are less likely to hold extreme political views. In contrast, friends who were politically dissimilar before Sciences Po would not follow this path, hence friendship has little effect on their political opinions.

Another important aspect of our results is the nonlinearity and heterogeneity of the friendship effect on political opinions. It is stronger for stronger friendship and closer social distance. It is likely to be asymmetric between network stars and non-stars, defined by network eigenvector centrality, in that stars are much less likely to be influenced. Those findings will be important in shaping how we model the process of belief formation and transmission in networks.

References

Acemoglu, Daron, Munther Dahleh, Ilan Lobel, and Asuman Ozdaglar, "Bayesian learning in social networks," *The Review of Economic Studies*, 2011, *78* (4), 1201–1236. 5

Advani, Arun and Bansi Malde, "Credibly identifying social effects: Accounting for network formation and measurement error," *Journal of Economic Surveys*, 2018, *32* (4), 1016–1044. 3

Alesina, Alberto, Paola Giuliano, and Nathan Nunn, "On the Origins of Gender Roles: Women and the Plough," *The Quarterly Journal of Economics*, 2013, *128* (2), 469–530. 5

Allcott, Hunt and Matthew Gentzkow, "Social Media and Fake News in the 2016 Election," *The Journal of Economic Perspectives*, 2017, *31* (2), 211–235. 2

Altonji, Joseph, Todd Elder, and Christopher Taber, "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 2005, *113* (1), 151–184. 8

Angrist, Joshua and Guido Imbens, "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity," *Journal of the American Statistical Association*, 1995, *90*, 430–442. 31

_ and Jörn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton, NJ: Princeton University Press, December 2008. 31

Angrist, Joshua D. and Kevin Lang, "Does school integration generate peer effects? Evidence from Boston's Metco program," *The American Economic Review*, 2004, *94* (5), 1613–1634. 4

Badev, Anton, "Discrete games in endogenous networks: Theory and policy," *Unpublished, University of Pensylvania*, 2018. 3

Bala, Venkatesh and Sanjeev Goyal, "Learning from Neighbours," *The Review of Economic Studies*, 1998, *65* (3), 595–621. 5

_ and _ , "Conformism and Diversity under Social Learning," *Economic Theory*, 2001, *17* (1), 101–120. 5

Biais, Bruno and Martin Weber, "Hindsight Bias, Risk Perception, and Investment Performance," *Management Science*, 2009, *55* (6), 1018–1029. 22

Bifulco, Robert, Jason M. Fletcher, and Stephen L. Ross, "The Effect of Classmate Characteristics on Post-secondary Outcomes: Evidence from the Add Health," *American Economic Journal: Economic Policy*, February 2011, *3* (1), 25–53. 3

Bisin, Alberto and Thierry Verdier, ""Beyond the Melting Pot": Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits," *The Quarterly Journal of Economics*, 2000, *115* (3), 955–988. 5

_ and _ , "The Economics of Cultural Transmission and the Dynamics of Preferences," *Journal of Economic Theory*, 2001, *97* (2), 298–319. 5

_ and _ , "The Economics of Cultural Transmission and Socialization," in Jesse Benhabib, Alberto

Bisin, and Matthew O. Jackson, eds., *Handbook of Social Economics*, Vol. 1, Amsterdam: North Holland, 2011, pp. 339–416. 5

Blume, Lawrence E., William A. Brock, Steven N. Durlauf, and Yannis M. Ioannides, "Identification of social interactions," in Jesse Benhabib, Alberto Bisin, and Matthew O. Jackson, eds., *Handbook of Social Economics*, Vol. 1, Amsterdam: North Holland, 2011, pp. 853 – 964. 3

Boisjoly, Johanne, Greg J. Duncan, Michael Kremer, Dan M. Levy, and Jacque Eccles, "Empathy or Antipathy? The Impact of Diversity," *American Economic Review*, December 2006, *96* (5), 1890–1905. 3

Boxell, Levi, Matthew Gentzkow, and Jesse Shapiro, "Is the Internet Causing Political Polarization? Evidence from Demographics," Working Paper 23258, National Bureau of Economic Research March 2017. 2

Bradburn, Norman M., Lance J. Rips, and Steven K. Shevell, "Answering Autobiographical Questions: The Impact of Memory and Inference on Surveys," *Science*, 1987, *236* (4798), 157–161. 21

Bramoullé, Yann, Andrea Galeotti, and Brian Rogers, *The Oxford handbook of the economics of networks*, Oxford, UK: Oxford University Press, 2016. 3

_ , Habiba Djebbari, and Bernard Fortin, "Identification of peer effects through social networks," *Journal of Econometrics*, 2009, *150* (1), 41 – 55. 3, 4, 26

Bruhn, Miriam and David McKenzie, "In Pursuit of Balance: Randomization in Practice in Development Field Experiments," *American Economic Journal: Applied Economics*, 2009, *1* (4), 200–232. 14

Burns, Justine, Lucia Corno, and Eliana La Ferrara, "Does Interaction Affect Racial Prejudice And Cooperation? Evidence From Randomly Assigned Peers In South Africa," *Unpublished*, 2016. 3

Cai, Jing and Adam Szeidl, "Interfirm Relationships and Business Performance," *The Quarterly Journal of Economics*, 2018, *133* (3), 1229–1282. 3

Calvó-Armengol, Antoni, Eleonora Patacchini, and Yves Zenou, "Peer Effects and Social Networks in Education," *The Review of Economic Studies*, 2009, *76* (4), 1239–1267. 3, 26

Camerer, Colin, George Loewenstein, and Martin Weber, "The Curse of Knowledge in Economic Settings: An Experimental Analysis," *Journal of Political Economy*, 1989, *97* (5), 1232–1254. 22

Cameron, Colin and Douglas Miller, "Robust inference for dyadic data," *Unpublished, University of California-Davis*, 2014. 11

_ , Jonah Gelbach, and Douglas Miller, "Robust inference with multiway clustering," *Journal of Business & Economic Statistics*, 2011, *29* (2), 238–249. 11

Campbell, Arthur, C. Matthew Leister, and Yves Zenou, "Social Media and Polarization," *CEPR Discussion Paper*, 2019. 26

Cantoni, Davide, Yuyu Chen, David Y. Yang, Noam Yuchtman, and Y. Jane Zhang, "Curriculum and Ideology," *Journal of Political Economy*, 2017, *125* (2), 338–392. 11

Carlsson, Magnus, Gordon B. Dahl, and Dan-Olof Rooth, "Backlash in Attitudes After the Election of Extreme Political Parties," Working Paper 21062, National Bureau of Economic Research 2015. 2

Carrell, Scott E., Bruce I. Sacerdote, and James E. West, "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation," *Econometrica*, 2013, *81* (3), 855–882. 3, 4, 19

Chandrasekhar, Arun and Randall Lewis, "Econometrics of sampled networks," *Unpublished, MIT*, 2011. 11

_ **, Horacio Larreguy, and Juan Pablo Xandri**, "Testing Models Of Social Learning On Networks: Evidence From Two Experiments," *Unpublished, MIT*, 2018. 5

Currarini, Sergio, Matthew O. Jackson, and Paolo Pin, "An Economic Model of Friendship: Homophily, Minorities, and Segregation," *Econometrica*, 2009, *77* (4), 1003–1045. 3

de Chaisemartin, Clement, "Tolerating defiance? Local average treatment effects without monotonicity," *Quantitative Economics*, 2017, *8* (2), 367–396. 9

De Paula, Aureo, "Econometrics of network models," in Bo Honoré, Ariel Pakes, Monika Piazzesi, and Larry Samuelson, eds., *Advances in Economics and Econometrics: Theory and Applications: Eleventh World Congress*, Cambridge: Cambridge University Press, 2017, pp. 268 – 323. 3

DeGiorgi, Giacomo, Michele Pellizzari, and Silvia Redaelli, "Identification of Social Interactions through Partially Overlapping Peer Groups," *American Economic Journal: Applied Economics*, 2010, *2* (2), 241–275. 3

DeGroot, Morris H., "Reaching a consensus," *Journal of the American Statistical Association*, 1974, *69* (345), 118–121. 5, 26, 30

DellaVigna, Stefano and Eliana La Ferrara, "Media and Social and Economic Outcomes," in Simon Anderson, Joel Waldfogel, and David Stromberg, eds., *Handbook of Media Economics*, Vol. 1, Elsevier, 2015, pp. 723–768. 5

_ **and Ethan Kaplan**, "The Fox News effect: Media bias and voting," *The Quarterly Journal of Economics*, 2007, *122* (3), 1187–1234. 2

_ **and Matthew Gentzkow**, "Persuasion: Empirical Evidence," *Annual Review of Economics*, 2010, *2* (1), 643–669. 2

DeMarzo, Peter, Dimitri Vayanos, and Jeffrey Zwiebel, "Persuasion bias, social influence, and unidimensional opinions," *The Quarterly journal of economics*, 2003, *118* (3), 909–968. 5

Epple, Dennis and Richard E. Romano, "Peer effects in education: A survey of the theory and evidence," in Jesse Benhabib, Alberto Bisin, and Matthew O. Jackson, eds., *Handbook of Social Economics*, Vol. 1, Amsterdam: North Holland, 2011, pp. 1053–1163. 3

Fafchamps, Marcel and Flore Gubert, "Risk Sharing and Network Formation," *American*

Economic Review, 2007, *97* (2), 75–79. 11

Fischhoff, Baruch and Ruth Beyth, "I knew it would happen ? Remembered probabilities of once-future things," *Organizational Behavior and Human Performance*, 1975, *13*, 1–16. 21, 22

Foster, Gigi, "It's not your peers, and it's not your friends: Some progress toward understanding the educational peer effect mechanism," *Journal of Public Economics*, 2006, *90*, 1455–1475. 3

Gabel, Matthew and Kenneth Scheve, "Estimating the effect of elite communications on public opinion using instrumental variables," *American Journal of Political Science*, 2007, *51* (4), 1013–1028. 2

Gentzkow, Matthew, "Television and voter turnout," *The Quarterly Journal of Economics*, 2006, *121* (3), 931–972. 2

_, Jesse Shapiro, and Michael Sinkinson, "The Effect of Newspaper Entry and Exit on Electoral Politics," *American Economic Review*, December 2011, *101* (7), 2980–3018. 2

Gerber, Alan, Dean Karlan, and Daniel Bergan, "Does the Media Matter? A Field Experiment Measuring the Effect of Newspapers on Voting Behavior and Political Opinions," *American Economic Journal: Applied Economics*, April 2009, *1* (2), 35–52. 2

Giuliano, Paola, "Living Arrangements in Western Europe: Does Cultural Origin Matter?," *Journal of the European Economic Association*, 2007, *5* (5), 927–952. 5

_ and Antonio Spilimbergo, "Growing Up in a Recession: Beliefs and the Macroeconomy," *Review of Economic Studies*, 2014, *8*, 787–817. 5

_ and Nathan Nunn, "Understanding Cultural Persistence and Change," Working Paper 23617, National Bureau of Economic Research, Inc 2017. 5

Goeree, Jacob, Margaret McConnell, Tiffany Mitchell, Tracey Tromp, and Leeat Yariv, "The 1/d law of giving," *American Economic Journal: Microeconomics*, 2010, *2* (1), 183–203. 11

Goldsmith-Pinkham, Paul and Guido Imbens, "Social networks and the identification of peer effects," *Journal of Business & Economic Statistics*, 2013, *31* (3), 253–264. 3

Golub, Ben and Evan Sadler, "Learning in Social Networks," in Yann Bramoullé, Andrea Galeotti, and Brian Rogers, eds., *The Oxford Handbook of the Economics of Networks*, Oxford, UK: Oxford University Press, 06 2016, chapter 19. 5

Golub, Benjamin and Matthew O. Jackson, "Naïve Learning in Social Networks and the Wisdom of Crowds," *American Economic Journal: Microeconomics*, 2010, *2* (1), 112–49. 5

_ and _ , "How homophily affects the speed of learning and best-response dynamics," *The Quarterly Journal of Economics*, 2012, *127* (3), 1287–1338. 3, 5, 26, 28, 30

Goyal, Sanjeev, "Learning in networks," in Jesse Benhabib, Alberto Bisin, and Matthew O. Jackson, eds., *Handbook of Social Economics*, Vol. 1, Amsterdam: North Holland, 2011, pp. 679–727. 5

Graham, Bryan S., "Methods of Identification in Social Networks," *Annual Review of Economics*, 2015, *7* (1), 465–485. 3

— **and Aureo De Paula**, *The Econometric Analysis of Network Data*, New York, NY: Academic Press, 2018. 3

Grimm, Veronika and Friederike Mengel, "Experiments on Belief Formation in Networks," *Journal of the European Economic Association*, forthcoming. 5

Harmon, Nikolaj, Raymond Fisman, and Emir Kamenica, "Peer effects in legislative voting," *American Economic Journal: Applied Economics*, forthcoming. 3

Imbens, Guido, "Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009)," *Journal of Economic Literature*, 2010, *48* (2), 399–423. 9

— **and Joshua Angrist**, "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 1994, *62* (2), 467–475. 8, 9

Ioannides, Yannis M., *From neighborhoods to nations: The economics of social interactions*, Princeton, NJ: Princeton University Press, 2013. 3

Jackson, Matthew O., *Social and Economic Networks*, Princeton, NJ: Princeton University Press, 2008. 4

— , "An overview of social networks and economic applications," in Jesse Benhabib, Alberto Bisin, and Matthew O. Jackson, eds., *Handbook of Social Economics*, Vol. 1, Amsterdam: North Holland, 2011, pp. 511–585. 3

— **, Brian Rogers, and Yves Zenou**, "The economic consequences of social-network structure," *Journal of Economic Literature*, 2017, *55* (1), 49–95. 3

Kendall, Chad, Tommaso Nannicini, and Francesco Trebbi, "How Do Voters Respond to Information? Evidence from a Randomized Campaign," *American Economic Review*, 2015, *105* (1), 322–53. 2

Kitagawa, Toru, "A Test for Instrument Validity," *Econometrica*, 2015, *83* (5), 2043–2063. 9, 42

Lazarsfeld, Paul and Robert Merton, "Friendship as a social process: A substantive and methodological analysis," in Morroe Berger, Theodore Abel, and Charles H. Page, eds., *Freedom and control in modern society*, New York: Van Nostrand, 1954, pp. 18–66. 2

— **, Hazel Gaudet, and Bernard Berelson**, *The People's Choice: How the Voter Makes Up His Mind in a Presidential Campaign.*, New York, NY: Columbia University Press, 1944. 2

Lee, David S. and Thomas Lemieux, "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 2010, *48* (2), 281–355. 10

Lee, Lung-Fei and Xiaodong Liu, "Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances," *Econometric Theory*, 2010, *26* (1), 187–230. 4

Leider, Stephen, Markus Möbius, Tanya Rosenblat, and Quoc-Anh Do, "Directed altruism and enforced reciprocity in social networks," *The Quarterly Journal of Economics*, 2009, *124* (4), 1815–1851. 2, 4, 7, 11, 12, 31

_ , _ , _ , and _ , “What do we expect from our friends?,” *Journal of the European Economic Association*, 2010, *8* (1), 120–138. 2

Lin, Xu, “Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables,” *Journal of Labor Economics*, 2010, *28* (4), 825–860. 4

List, John, Fatemeh Momeni, and Yves Zenou, “Are estimates of early education programs too pessimistic? Evidence from a large-scale field experiment that causally measures neighbor effects,” *CEPR Discussion Paper No. 13725*, 2019. 3

Liu, Xiaodong, Eleonora Patacchini, and Yves Zenou, “Endogenous peer effects: local aggregate or local average?,” *Journal of Economic Behavior & Organization*, 2014, *103*, 39–59. 4

Manski, Charles F., “Identification of Endogenous Social Effects: The Reflection Problem,” *The Review of Economic Studies*, 1993, *60* (3), 531–542. 3, 4

McPherson, Miller, Lynn Smith-Lovin, and James M Cook, “Birds of a feather: Homophily in social networks,” *Annual Review of Sociology*, 2001, *27* (1), 415–444. 3

Mele, Angelo, “A structural model of dense network formation,” *Econometrica*, 2017, *85* (3), 825–850. 3

Möbius, Markus and Tanya Rosenblat, “Social Learning in Economics,” *Annual Review of Economics*, 2014, *6* (1), 827–847. 5, 26, 30

_ , Tuan Phan, and Adam Szeidl, “Treasure hunt: Social learning in the field,” Technical Report, National Bureau of Economic Research 2015. 5

Pariser, Eli, *The filter bubble: What the Internet is hiding from you*, Penguin UK, 2011. 2

Patacchini, Eleonora and Yves Zenou, “Social networks and parental behavior in the intergenerational transmission of religion,” *Quantitative Economics*, 2016, *7* (3), 969–995. 3

Rao, Gautam, “Familiarity Does Not Breed Contempt: Generosity, Discrimination and Diversity in Delhi Schools,” *American Economic Review*, 2019, *109* (3), 774–809. 3, 5

Rouban, Luc, “Sociologie politique des députés de la Ve République, 1958–2007,” *Cahiers du CEVIPOF*, 2011, *55*, 1–99. 5

_ , *La fonction publique en débat* number 5396-97. In ‘Les Études.’, La Documentation française, 2014. 5

Sacerdote, Bruce, “Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?,” in Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, eds., *Handbook of the Economics of Education*, Vol. 3, Amsterdam: North Holland, 2011, pp. 249 – 277. 3, 19

_ , “Experimental and Quasi-Experimental Analysis of Peer Effects: Two Steps Forward?,” *Annual Review of Economics*, 2014, *6* (1), 253–272. 3, 19

Sunstein, Cass, *Republic. com 2.0*, Princeton, NJ: Princeton University Press, 2009. 2

_ , *# Republic: Divided democracy in the age of social media*, Princeton, NJ: Princeton University

Press, 2018. 2

Tiberj, Vincent, "Sciences Po, dix ans après les Conventions Education Prioritaire," Working Paper, Sciences Po 2011. 7

Topa, Giorgio and Yves Zenou, "Neighborhood versus network effects," in Gilles Duranton, J. Vernon Henderson, and William C. Strange, eds., *Handbook of Regional and Urban Economics*, Vol. 5, Amsterdam: North Holland, 2015. 3

Wagenaar, Willem, "My Memory: A Study of Autobiographical Memory over Six Years," *Cognitive Psychology*, 1986, *18* (2), 225–252. 21

A Appendix: Analysis of compliers

In this appendix, we detail the technical derivations of formulae that describe the subsample of compliers in a LATE setting. Denote the three sets of compliers, never-takers, and always-takers respectively as $\mathcal{C}$, $\mathcal{N}$, and $\mathcal{A}$.

Appendix Table A4 illustrates the four cases corresponding to the values of friendship L (the treatment variable) and same integration-group IG (the instrument), in a standard LATE setting where both treatment and instrument are binary variables. Among the four cases, it is clear that $\{(i,j)|L=0, IG=1\} \subseteq \mathcal{N}$ (pairs who would not befriend even if assigned into the same IG must be never-takers), and $\{(i,j)|L=1, IG=0\} \subseteq \mathcal{A}$ (pairs who would befriend even without being in the same IG must be always-takers). The other two cases are composite, and include both compliers and never-takers/always-takers: $\{(i,j)|L=0, IG=0\} \subseteq \mathcal{N} \cup \mathcal{C}$ and $\{(i,j)|L=1, IG=1\} \subseteq \mathcal{A} \cup \mathcal{C}$ (there are no defiers, given the monotonicity assumption). For simplicity, we denote $Pr((i,j) \in \mathcal{N}|L, IG)$ as $Pr(\mathcal{N}|L, IG)$.

The unconditional shares of never-takers and always-takers can be simply calculated thanks to the exogeneity of the instrument IG as: $Pr(\mathcal{N}) = Pr(\mathcal{N}|IG=1) = \frac{Pr(L=0,IG=1)}{Pr(IG=1)}$, and similarly $Pr(\mathcal{A}) = Pr(\mathcal{A}|IG=0) = \frac{Pr(L=1,IG=0)}{Pr(IG=0)}$. We thus deduce the share of compliers as $Pr(\mathcal{C}) = 1 - \frac{Pr(L=0,IG=1)}{Pr(IG=1)} - \frac{Pr(L=1,IG=0)}{Pr(IG=0)}$.

Remark 1 $Pr(\mathcal{N}) = \frac{Pr(L=0,IG=1)}{Pr(IG=1)}$, $Pr(\mathcal{A}) = \frac{Pr(L=1,IG=0)}{Pr(IG=0)}$, $Pr(\mathcal{C}) = 1 - \frac{Pr(L=0,IG=1)}{Pr(IG=1)} - \frac{Pr(L=1,IG=0)}{Pr(IG=0)}$.

Also using the exogeneity of the instrument IG, we can estimate the conditional share of never-takers in the case $\{(i,j)|L=0, IG=0\}$ by observing that $Pr(\mathcal{N}|IG=0) = Pr(\mathcal{N}|IG=1)$, therefore:

$$Pr(\mathcal{N}|L=0, IG=0) = \frac{Pr(\mathcal{N}|IG=0)}{Pr(L=0|IG=0)} = \frac{Pr(\mathcal{N}|IG=1)}{Pr(L=0|IG=0)} = \frac{Pr(L=0|IG=1)}{Pr(L=0|IG=0)},$$

which is readily available from the data. Similarly, $Pr(\mathcal{A}|L=1, IG=1) = \frac{Pr(L=1|IG=0)}{Pr(L=1|IG=1)}$. The conditional share of compliers in the two cases $\{(i,j)|L=0, IG=0\}$ and $\{(i,j)|L=1, IG=1\}$ are thus $1 - \frac{Pr(L=0|IG=1)}{Pr(L=0|IG=0)}$ and $1 - \frac{Pr(L=1|IG=0)}{Pr(L=1|IG=1)}$, respectively.

Remark 2 $Pr(\mathcal{C}|L=0, IG=0) = 1 - \frac{Pr(L=0|IG=1)}{Pr(L=0|IG=0)}$, $Pr(\mathcal{C}|L=1, IG=1) = 1 - \frac{Pr(L=1|IG=0)}{Pr(L=1|IG=1)}$.

An analogous argument applies to any statistics defined on the population, such as the mean of a random variable S, $\mathbb{E}[S]$. The exogeneity and monotonicity assumptions imply that $\mathbb{E}[S|\mathcal{N}] = \mathbb{E}[S|\mathcal{N}, IG=1] = \mathbb{E}[S|L=0, IG=1]$, and similarly $\mathbb{E}[S|\mathcal{A}] = \mathbb{E}[S|\mathcal{A}, IG=0] = \mathbb{E}[S|L=1, IG=0]$, both quantities readily estimable as subsample means in the data. It follows further from

$$\mathbb{E}[S] = Pr(\mathcal{N})\mathbb{E}[S|\mathcal{N}] + Pr(\mathcal{A})\mathbb{E}[S|\mathcal{A}] + Pr(\mathcal{C})\mathbb{E}[S|\mathcal{C}],$$

that the unconditional mean of S among compliers is:

$$\mathbb{E}[S|\mathcal{C}] = \frac{\mathbb{E}[S] - Pr(\mathcal{N})\mathbb{E}[S|\mathcal{N}] - Pr(\mathcal{A})\mathbb{E}[S|\mathcal{A}]}{1 - Pr(\mathcal{N}) - Pr(\mathcal{A})}. \tag{1}$$

Regarding conditional means, observe that $\mathbb{E}[S|IG = 1, \mathcal{N}] = \mathbb{E}[S|IG = 0, \mathcal{N}]$, therefore $\mathbb{E}[S|L = 0, IG = 0, \mathcal{N}] = \mathbb{E}[S|L = 0, IG = 1, \mathcal{N}] = \mathbb{E}[S|L = 0, IG = 1]$. It follows from

$$\begin{aligned}\mathbb{E}[S|L = 0, IG = 0] = Pr(\mathcal{N}|L = 0, IG = 0)\mathbb{E}[S|L = 0, IG = 0, \mathcal{N}] +\\ +Pr(\mathcal{C}|L = 0, IG = 0)\mathbb{E}[S|L = 0, IG = 0, \mathcal{C}],\end{aligned}$$

$$\begin{aligned}\Rightarrow \mathbb{E}[S|L = 0, IG = 0, \mathcal{C}] &= \frac{\mathbb{E}[S|L = 0, IG = 0] - Pr(\mathcal{N}|L = 0, IG = 0)\mathbb{E}[S|L = 0, IG = 0, \mathcal{N}]}{Pr(\mathcal{C}|L = 0, IG = 0)} =\\ &= \frac{Pr(L = 0|IG = 0)\mathbb{E}[S|L = 0, IG = 0] - Pr(L = 0|IG = 1)\mathbb{E}[S|L = 0, IG = 1]}{Pr(L = 0|IG = 0) - Pr(L = 0|IG = 1)}. \end{aligned}\tag{2}$$

We thus obtain an estimate of the mean of S among non-treated compliers $\mathbb{E}[S|L = 0, IG = 0, \mathcal{C}]$ as a "weighted average" of its averages over two cases $\{(i, j)|L = 0, IG = 0\}$ and $\{(i, j)|L = 0, IG = 1\}$ (but with some negative weights). A similar estimate of the mean of S among treated compliers $\mathbb{E}[S|L = 1, IG = 1, \mathcal{C}]$ is obtained analogously by:

$$\frac{Pr(L = 1|IG = 1)\mathbb{E}[S|L = 1, IG = 1] - Pr(L = 1|IG = 0)\mathbb{E}[S|L = 1, IG = 0]}{Pr(L = 1|IG = 1) - Pr(L = 1|IG = 0)}. \tag{3}$$

Notably, the difference between the formulae in equations (2) and (3) yields the LATE estimand on the variable S, namely $\frac{\mathbb{E}[S|IG=1]-\mathbb{E}[S|IG=0]}{Pr(L=1|IG=1)-Pr(L=1|IG=0)}$.

Remark 3 *Unconditional and conditional means $\mathbb{E}[S|\mathcal{C}]$, $\mathbb{E}[S|L = 0, IG = 0, \mathcal{C}]$, and $\mathbb{E}[S|L = 1, IG = 1, \mathcal{C}]$ are given in equations (1), (2), and (3).*

In our benchmark regression, in presence of control variables, in order to use those resuklts we can first partial out all the controls to obtain the residual $\widetilde{DY}$ (that is, the residual of regressing the outcome DY, namely the pairwise difference in opinion after 6 months, on the set of controls). Then, we can apply the formulae in equations (2) and (3) to compute the potential outcomes with and without treatment, the difference between which yields the LATE estimate.

We can further apply those formulae to any statistic based on pre-intervention characteristics, including quantiles of the distribution of a measured trait S (e.g., the median and quartiles of tuition fees, a proxy for parents' income), and compare it between the compliers, the never-takers, and the always-takers, in order to understand whether and how compliers are particularly different from the rest of the population.[54]

[54]This derivation is used in Kitagawa's (2015) test to infer the conditional distributions of outcomes among the two subgroups of compliers, and then test whether their density is always nonnegative on its support.

B Appendix: Additional tables

Table A1: Description of Variables in Dyadic Data

Variable	Description
Friendship	1 if at least one of the two individual has named the other as one of her friends (i.e. undirected friendship), zero otherwise.
Same Integration Group	1 if the two individuals have attended the same integration group over the 2013 summer, 0 otherwise.
Same Tutorial Group	1 if the two individuals have been part of the same "triplette" group, 0 otherwise.
Difference in Political Opinion (March 2014)	Absolute difference between the political opinions of the two individuals, as declared on a 1-10 scale.
Initial difference in Political Opinion (August 2013)	Absolute difference between the political opinions of the two individuals, as declared on a 1-10 scale.
Difference in Political Opinion in 2015	Absolute difference between the political opinions of the two individuals, as declared on a 1-10 scale.
Difference in Political Opinion in 2014	Absolute difference between the political opinions of the two individuals, as declared on a 1-10 scale.
Membership in Some Political Party in 2014	0 if the two individuals are enrolled in a political party. Missing if at least one of them did not answer to this question. 0 otherwise.
Membership in Some Political Party in 2013	0 if the two individuals are enrolled in a political party. Missing if at least one of them did not answer to this question. 0 otherwise.
Membership in the Same Political Party in 2014	0 if the two individuals are enrolled in the same political party. Missing if at least one of them did not answer to this question. 0 otherwise.
Membership in Some Association	0 if the two individuals are enrolled in the a student association. Missing if at least one of them did not answer to this question. 0 otherwise.
Membership in the Same Association	0 if the two individuals are enrolled in the same student association. Missing if at least one of them did not answer to this question. 0 otherwise.
Sum of Movements	Sum of absolute movements in political opinion among the two individuals between August 2013 and March 2014.
Staying in Same Position	1 if both individuals have not changed political opinion between August 2013 and March 2014
Moving in Same Direction	1 if both individuals have changed their political opinion between August 2013 and March 2014 and their new political opinion have moved in the same direction relative to their initial one, 0 otherwise.
Moving in Non-Opposing Direction	1 if between August 2013 and March 2014 the two individuals have not changed their political opinion in opposing directions (e.g. one has moved to the right and the other to the left), 0 otherwise.
Moving in Same Converging Direction	1 if both individuals have changed their political opinion between August 2013 and March 2014, their new political opinion have moved in the same direction relative to their initial one and the difference in their political opinions has diminished, 0 otherwise.
Strong Convergence	1 if the two individuals have different initial political positions, none of them have moved away from and at least one of them has moved towards the other initial political opinion relative to her own initial position. Missing if the two individuals have the same initial political opinion. 0 otherwise.
Weak Convergence	1 if the two individuals have different initial political positions, none of them have moved away from the other initial political opinion relative to her own initial position. Missing if the two individuals have the same initial political opinion. 0 otherwise.

Strong Divergence	1 if the two individuals have both moved away from each others initial political position relative to their own initial position, 0 otherwise.
Weak Divergence	1 if the two individuals have not moved towards each others political position relative to their own initial position, 0 otherwise.
Friendship Strength 1	1 1 if at least one of the two individual has named the other as one of her friends and has stated that their friendship is at least as intense as a “mere relationship”, 0 otherwise.
Friendship Strength 2	1 1 if at least one of the two individual has named the other as one of her friends and has stated that their friendship is at least as intense as a “friendship link”, 0 otherwise.
Friendship Strength 3	1 1 if at least one of the two individual has named the other as one of her friends and has stated that their friendship is at least as intense as a “close friendship”, 0 otherwise.
Friendship Strength 4	1 1 if at least one of the two individual has named the other as one of her friends and has stated that their friendship is at least as intense as a “very close friendship”, 0 otherwise.
Shortest Path	Shortest path between the two individuals.
1st vs 2nd order only	Equal to the shortest path if this is either 1 or 2, missing otherwise.
2nd vs 3rd order only	Equal to the shortest path if this is either 2 or 3, missing otherwise.
3rd vs more order only	Equal to the shortest path if this is either 3 or more, missing otherwise.
Difference in Differences in Political Opinion	Difference in Political Opinion in March 2014 minus Difference in Political Opinion in August 2013
Same Gender	1 if the two individuals are of the same gender, 0 otherwise.
Both Female	1 if the two individuals are both female, 0 otherwise.
Same Nationality	1 if the two individuals share a common nationality, 0 otherwise.
Same French	1 if the two individuals are both french, 0 otherwise.
Same Double French Nationality	1 if the two individuals are both french and they both have a second nationality.
Same Admission Type	1 if the two individuals have been admitted through the same admission procedure, 0 otherwise.
Both Priority Admission	1 if the two individuals have both been admitted through the priority admission procedure, 0 otherwise.
Same Department of High School	1 if the two individuals have completed their high school diploma in the same french department, 0 otherwise.
Same Region of High School	1 if the two individuals have completed their high school diploma in the same french region, 0 otherwise.
Same High School Major	1 if the two individuals have a high school diploma with the same major, 0 otherwise.
Difference in Tuition Fees	Absolute difference in tuition fees among the couple (proxy for family income).
Both Free Tuition	1 if both individuals do not pay tuition fees, 0 otherwise.
Same Parents’ Profession	1 if at least one of an individual’s parents has common profession with the parents of the other individual, 0 otherwise.
Same ZIP code	1 if the two individuals live in the same ZIP code area, 0 otherwise.
Same Program	1 if the two individuals are enrolled in the same program, 0 otherwise.

Notes on sample construction: The sample excludes observations (pairs of students) in which any of the above-mentioned variables is missing, when at least one of the two individuals in the couple did not answer to the related question in the survey. We also drop pairs that contain at least one individual in the top 5 percent of the distribution of time taken to name each friend (about 82 seconds per friend or 13.5 minutes for individuals with 10 friends).

Table A2: Additional Descriptive Statistics of Covariates

Panel A: Independent Variables by Individual

Variable	(1) Full Sample			(2) Benchmark Sample		
	Mean	Standard deviation	Mean	Obs.	Standard deviation	Obs.
Gender (1= Female)	0.592	(0.492)	796	0.583	(0.494)	331
Honors Graduation	0.754	(0.431)	796	0.831	(0.375)	331
Tuition Fees	3602	(3495)	713	3826	(3328)	331

Notes: Summary statistics (1) refer to the full individual sample, where the full sample is made of all the individual observations for which the variable described is not missing. Summary statistics (2) refer to the benchmark sample, where the benchmark sample is defined as the individual sample containing all the individuals that are present in our benchmark dyadic sample as detailed in Table A1.

Panel B: Independent Variables by Dyad

Variable	(1) Full Sample			(2) Benchmark Sample		
	Mean	Standard deviation	Obs.	Mean	Standard deviation	Obs.
Same Gender	0.522	(0.4995)	294,306	0.512	(0.4998)	54,615
Both Female	0.369	(0.483)	294,306	0.339	(0.473)	54,615
Same Nationality	0.928	(0.259)	291,060	0.964	(0.186)	54,615
Same Double French Nationality	0.774	(0.418)	294,680	0.778	(0.416)	54,615
Same Admission Type	0.565	(0.496)	294,306	0.675	(0.468)	54,615
Both Priority Admission	0.029	(0.168)	294,306	0.015	(0.122)	54,615
Same Department of High School	0.052	(0.221)	265,740	0.059	(0.237)	54,615
Same Region of High School	0.253	(0.435)	264,710	0.251	(0.434)	54,615
Same High School Major	0.363	(0.481)	294,306	0.375	(0.484)	54,615
Difference in Tuition Fees	3879	(3005)	245,520	3759	(2832)	54,615
Both Free Tuition	0.476	(0.499)	294,306	0.614	(0.487)	54,615
Same Parents' Profession	0.422	(0.494)	238,632	0.442	(0.497)	54,615
Same ZIP code	0.026	(0.160)	293,222	0.026	(0.159)	54,615
Same Program	0.520	(0.4995)	294,306	0.508	(0.4999)	54,615

Notes: Summary statistics (1) refer to the full dyadic sample, where full sample is defined as the set of all pairs for which both members named at least one friend or stated that they have no friends in Sciences Po. Summary statistics (2) refer to the benchmark dyadic sample as detailed in Table A1.

Table A3: Permutation tests of integration group assignment's randomness

Variable	Within-Group Statistics	Actual value	p-value
Initial Political Opinion (Aug. 2013)	Standard Deviation	1.859	0.937
Tuition Fees	Standard Deviation	3289.036	0.420
Gender	Herfindahl-Hirschman Index	0.108	0.843
First Nationality	Herfindahl-Hirschman Index	0.099	1.000
Second Nationality	Herfindahl-Hirschman Index	0.500	0.923
Admission Type	Herfindahl-Hirschman Index	0.114	0.997
Program	Herfindahl-Hirschman Index	0.140	0.613
Parents' Profession	Herfindahl-Hirschman Index	0.139	0.843
High School Major	Herfindahl-Hirschman Index	0.099	1.000
Département of High School	Herfindahl-Hirschman Index	0.121	0.857
Region of High School	Herfindahl-Hirschman Index	0.116	0.783

Notes: Permutation tests over the full sample are performed over 300 Monte Carlo draws. The within-group statistics, either the standard deviation or the Herfindahl-Hirschman index, is measured as group-size-weighted average of its value in each (permuted) integration group. Left-tail tests are reported for the within-group standard deviation, and right-tail tests for the Herfindahl-Hirschman Index, for both of which the alternative hypothesis is that there is too little within-group variation.

Table A4: Analysis of compliers, never-takers, and always-takers

			Instrument: Same Integration Group	
			IG = 0	IG = 1
Treatment: Friendship	L=0	Composition	Compliers, Never-takers	Never-takers
		Observations	8,924 Cs, 43,938 Ns	807 Ns
		Shares	16.3% Cs, 80.4% Ns	1.5% Ns
		$\mathbb{E}[\widetilde{DY}]$	Cs: 0.31, Ns: -0.06	Ns: -0.06
		Pr[AA and non-AA]	Cs: 0.29, Ns: 0.20	Ns: 0.20
	L=1	Composition	Always-takers	Compliers, Always-takers
		Observations	768 As	164 Cs, 14 As
		Shares	1.4% As	0.3% Cs, 0.03% As
		$\mathbb{E}[\widetilde{DY}]$	As: -0.07	Cs: -0.20, As: -0.07
		Pr[AA and non-AA]	As: 0.11	Cs: 0.23, As: 0.11

Notes: This table reports the analysis of compliers, including the subsample sizes and statistics among compliers (Cs), never-takers (Ns), and always-takers (As) in the for cases based on the values of the treatment L_{ij} and the instrument IG_{ij}.

Table A5: Descriptive Statistics on Recall Bias

		Percentages (Actual Pol. Op. in 2014 as a reference)										N
		Actual (Individual) Political Opinion in 2014										
		1	2	3	4	5	6	7	8	9	10	Total
Recalled P. Op. in 2014	1	0	1	0	0	0	0	0	0	0	0	1
	2	0	5	1	2	0	0	0	0	0	0	8
	3	1	6	19	7	3	1	0	0	0	0	37
	4	0	0	7	16	21	4	1	0	0	0	49
	5	0	0	2	7	25	6	1	0	0	0	41
	6	1	0	0	1	6	21	8	3	0	0	40
	7	0	0	0	1	0	6	12	5	0	0	24
	8	0	0	0	0	0	1	6	6	1	0	14
	9	0	0	0	0	0	0	2	1	0	1	4
	10	0	0	0	0	0	0	0	0	0	0	0
	Total	2	12	29	34	55	39	30	15	1	1	218

Notes: The table reports the joint empirical distribution of actual (horizontal axes) and recalled (vertical axes) individual political opinion in 2014. The table makes use of the individual linked 2014-2015 dataset. The sample includes only those individuals present both in the 2014 and in the 2015 survey and for which the variables "actual political opinion in 2014" and "recalled political opinion in 2014" are not missing.

Table A6: Recall Bias Regression on Individual Data

Dependent Variable:	Absolute Recall Bias	Recall Bias
	(1)	(2)
Actual Political Opinion in 2015	0.00426	-
	(0.116)	
Actual Political Opinion in 2014	0.00609	-
	(0.137)	
Diff. in Actual Political Opinion Between 2015 and 2014	-	0.574***
		(0.0437)
Observations	216	216
Double Group Clust.	Yes	Yes

Notes: These OLS regressions aim to predict recall bias based on actual opinions, on the individual linked 2014-2015 sample, including individuals present in both the 2014 and the 2015 survey, for which the variables "political opinion in 2015", "actual political opinion in 2014" and "recalled political opinion in 2014" are not missing. The outcome variable "Recall Bias" is calculated as recalled political opinion of 2014, as answered in the 2015 survey, minus actual political opinion in 2014, as answered in the 2014 survey. "Absolute Recall Bias" is the absolute value of Recall Bias. Standard errors are clustered at the group level.

Table A7: Convergence of Political Opinion - Alternative Specification

Dependent Variable:	Difference in Differences in Political Opinion (March 2014 - August 2013)			
	(1)	(2)	(3)	(4)
	IV	IV	IV	IV
Friendship	-0.841***	-0.902***	-0.928***	-0.940***
	(0.275)	(0.286)	(0.298)	(0.330)
Same Tutorial Group	-	-	0.310***	0.271
			(0.101)	(0.191)
Friendship * Same Tutorial Group	-	-	-	0.117
				(0.615)
Observations	54,615	54,615	54,615	54,615
IV	Same Integration Group			
Controls	No	Yes	Yes	Yes
Double Group Clust.	Yes	Yes	Yes	Yes
WeakIV test stat.	114.2	113.2	108.8	20.07

Notes: Standard errors are two-way clustered by individual 1's group and by individual 2's group. The benchmark sample is detailed in Table A1. Weak IV stat reports the Kleibergen-Paap cluster-robust statistic, distributed as a Chi-squared under the null hypothesis of weak identification.

Table A8: Initial Opinion Gap and Convergence of Political and Associative Activities

Panel A: Convergence of Political and Associative Activities - Conditional on Similar Initial Opinion

Dependent Variables:	Membership in Student Associations	Membership in Same Association*	Membership in Same Association	Membership in a Political Party	Membership in the Same Political Party*
Condition:	Difference in Initial Political Opinion < 2				
	(1)	(2)	(3)	(4)	(5)
Friendship	-0.267*	0.223	0.0788	0.0424	0.185
	(0.157)	(0.146)	(0.0665)	(0.0949)	(0.336)
Political Membership	-	-	-	0.449***	-
(August 2013)				(0.0769)	
Observations	19,073	9,727	21,949	21,274	305
IV	Same Integration Group				
Controls	Yes	Yes	Yes	Yes	Yes
Double Group Clust.	Yes	Yes	Yes	Yes	Yes
WeakIV test stat.	51.83	36.26	59.77	62.83	2.618

Panel B: Convergence of Political and Associative Activities - Conditional on Different Initial Opinion

Dependent Variables:	Membership in Student Associations	Membership in Same Association*	Membership in Same Association	Membership in a Political Party	Membership in the Same Political Party*
Condition:	Difference in Initial Political Opinion ≥ 2				
	(1)	(2)	(3)	(4)	(5)
Friendship	0.134	0.140	0.0742*	0.0355	-0.687
	(0.136)	(0.104)	(0.0434)	(0.0863)	(0.433)
Political Membership	-	-	-	0.480***	-
(August 2013)				(0.0749)	
Observations	28,822	14,583	32,666	31,376	475
IV	Same Integration Group				
Controls	Yes	Yes	Yes	Yes	Yes
Double Group Clust.	Yes	Yes	Yes	Yes	Yes
WeakIV test stat.	71.18	43.36	91.45	83.85	1.506

Panel C: Convergence on Associative Activities - Conditional on Different Initial Opinion

	Membership in						
Dependent Variable:	Same Association					Associations of Same Type	Different Associations of Same Type
Type of Association:	Strictly Political or Student Union	Political	Related to One's Identity	Related to Policy Issues	Sports	Political	Political
Condition:	Difference in Initial Political Opinion ≥ 2						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Friendship	-0.0108	0.0500	-0.0728	-0.123	0.057	-0.0252***	0.00575
	(0.0299)	(0.0802)	(0.162)	(0.146)	(0.2378)	(0.00893)	(0.0358)
Observations	14,583	14,583	14,583	14,583	14,583	14,583	14,583
IV	Same Integration Group						
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Double Group Clust.	Yes	Yes	Yes	Yes	Yes	Yes	Yes
WeakIV test stat.	43.36	43.36	43.36	43.36	43.36	43.36	43.36

* conditional on the individuals in the couple being enrolled in some association/political party.

Notes: The dependent dummy variables are equal to 0 if the characteristic is the same for the two individuals in the pair and 1 if the characteristic is not the same. Standard errors are two-way clustered by individual 1's group and by individual 2's group. The benchmark sample is detailed in Table A1. Weak IV stat reports the Kleibergen-Paap cluster-robust statistic, distributed as a Chi-squared under the null hypothesis of weak identification.

Table A9: Initial Opinion Gap, Convergence, Divergence, and Co-movement

Panel A: Weak Divergence and Initial Differences in Political Opinion

Dependent Variables:	Weak Divergence in Political Opinion					
Conditions		Difference in Initial Political Opinion: =0	=1	=2	> 2	< 2
	(1)	(2)	(3)	(4)	(5)	(6)
Friendship	-0.231	-0.141	-0.193	-0.104	-0.154*	-0.173*
	(0.147)	(0.213)	(0.124)	(0.199)	(0.0794)	(0.101)
Friendship*Initial Diff.	0.0309	-	-	-	-	-
in Political Opinion	(0.0411)					
Initial Diff. in Political	-0.0838***	-	-	-	-0.0311***	-0.371***
Opinion (August 2013)	(0.00295)				(0.00447)	(0.00524)
Observations	54,615	7,714	14,235	11,644	21,022	21,949
IV	Same Integration Group					
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Double Group Clust.	Yes	Yes	Yes	Yes	Yes	Yes
WeakIV test stat.	30.97	46.36	41.59	18.14	68.69	59.77

Panel B: Strong Convergence, Weak Convergence, and Initial Differences in Political Opinion

Dependent Variables:	Strong Convergence in Political Opinion				Weak Convergence in Political Opinion			
Conditions		Difference in Initial Political Opinion: =1	=2	> 2		Difference in Initial Political Opinion: =1	=2	> 2
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Friendship	-0.17	-0.0056	-0.212	0.00387	0.0388	0.0128	-0.0588	0.0892
	-0.167	-0.115	-0.129	-0.176	-0.189	-0.149	-0.233	(0.135)
Friendship*Initial Diff.	0.0486	-	-	-	-0.00186	-	-	-
in Political Opinion	-0.0803				-0.067			
Initial Diff. in Political	0.0590***	-	-	0.0531***	0.110***	-	-	0.0725***
Opinion (August 2013)	-0.0064			-0.0129	-0.00667			(0.0101)
Observations	46,901	14,235	11,644	21,022	46,901	14,235	11,644	21,022
IV	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Double Group Clust.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
WeakIV test stat.	22.53	41.59	18.14	68.69	22.53	41.59	18.14	68.69

Panel C: Co-movement and Initial Differences in Political Opinion

Dependent Variables:	Moving in the Same Direction					
Conditions		Difference in Initial Political Opinion: =0	=1	=2	>2	< 2
	(1)	(2)	(3)	(4)	(5)	(6)
Friendship	0.177*	0.270	0.0295	0.143	0.0650	0.111
	(0.103)	(0.187)	(0.0804)	(0.255)	(0.0967)	(0.0722)
Friendship*Initial Diff.	-0.0357	-	-	-	-	-
in Political Opinion	(0.0386)					
Initial Diff. in Political	-0.0271***	-	-	-	-0.0265***	-0.00642
Opinion (August 2013)	(0.00601)				(0.00503)	(0.00403)
Observations	54,615	7,714	14,235	11,644	21,022	21,949
IV	Same Integration Group					
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Double Group Clust.	Yes	Yes	Yes	Yes	Yes	Yes
WeakIV test stat.	30.97	46.36	41.59	18.14	68.69	59.77

Notes: Denoting ΔY_i as i's signed change in opinion from $t = 0$ (before Sciences Po) to $t = 1$ (survey time), weak divergence means $\Delta Y_i(Y_{j0} - Y_{i0}) \leq 0$ & $\Delta Y_j(Y_{i0} - Y_{j0}) \leq 0$, strong convergence means $\Delta Y_i(Y_{j0} - Y_{i0}) > 0$ & $\Delta Y_j(Y_{i0} - Y_{j0}) > 0$, weak convergence means $\Delta Y_i(Y_{j0} - Y_{i0}) \geq 0$ & $\Delta Y_j(Y_{i0} - Y_{j0}) \geq 0$, and co-movement means $\Delta Y_i \Delta Y_j \geq 0$. Standard errors are two-way clustered by individual 1's group and by individual 2's group. The benchmark sample is detailed in Appendix Table A1. Weak IV stat reports the Kleibergen-Paap cluster-robust statistic, distributed as a Chi-squared under the null hypothesis of weak identification.

Table A10: Main Result after Excluding Nationalities

Dependent Variable:	Difference in Political Opinion			
Excluding:	Algeria	Germany	Belgium	Spain
Friendship	-0.436*	-0.522**	-0.534**	-0.512**
	(0.241)	(0.238)	(0.233)	(0.220)
Observations	53,301	52,975	53,956	54,285
Excluding:	Italy	Madagascar	Morocco	Senegal
Friendship	-0.543**	-0.546**	-0.593***	-0.546**
	(0.233)	(0.228)	(0.228)	(0.229)
Observations	53,628	54,285	53,301	54,285

Notes: Standard errors are two-way clustered by individual 1's group and by individual 2's group. The benchmark sample is detailed in Appendix Table A1. Weak IV stat reports the Kleibergen-Paap cluster-robust statistic, distributed as a Chi-squared under the null hypothesis of weak identification.

Table A11: Main Result after Excluding Names Starting with a Given Alphabet Letter

Dependent Variable:	Difference in Political Opinion								
Excluding:	A	B	C	D	E	F	G	H	I
Friendship	-0.586**	-0.472*	-0.376	-0.843***	-0.574***	-0.554**	-0.449**	-0.542**	-0.597**
	(0.243)	(0.268)	(0.290)	(0.269)	(0.216)	(0.229)	(0.220)	(0.230)	(0.239)
Observations	50,403	40,186	46,056	45,150	53,628	50,721	48,205	49,770	53,301
Excluding:	J	K	L	M	N	O	P	Q	R
Friendship	-0.545***	-0.501**	-0.343*	-0.411*	-0.532**	-0.624**	-0.599***	-0.557**	-0.573**
	(0.204)	(0.229)	(0.201)	(0.222)	(0.226)	(0.247)	(0.222)	(0.224)	(0.230)
Observations	51,681	53,628	43,071	46,056	53,956	53,301	50,086	54,285	49,141
Excluding:	S	T	U	V	W	X	Y	Z	"DE " , "D ", "DU "
Friendship	-0.548**	-0.526**	-0.546**	-0.476*	-0.535**	-0.538**	-0.545**	-0.531**	-0.654**
	(0.235)	(0.235)	(0.234)	(0.243)	(0.233)	(0.228)	(0.235)	(0.231)	(0.264)
Observations	48,205	50,403	54,285	52,003	54,285	54,615	53,956	54,285	51,681

Notes: Standard errors are two-way clustered by individual 1's group and by individual 2's group. The benchmark sample is detailed in Appendix Table A1. Weak IV stat reports the Kleibergen-Paap cluster-robust statistic, distributed as a Chi-squared under the null hypothesis of weak identification.

Table A12: Friendship Effect and Network Centrality by Quartile

Dependent Variable:		Difference in Political Opinion (March 2014)			
		Quartile of Eigenvector Centrality			
		1st	2nd	3rd	4th
Quartile	1st	-1.310***	-1.138	-0.546	-2.960
		(0.457)	(1.555)	(1.371)	(4.484)
	2nd	-	-1.512	0.110	-0.753
			(1.222)	(0.280)	(0.684)
	3rd	-	-	-0.733**	-0.651
				(0.329)	(0.576)
	4th	-	-	-	0.149
					(0.229)

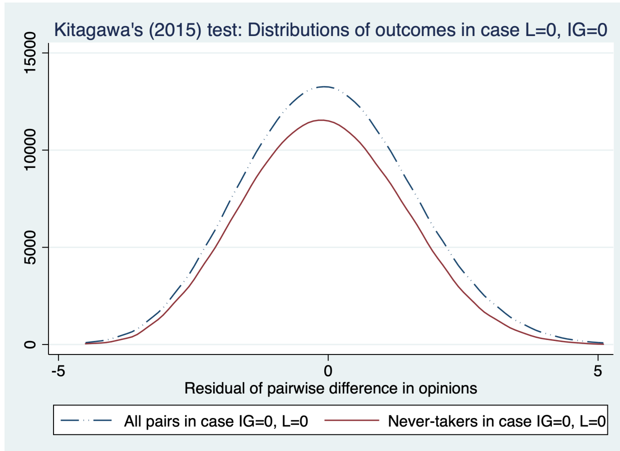
Notes: The table presents a matrix of results for different sub-samples of couples selected based on their eigenvector centrality. Entry (1,1) in the matrix reports the coefficient on friendship in specification (2) - Table 4 for the sub-sample of couples where one both individuals are in the first quartile of eigenvector centrality. Entry (1,2) reports the same coefficient for the sub-sample of couples where one individual is in the first quartile of eigenvector centrality and the other is in the second quartile and so on. Standard errors are two-way clustered by individual 1's group and by individual 2's group. The benchmark sample is detailed in Appendix Table A1. Weak IV stat reports the Kleibergen-Paap cluster-robust statistic, distributed as a Chi-squared under the null hypothesis of weak identification.

Table A13: Heterogeneous Effects of Friendship

Dependent Variable:	Difference in Political Opinion				
	(1)	(2)	(3)	(4)	(5)
Friendship	-0.746*	-0.850**	-0.495**	-0.824*	-0.807*
	(0.413)	(0.392)	(0.247)	(0.422)	(0.425)
Friendship * Same Gender	0.346	-	-	-	-
	(0.555)				
Same Gender	0.0485	-	-	-	-
	(0.0491)				
Friendship * Same Admission	-	0.421	-	-	-
		(0.440)			
Same Admission	-	-0.00543	-	-	-
		(0.0797)			
Friendship * Same Department of High School	-	-	-0.438	-	-
			(0.692)		
Same Department of High School	-	-	0.0102	-	-
			(0.0512)		
Friendship * Number of Common Friends	-	-	-	0.112	-
				(0.208)	
Number of Common Friends	-	-	-	0.0474	-
				(0.0816)	
Friendship * Share of Common Friends	-	-	-	-	1.134
					(1.943)
Share of Common Friends	-	-	-	-	0.314
					(0.723)
Observations	54,615	54,615	54,615	54,615	54,615
IV	Same Group and Interactions				
Controls	Yes	Yes	Yes	Yes	Yes
Double Group Clustering	Yes	Yes	Yes	Yes	Yes
WeakIV test stat.	32.94	60.96	56.83	56.49	29.68

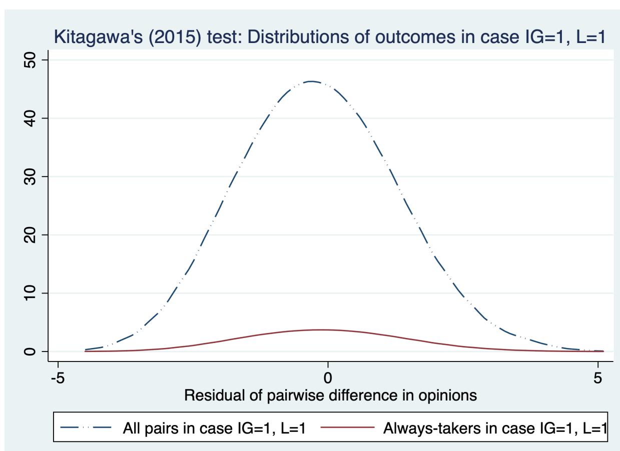
Notes: Standard errors are two-way clustered by individual 1's group and by individual 2's group. The benchmark sample is detailed in Appendix Table A1. Weak IV stat reports the Kleibergen-Paap cluster-robust statistic, distributed as a Chi-squared under the null hypothesis of weak identification.

Figure A1: Distributions of residual outcomes in case $IG = 0$ and $L = 0$



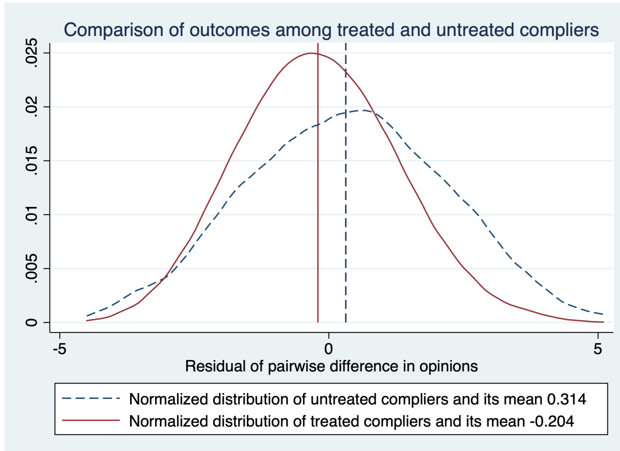
Notes: The two curves are kernel estimates of the distributions of outcomes (the residual of pairwise difference in opinions after partialing out controls, as described in Table 2), drawn for all observations in case IG=0 and L=0, and for never-takers among them. Kitagawa (2015) proposes a test that rejects the validity of the LATE assumptions on the instrumental variable if those two distributions intersect

Figure A2: Distributions of residual outcomes in case $IG = 1$ and $L = 1$



Notes: The two curves are kernel estimates of the distributions of outcomes (the residual of pairwise difference in opinions after partialing out controls, as described in Table 2), drawn for all observations in case IG=1 and L=1, and for always-takers among them. Kitagawa (2015) proposes a test that rejects the validity of the LATE assumptions on the instrumental variable if those two distributions intersect

Figure A3: Comparison of treated and untreated compliers' outcomes



Notes: The two curves are kernel estimates of the distributions of outcomes (the residual of pairwise difference in opinions after partialing out controls, as described in Table 2), normalized to total weight equal 1, drawn separately for untreated and treated compliers. The vertical lines represent their respective means (0.314 among untreated compliers, -0.204 among treated compliers).