



Note

What drives segregation? Evidence from social interactions among students[☆]Emilio Calvano^{a,b,e,c}, Giovanni Immordino^{d,e}, Annalisa Scognamiglio^{d,e,*}^a Università di Roma Tor Vergata, Italy^b Toulouse School of Economics, France^c CEPR, United Kingdom^d Università di Napoli - Federico II, Italy^e Center for Studies in Economics and Finance, Italy

ARTICLE INFO

Keywords:

Peer pressure

Education

Network formation

ABSTRACT

We study the dynamics of group formation using data on the online social networks of a cohort of undergraduate students. We document gradual endogenous segregation along the ability dimension. We show that “high ability” students interact more and more with high ability students over time (active segregation). Instead, “low ability” students “reach out” to high ability students but are not reciprocated (passive segregation). Exploiting our administrative records on student performance we provide evidence that information about ability is an additional driver of social interactions in educational settings, besides the homophily hypothesis.

1. Introduction

A number of studies in the economic literature showed that the size and characteristics of one's peer group is a key driver of a variety of economic outcomes (De Giorgi et al., 2012). These findings sparked a large debate on public policies to support members of disadvantaged categories by actively sorting individuals into purposefully crafted groups.

A major challenge in implementing these policies is the endogenous response of individuals to these interventions. For instance, in a large controlled field experiment, Carrell et al. (2013) implemented policies designed to leverage peer effects in educational settings to increase the performance of the lowest ability students. These policies typically prescribe to sort students into classrooms with a bi-modal skill distribution. They show that the intervention backfired due to endogenous segregation. Within treated classrooms, students sorted into smaller peer groups of “similar” classmates. They conclude that understanding the endogenous pattern of social interaction is essential to be able to implement these affirmative public policies.

The evidence of segregation begs the question of *why* low ability individuals segregate in the face of the gains from blending with their classmates documented in the peer effects literature. One hypothesis is that homophily (that is taste for like minded people) swamps

material incentives.¹ Alternatively, segregation (in the form of assortative friendships) may be an incentive driven equilibrium outcome, with individuals rejecting to be friends with individuals with lower skills (Shimer & Smith, 2000).

This paper contributes to the debate by studying the *dynamics* of group formation in contexts with heterogeneous agents using observational data on the social network of a cohort of undergraduate students majoring in Economics and Business at the University of Naples Federico II. We provide empirical support for the latter hypothesis documenting that social efforts from low types are not “reciprocated” by higher types in a sense made precise below.

Specifically, we use data from a popular social networking site gathered throughout a key juncture of students' social life: right at the beginning of their undergraduate degrees. This stage is, arguably, particularly suitable to study and test hypothesis on network formation, as most students start their undergraduate life with a clean social slate. Our measures of online social interactions are *directional*. An individual can engage with some other individual's posts (say by “liking”, “sharing” or “commenting” its content) and not be reciprocated. This gives us a time varying measure of social interactions within our network of students, hinging on the hypothesis that online interactions reflect offline ones. However, given how pervasive and diffused this platform

[☆] This project has been developed thanks to a research grant awarded by the Einaudi Institute for Studies in Economics and Finance (EIEF), Italy whose support is gratefully acknowledged.

* Correspondence to: Dipartimento di scienze economiche e statistiche, Università degli Studi di Napoli Federico II, Via Cinthia Monte Sant'Angelo, 80126 Napoli, Italy.

E-mail address: annalisa.scognamiglio2@unina.it (A. Scognamiglio).

¹ Bramoullé et al. (2012), Currarini et al. (2009, 2010), Mele (2017) and Tarbush and Teytelboym (2017) study homophily in networks assuming a dynamic and stochastic matching.

is in the age group of interest, this is arguably a good measure of actual social behavior in peer groups. As detailed in Section 3, the data allows to observe both the intensive (“closeness”) and the extensive margins (“mere friendship”) over time. Crucially, to identify “ability” as a driver of friendship formation, we are able to match these data with administrative records from the University on student performance before and after these social interactions took place, among other things.

Consistent with previous findings, in the first part of the paper we show that students segregate in an assortative manner along the ability dimension. High skill students hang out more and more with high skill peers. Exploiting the panel nature of our data we then document that this segregation occurs *gradually* over the course of the year. This finding is consistent with the hypothesis that, as information about ability naturally reveals, individuals sort accordingly. To support the hypothesis that individuals respond to ability we exploit the fact that first term exams – being publicly held – generate a discontinuity in the revelation of information about students’ types. We show that the grade point average obtained at the end of the first term has a significant positive effect on popularity controlling for possible differential changes in the level of engagement in online activity.

To understand what drives the increase in popularity of high types relative to low types we exploit the directional nature of our data. We show that this increased popularity is driven exclusively by lower types increasing their social effort selectively towards higher types. Interestingly, for the high types we find no differential change in the social effort directed towards high versus low types. These findings are consistent with the hypothesis that low types react to information on ability by “reaching out” to high types, but they are not reciprocated.

Our measure of social interactions hinges on the assumption that online interactions on a particular social media reflect offline ones. Needless to say, online platform use is self-selected and endogenous. In Section 2.1 we document how our sample of “online” students compares to that of other students either not registered on the platform or registered with “private” profiles, highlighting differences. These differences may raise concerns about the external validity of the results, but the findings hold for students that use the platform, and, given how pervasive and diffused this platform is in the age group of interest, our study captures an important part of social interactions and provides valuable evidence on how such interactions evolve over time. Indeed, selection into the sample does not seem to change over time: as shown in Fig. 1 in the Appendix, there is no clear pattern in the number of new registrations to the platform over the relevant time period.²

A related literature consists of empirical papers that try to understand social interactions or the dynamic learning of ability. Farber and Gibbons (1996) develop a dynamic model of learning about worker ability in a competitive labor market. The model produces two testable implications regarding wage dynamics close in spirit, although in a very different setting, to what we find in our results. First, the role of schooling in the labor market’s inference process declines as performance observations accumulate; second, time invariant variables correlated with ability but unobserved by employers (such as certain test scores) are increasingly correlated with wages as experience increases. Similarly to us, Marmaros and Sacerdote (2006) study the interaction between first-year students proxying friendships through the volume of emails exchanged. They focus on the role of geographic proximity, common interests, majors and family background. They pay no attention to the effect of students learning about each others’ hidden attributes on the dynamics of social interactions which is instead our focus. A more recent study by Conti et al. (2013) shows that high ability students, i.e., those with higher IQ, are more popular among their peers

and documents a positive effect of popularity on future earnings.³ Our study differs from Conti et al. (2013) as we analyze whether high achievers become more popular around the time information about their ability is revealed and whether this increase in popularity is driven by interactions with other high achievers or by low ability students reaching out.

Our findings are also closely related to the literature on social peer pressure in educational settings (Austen-Smith & Fryer, 2005; Bursztyn et al., 2019; Bursztyn & Jensen, 2015, among many others). In particular, our findings support the “cool to be smart” hypothesis put forward in Bursztyn et al. (2019). In a large field experiment (Bursztyn et al., 2019) show that students behave in a way that is consistent with the hypothesis that signaling ability is “cool”. The existence of this reward is indirectly inferred through behavior which would be difficult to rationalize otherwise. In this paper we actually document how ability (or its perception as revealed through grades) is indeed rewarded by a greater social effort from low achievers.

Finally, segregation may have an effect on school outcomes of minorities (Ananat, 2011; Angrist & Lang, 2004; Cutler & Glaeser, 1997; Echenique & Fryer, 2007; Echenique et al., 2006) making appealing to re-engineer this process by exposing students of different racial backgrounds to each other. Therefore, studying the dynamics of group formation in contexts with heterogeneous agents is key for desegregation plans to be properly designed.

The paper proceeds as follows: Section 2 describes the data and discusses the representativeness of the sample used in the empirical analysis; Section 3 presents patterns of segregation in a static setting; Section 4 analyzes the dynamics of the social interactions among students in the same class, around the first term exams; Section 5 presents robustness checks for the main findings; and Section 6 concludes.

2. Data

We collect Instagram data for a sub-sample of students enrolled in 2015 in Economics or Business at the University of Naples Federico II.

Instagram is a social networking service built around posting photos. It launched in October 2010 on iOS first, and became available on Android in April 2012. Once posted, a photo is shared with one’s friend or made “public”, depending on one’s privacy settings. Like most social media apps, Instagram allows you to follow users that you are interested in. This creates a feed on the homepage showing recent posts from everyone you follow. You can like posts and comment on them. For each student that has a public Instagram profile we observe all posts, likes and comments received by other students in the sample or from non-students, and likes and comments sent to other students in the sample.^{4,5} A key feature of our data is that we can observe the “direction” of interactions — i.e., the identity of the sender and of the receiver. Social interactions data for students enrolled in 2015 spans the period September 2014 to December 2017, but in most of the empirical analysis we focus on interactions occurring during the academic year 2015–2016.

We merge this data with administrative records on students’ undergraduate performance, academic curriculum and gender. Italian public schools are subdivided in a number of broad categories (e.g., classical studies, professional studies, scientific studies). Students attending different schools belonging to the same category share a common

³ Fletcher (2014) shows that the popularity premium disappears when considering comparisons among siblings.

⁴ We link Instagram accounts to individual students exploiting a feature of Instagram that allows to retrieve any profile associated to a given email address. Students registered to the service with an email address different than the one supplied to the University’s registrar office are missing.

⁵ Students and Instagram accounts have been linked in the spring of 2018, right at the end of our sampling period. Users that canceled their account in the meantime are thus not recorded as users.

² As we do not directly observe new registrations to the platform, we proxy for them using the first time a student posts anything on her page.

curriculum. Within the same school, students of the same cohort are grouped into classes of up to 30 students with whom they spend most of the day for the whole 5 year program. At the end of the high school program, students take a battery of final exams. Performance is summarized by a grade reflecting both their performance in the finals and their previous high school academic record.

Table 1 in the Appendix presents summary statistics on interactions over the period September 2014–December 2017. The three main variables are the number of posts, likes and comments. As posting requires more effort than commenting and commenting requires more effort than liking, individuals mainly interact by publicly “liking” others’ posts. Posting new photos and commenting others’ posts are much less frequent: on average students in our sample post 2 times per month, with each post receiving about 37 likes and 3 comments. The total average number of likes received in a month is 148, whereas the total number of comments received in a month on average is only 10.

As one would expect, the fraction of likes received by other students increases over time, suggesting that students-to-students interactions account for an increasingly larger fraction of the total number of interactions (see Fig. 2 in the Appendix).

2.1. How does our sample compare to “non-instagram” students?

We compared, along several characteristics, the sample of 280 students for whom we have information on their social interactions on Instagram, with the 627 who do not have an Instagram profile and with the 117 that have a private one. One might imagine that students without an Instagram profile and those with private profiles represent opposite extremes: of no engagement at all; and of very personal and meaningful posts. Hence, the sample of students that have a public profile (those that populate our estimation sample) would not be representative of the population of interest. To test whether this is the case, we regressed each observed student characteristic, namely their high school graduation mark, gender, academic major, type of high school (scientific, vocational or classic) and academic performance onto a constant and indicators for private and public profiles. We found a statistically significant difference in high school grade, gender and type of high school. When considering all the characteristics together, the p -value of a test for the joint significance of “Public profile” and “Private profile” is 0.0255. Specifically, students with public profiles have a slightly lower average high school grade than those with no profile at all or with a private one. The share of males (and scientific diploma holders) in the “Public profile” sample is higher than in the “No Instagram” and the “Private profile” samples. The three groups do not differ significantly in terms of their majors (economics or business) and the students’ distribution in classes defined by last name initial and major does not differ either. Finally and most importantly, the fraction of high achievers (to be defined in the next section) does not differ between groups. A table with the estimated coefficients is relegated in the appendix (Table 2).

3. Do students segregate into separate social networks?

Data on students with public Instagram profiles from the academic year 2015–2016 are used in a pooled cross-section regression analysis. For each pair of students, we construct measures of social interaction which we then regress on variables derived from administrative records (see Table 1). “Any like” (columns 1 and 2) is a dummy equal to 1 if at least one like is exchanged within the pair during that month. Columns 3 and 4 instead use the total number of likes exchanged within a pair. FF is equal to one if both students in the pair are female, and zero otherwise, and MM is an indicator for pairs formed by two male students. Therefore the coefficients of FF and MM have to be interpreted as deviations from the benchmark case in which the pair is formed by a female and a male student (FM). We control for the

Table 1
Segregation.

	Any like		No. of likes	
	Linear (1)	Logit (2)	Linear (3)	Poisson (4)
GPA difference	−0.0002*** (0.0000)	−0.0003*** (0.0001)	−0.0006*** (0.0001)	−0.0008*** (0.0002)
FF	0.0019*** (0.0005)	0.0014*** (0.0003)	0.0049*** (0.0016)	0.0037*** (0.0010)
MM	0.0001 (0.0002)	0.0001 (0.0003)	0.0007 (0.0008)	0.0007 (0.0010)
same_hs	0.0468*** (0.0056)	0.0067*** (0.0005)	0.1087*** (0.0192)	0.0152*** (0.0017)
same_hs_type	0.0006*** (0.0002)	0.0007*** (0.0002)	0.0015** (0.0007)	0.0019** (0.0008)
same_class	0.0046*** (0.0004)	0.0034*** (0.0003)	0.0095*** (0.0014)	0.0072*** (0.0010)
No. of posts	0.0005*** (0.0000)	0.0003*** (0.0000)	0.0014*** (0.0002)	0.0007*** (0.0001)
ymean	0.0019		0.0042	
ysd	0.0439		0.1300	
N	643,753	643,753	643,753	643,753
N_clust	58,523			

Note: *GPA difference* is the absolute value of the difference between students’ GPA in a pair; *FF* indicates that both students are female; *MM* indicates that both students are male. *same_hs* and *same_hs_type* indicate that the two students in the pair graduated from the same high school or from the same type of school; *same_class* indicates that both students in the pair are in the same university class; *No. of posts* is the number of posts published by either student in the pair in a given month. *Any like* is a dummy that takes value equal to one if the pair exchanges at least one like in a given month; *No. of likes* is the number of likes exchanged in a month within the pair. All specifications include month dummies. Standard errors clustered at the pair level in parentheses. The coefficients reported in columns (2) and (4) are the estimated average marginal effects. * indicates $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

overall level of activity of the pair (No. of posts) and for month fixed effects. Results are presented in Table 1.

Using a linear probability model (column 1) we find that FF interactions are more likely than interactions in MM and FM pairs. The number of interactions (column 3) is highly positively correlated with variables reflecting some common background. In particular, a common high school (*same_hs*), being in the same class at the university (*same_class*) or merely sharing the high school curriculum (*same_hs_type*).

Importantly, students with similar ability tend to interact more: the coefficient of the absolute value of the difference between students’ GPA in a pair (*GPA difference*) is negative and significantly different from zero, meaning that, the likelihood of students’ interaction within a pair is lower the higher is the difference in students’ performance. Specifically, a one standard deviation change in *GPA difference* (the standard deviation of this variable is 2.31) is associated with a 24 percent reduction in the likelihood of an interaction and a 33 percent reduction in the number of likes exchanged in a pair.

Because of the very small probabilities and counts of posts and likes, and large proportion of zeroes, we also present marginal effects from Logit (column 2) and Poisson (column 4) models. The results are unchanged.

4. Interaction dynamics

We now turn to study the dynamics of social interactions, exploiting the fact that our measure is directional. While in the previous section we aggregated interactions at the pair level, here we use separate measures according to the identity of the sender and the receiver, i.e., our unit of observation is now a pair ij , where i is the receiver and j is the sender. This allows us to analyze the dynamics leading to a friendship, rather than just the final outcome of such process, and identify whether the segregation along the ability dimension (as proxied by academic achievement) documented in the previous section results only from active segregation or also from passive segregation. Specifically, we

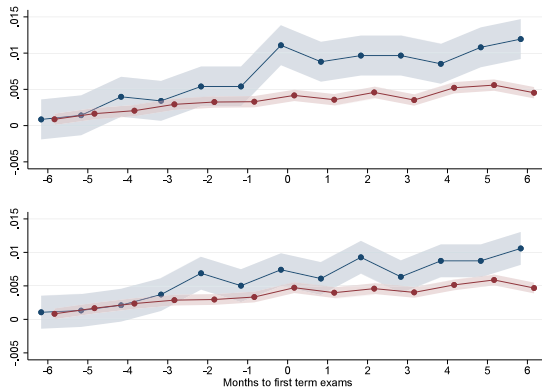


Fig. 1. Popularity of high and low achievers over time. Note: The top panel shows the probability of *receiving* a like for high achievers (blue) and low achievers (red). The bottom panel shows the evolution over time of the probability of *sending* a like for high achievers (blue) and other students (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

test whether “low achievers” reach out to ‘high achievers’, while high achievers mainly interact with other high achievers.

To this purpose a student is considered a high achiever if he passes all first term exams with average grade greater or equal to the 75th percentile (in our sample this translates into a GPA of at least ≥ 25 over 30). Other students are students that either do not pass all first term exams or pass them but with an average between 18 (which is the minimum grade) and 25. The large majority of students that fall in the low achievers category did not pass all three exams. Fig. 3 in the Appendix shows the distribution of average first-term grades among students that passed one exam, two exams and all three first-term exams.

The top panel of Fig. 1 shows the likelihood of *receiving* a like for high achievers (blue line) and low achievers (red line). Time is normalized relative to first term exams (0 is March 2016, -6 is September 2015, namely the month of enrollment). A first interesting finding is that high achievers become gradually more popular than low achievers over time. This finding is consistent with the hypothesis that sorting is driven by an underlying process of information revelation. As students learn about each others’ ability over time, they sort on that dimension.

To provide further support for the hypothesis that ability is the main driver of the observed evolution in friendship formation we exploit the release of public information that occurs in connection with first term exams at time $t = 0$. The figure documents a concurrent visible jump in the popularity of high types (but not of low types).

One possible explanation, that we can rule out immediately, is that high achievers simply interact more between themselves after the exams, perhaps because they have more time available. The bottom panel of Fig. 1 shows the evolution from -6 to +6 of the probability of *sending* a like for high achievers (blue line) and other students (red line): the increase in the probability of receiving a like by high achievers is not accompanied by an equal increase in the probability of sending a like. Indeed, high achievers do not start sending more likes after the exams (relative to other students).

Furthermore, the number of posts from high achievers does not differ from low achievers around the first-term exams period, nor there are significantly increasing trends (see Fig. 4 in the Appendix). However, in March 2016 there is a spike in the number of posts published by high achievers. Hence, a possible concern is that the relative increase in the popularity of high achievers after the first-term exams may be driven by dynamic effects related to the number of posts. Indeed, one may think that the number of posts at time t influences the number of likes received thereafter. To address this issue we estimate a dynamic

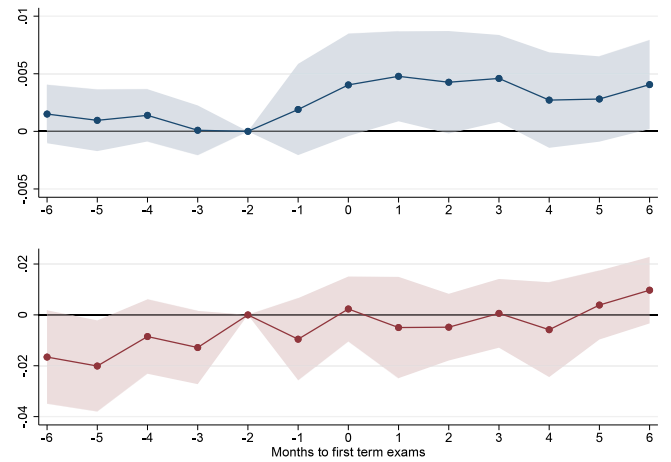


Fig. 2. Popularity of high achievers relative to low achievers, by sender's type. Note: the figure reports the changes in the estimated probability that a high achiever receives a like, relative to a low achiever, respectively by a low achiever sender (top panel), and a high achieving sender (bottom panel). The top panel of the figure reports the estimated γ_i coefficients and the 95% confidence intervals from Eq. (1), restricting the sample to low-achieving senders. The bottom panel of the figure reports the estimated γ_i coefficients and the 95% confidence intervals from Eq. (1), restricting the sample to high-achieving senders. In both panels the receiver can be either a high or a low achiever.

model relating the number of likes at time t with 2 leads and 2 lags of the number of posts. The results, reported in Fig. 5 in the Appendix, show that only the current number of posts correlates positively and significantly with the number of likes received in each month.

To understand what drives the increase in popularity of high types relative to low types we look at dynamics within and across high and low achievers’ groups and ask whether information released after first-term exams affects the number of likes that low achievers direct towards high achievers. Specifically, we restrict the sample to low-achiever senders and to pairs in the same university class and test whether, following first-term exams, the likelihood of receiving a like from a low achiever in period t increases significantly for high achievers relative to low achievers, estimating the following specification:

$$Y_{ijt} = \alpha_{ij} + \lambda_t + \gamma_i H_i + \beta Nposts_{it} + \epsilon_{ijt} \quad (1)$$

where the receiver, indexed by i can be either a high or a low achiever (with $H_i = 1$ if high), while the sender, indexed by j is a low achiever. This specification controls for any unobserved heterogeneity at the pair level (pairs fixed effects, α_{ij}); time patterns in probability of receiving a like (time fixed effects, λ_t); and, to account for possible changes of the overall level of online activity of the receiver, the number of posts that receivers publish on their Instagram page ($Nposts_{it}$). γ_i are the coefficients of interest that capture the differential change, between time t and time $t = -2$, in the probability that a low achiever sends a like to a high achiever, relative to the change in probability that a low achiever sends a like to another low achiever.

Until time $t = -1$ there is no significant difference in trends between high and low achievers in the probability of receiving a like from a low achiever (top panel of Fig. 2). This shows that the two groups are not on diverging trends before the information revelation that occurs during the first term exams.

Crucially, the γ_i coefficients become positive and significantly different from zero right after the first term exams, suggesting that high achievers become relatively more likely to be “liked” by low achievers after the exams (top panel of Fig. 2). The coefficient of 0.005 means that for high achievers the probability of receiving a like from a low achiever increases by 50 basis points more than the probability of a low achiever receiving a like from another low achiever. This effect is both statistically and economically significant, considering that the

likelihood that a pair interacts at least once in a month is five times lower (0.001).

One possible concern is that this result is consistent with low types changing their behavior after first-term exams, for instance by attending classes more often and thus intensifying their online and offline interactions with high achievers. To address this issue, we repeat the exercise focusing on high-achieving senders: if the relative increase in low-to-high interactions is driven simply by low types going to the university more often, one would expect a similar change in the likelihood of high-to-low interactions. In the bottom panel of Fig. 2 we report the coefficients of a version of Eq. (1) estimated using only the sample of high-achieving senders (j is restricted to be a high achiever, while the receiver i can be either a high or a low achiever).

We find that the likelihood of high-to-high interactions (relative to that of high-to-low) does not change after first-term exams, rather the probability of receiving a like from high-achieving senders is higher for other high achievers, and it increases over time, as witnessed by the fact that the estimated coefficient of the interaction between our measure of high achievement and a linear time trend is positive and significantly different from zero (the point estimate is 0.002 and the standard error is about 0.001). One possible explanation is that high achievers interact more with other high achievers than with low achievers from the beginning of the academic year, because they have common traits that correlate with their academic success. For instance, high achievers may attend more lectures and spend more time at the university than students that eventually do not perform well in exams. As friendship among high achievers develops over time, students also update their beliefs about their friends' ability so that the exams are relatively more revealing for low types than for high types.

Another potential concern with this analysis is that peer effects themselves may determine who becomes a high or a low achiever. However, this is not an issue because our definition of low/high achievers relies only on first-term exams, thus being predetermined with respect to behavioral changes occurring after them. Even if low achievers are such because of their peer group, our result can still be interpreted as evidence that they reach out to another peer group hoping to improve their future performance.

Finally, as our measure of ability is to some extent an outcome variable, it raises concerns on the reflection issue: students' Instagram use habits and interactions with others can affect their exam performance. In the appendix (Fig. 6) we report γ_t coefficients estimated using high school grades as the ability indicator⁶ and find similar results, although less precise.

We interpret this as suggestive evidence that homophily is not the only driver of segregation: all types eventually increase their social effort towards high achievers, consistently with the hypothesis that individuals react also to material incentives.

5. Robustness checks

In the previous section we showed that high achievers receive disproportionately more likes from low achievers after the first term exams. We interpret this finding as evidence in support of the hypothesis that, upon the revelation of information about students' ability triggered by the first term exams, low achievers reach out to high achievers in order to join their network, possibly to obtain returns later on in their academic career.

In this section we provide robustness tests in support of both our definition of high achievers and the interpretation of our main result. As for the definition of high achievers, first, we exploit the availability in our data of both high-school final graduation mark and GPA in the first semester, and test whether the effect of high school graduation

⁶ More specifically, we define a high achiever as a student that attained the highest possible grade in high school (i.e., 100).

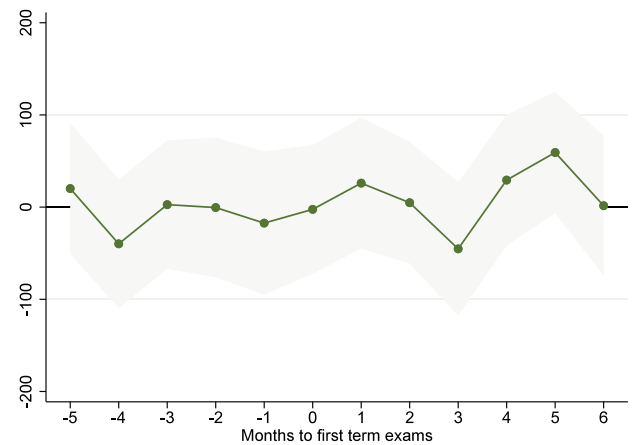


Fig. 3. Likes from non-students to high achievers. Note: point estimates and 95% confidence intervals of changes in the monthly number of likes received by high achievers from non-students from the month of enrollment (−6) to month t relative to first term exams.

mark on popularity fades away once information on GPA is revealed, in the spirit of the “employer learning” model in Altonji and Pierret (2001) and Farber and Gibbons (1996). Specifically, we estimate the following model:

$$TotLikes_{it} = \alpha_i + \beta_1 H_i \times H_i^{HS} \times Post_t + \beta_2 H_i \times Post_t + \beta_3 H_i^{HS} \times Post_t + \beta_4 Post_t + \beta_5 Nposts_{it} + \eta_{it} \quad (2)$$

where $TotLikes$ is the total number of likes received by individual i in month t , α_i are individual fixed effects, H_i is an indicator for high-achieving student (as defined in the previous sections), H_i^{HS} is a dummy equal to one if the high school graduation mark is 100/100, $Post$ is an indicator for the period following first-term exams, and $Nposts$ is the number of posts of individual i in month t . Interestingly, the estimate of β_1 is 1.88 with a robust standard error of 0.98, the point estimate of β_2 is 0.48, with a robust standard error of 0.23, whereas the coefficient of the interaction term between high-school achievement and post is not statistically different from zero ($\hat{\beta}_3$ is 0.38 with robust standard error of 0.31). These results show that, while the popularity of high achievers at university increases over time, that of students that performed well in high school (but do not end up being high achievers in the first-term exams) does not change over time.

Second, we perform sensitivity analysis changing the definition of high achievers. Specifically, we build an adjusted measure of average GPA, that accounts for the fraction of exams actually taken by the students in the first term. Namely, we multiply the first-term weighted average GPA by the fraction of credits obtained over the total achievable number of credits (30) in the first term. Then, we define high achievers as those with adjusted GPA higher than the 75th percentile (column 1 in Table 3 in the Appendix) or the median (column 2), and estimate difference-in-differences models to evaluate the relative change in the probability of receiving a like (from a low achiever) between high- and low-achieving receivers, before and after first-term exams. The results, reported in Table 3 in the Appendix, show that the probability of receiving a like from a low achiever significantly increases for high achievers relative to low achievers, following first-term exams. Furthermore, in Column (2) we perform an additional robustness test, by excluding receivers with adjusted GPA between the 25th and 75th percentiles; results are unchanged.

One potential concern is that, following the first term exams, high achievers have more free time, they can dedicate more time to social relations and in turn they may become more popular also online. To address this threat to identification we take advantage of the fact that we observe all the likes received by the students in the sample on

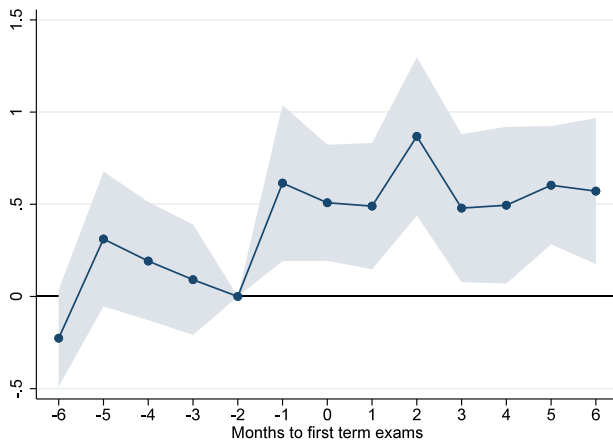


Fig. 4. Analysis at the post level. Note: point estimates and 95% confidence intervals of changes in the monthly number of likes received by high achievers' posts from low achievers from the month of enrollment (−6) to month t relative to first term exams.

their Instagram page (not only those that involve other students) and analyze the dynamics of the likes received by high achievers from non-students. In Fig. 3 we show that there are no significant changes in the number of likes received by high achievers from non-students from the month of enrollment to the months immediately following the first term exams.

As an additional test to support our interpretation of the results shown in the previous section, we analyze the evolution of the likes received by high and low achievers from low achievers *that are not in the same class*. The rationale for this analysis is that first term exams should trigger a lower degree of information revelation to students in other classes. Indeed we find that the dynamics of likes sent from low achievers to students in other classes do not differ between high- and low-achieving receivers (see Fig. 7 in the Appendix).

Another potential concern with the analysis reported so far is that students who are not posting in a month clearly have zero chance to receive likes in that given month. Controlling for the number of posts in a period may not fully account for this constraint. To address this concern we repeat the analysis using as unit of observation each post (rather than each pair). More specifically we estimate the following model:

$$\text{LikesFromLow}_{pt} = \alpha H_p + \lambda_t + \gamma_t H_p + \epsilon_{pt}$$

where LikesFromLow_{pt} is the number of likes given by low achievers to a post p published in month t ; H_p is a dummy equal to one if the post is published by a high achiever, and λ_t are time dummies. The series of coefficients γ_t thus capture the change in the number of likes sent by low achievers to posts published by high achievers relative to those sent to posts published by other low achievers. Fig. 4 reports the estimated γ_t and confirms the results presented in the main analysis: around the first term exams, the popularity of high achievers among low achievers increases more than that of other low achievers.

6. Conclusions

This paper exploits panel data on social interactions taking place on Instagram among a cohort of undergraduate students to analyze whether individuals' effort in developing social ties responds to material incentives. Specifically, we study whether low-achieving students reach out to high-achieving ones.

The main finding is a behavioral asymmetry between types that may shed light on the drivers of segregation. Exploiting the fact that

our data record the direction of social effort, we document that low types do indeed reach out to high types. This finding is consistent with segregation being an equilibrium phenomenon as in search theoretic models of the labor market (Shimer & Smith, 2000).

This study has two main limitations. First, our findings hinge on the hypothesis that online interactions reflect offline ones. That is the hypothesis that the likelihood of interacting offline increases with online interactions. Second, because of the very limited number of interactions among students, our estimates are imprecise. This may be due to the fact that by the time students enroll at the university their social networks are already well established, so that it may take more than just the first year of enrollment to significantly increase the interactions with other students.

CRedit authorship contribution statement

Emilio Calvano: Conceptualization, Funding acquisition, Methodology, Writing – original draft. **Giovanni Immordino:** Conceptualization, Methodology, Writing – original draft. **Annalisa Scognamiglio:** Conceptualization, Methodology, Formal analysis, Writing – original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.econedurev.2022.102290>.

References

- Altonji, J. G., & Pierret, C. R. (2001). Employer learning and statistical discrimination. *Quarterly Journal of Economics*, 116(1), 313–350.
- Ananat, E. O. (2011). The wrong side (s) of the tracks: The causal effects of racial segregation on urban poverty and inequality. *American Economic Journal: Applied Economics*, 3(2), 34–66.
- Angrist, J. D., & Lang, K. (2004). Does school integration generate peer effects? Evidence from boston's metco program. *American Economic Review*, 94(5), 1613–1634.
- Austen-Smith, D., & Fryer, R. G., Jr. (2005). An economic analysis of "acting white". *Quarterly Journal of Economics*, 120(2), 551–583.
- Bramoullé, Y., Currarini, S., Jackson, M. O., Pin, P., & Rogers, B. W. (2012). Homophily and long-run integration in social networks. *Journal of Economic Theory*, 147(5), 1754–1786.
- Bursztn, L., Egorov, G., & Jensen, R. (2019). Cool to be smart or smart to be cool? Understanding peer pressure in education. *Review of Economic Studies*, 86(4), 1487–1526.
- Bursztn, L., & Jensen, R. (2015). How does peer pressure affect educational investments? *Quarterly Journal of Economics*, 130(3), 1329–1367.
- Carrell, S. E., Sacerdote, B. I., & West, J. E. (2013). From natural variation to optimal policy? The importance of endogenous peer group formation. *Econometrica*, 81(3), 855–882.
- Conti, G., Galeotti, A., Müller, G., & Pudney, S. (2013). Popularity. *Journal of Human Resources*, 48(4), 1072–1094.
- Currarini, S., Jackson, M., & Pin, P. (2009). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, 77(4), 1003–1045.
- Currarini, S., Jackson, M. O., & Pin, P. (2010). Identifying the roles of race-based choice and chance in high school friendship network formation. *Proceedings of the National Academy of Sciences*, 107(11), 4857–4861.
- Cutler, D. M., & Glaeser, E. L. (1997). Are ghettos good or bad? *Quarterly Journal of Economics*, 112(3), 827–872.
- De Giorgi, G., Pellizzari, M., & Woolston, W. G. (2012). Class size and class heterogeneity. *Journal of the European Economic Association*, 10(4), 795–830.
- Echenique, F., & Fryer, R. G., Jr. (2007). A measure of segregation based on social interactions. *Quarterly Journal of Economics*, 122(2), 441–485.

- Echenique, F., Fryer, R. G., Jr., & Kaufman, A. (2006). Is school segregation good or bad? *American Economic Review*, 96(2), 265–269.
- Farber, H. S., & Gibbons, R. (1996). Learning and wage dynamics. *Quarterly Journal of Economics*, 111(4), 1007–1047.
- Fletcher, J. (2014). Friends or family? Revisiting the effects of high school popularity on adult earnings. *Applied Economics*, 46(20), 2408–2417.
- Marmaros, D., & Sacerdote, B. (2006). How do friendships form? *Quarterly Journal of Economics*, 121(1), 79–119.
- Mele, A. (2017). A structural model of dense network formation. *Econometrica*, 85(3), 825–850.
- Shimer, R., & Smith, L. (2000). Assortative matching and search. *Econometrica*, 68(2), 343–369.
- Tarbush, B., & Teytelboym, A. (2017). Social groups and social network formation. *Games and Economic Behavior*, 103(286–312).