Socio-Economic Planning Sciences 73 (2021) 100865

Contents lists available at ScienceDirect

Socio-Economic Planning Sciences

journal homepage: http://www.elsevier.com/locate/seps

Does joining social media groups help to reduce students' dropout within the first university year?

Lucio Masserini [a,*], Matilde Bini [b]

[a] Department of Economics and Management, University of Pisa, Pisa, Italy
[b] Department of Human Sciences, European University of Rome, Rome, Italy

ARTICLE INFO

Keywords:
First year students' dropout
Selection bias
Propensity score matching
Social networks
Average treatment effect

ABSTRACT

Using observational data, the aim of our study paper was to investigate whether university students' dropout within the first year is influenced by participation in social media groups such as Facebook pages created and run by other students. Specifically, in this paper such participation is considered as a treatment and represents a means to help promote and strengthen social relationships amongst students but also to help share information on courses and other material useful for studying and preparing for exams. For this purpose, data from a sample survey of students enrolled in a major Italian university were used. Given a non-random treatment assignment, analysis was carried out using propensity score matching (PSM) in order to correct for selection bias due to a set of observable pretreatment covariates. Several matching techniques and sensitivity analyses suggested that the results were robust for estimating an average treatment effect on the treated group. The estimated effect indicated that participation in social media groups is effective for lowering the dropout rate.

1. Introduction

University dropout is one of the most serious issues that can occur in a student's university career, particularly during the first year of studies [1]. Over the last decades, this issue has become quite relevant, especially in those countries where admission to university was opened to all upper secondary school graduates and attending university shifted from elitism to mass opportunity. In particular, open access policies and absence of selective admission procedures has encouraged university enrolment of students with very different social and cultural backgrounds, but also with diverse life experiences and varying levels of preparation and motivation. These differences inevitably have resulted in a great variety of needs, expectations, opportunities or academic potential [2] and have determined varying levels of academic success. Indeed, less motivated students or those who do not meet the academic prerequisites to cope successfully with university studies are more likely to have difficulties during their university career and consequently are more susceptible to dropout.

The issue of withdrawal from studies has important effects on universities, in particular for the social implications that this entails for students and their families, but also because part of public funding is now allocated based on the students' career progression. In Italy, for instance, the withdrawal rate after the first year is about 20% on average. As a consequence, universities are committed to collect data in order to understand the factors that can influence delays in university careers and the reasons that may lead to dropping out of the university in order to design policies that can help to reduce these problems. The literature has put forth a variety of studies that dealt with the determinants of dropping out, particularly within the first year. They highlighted that the most important factors affecting the decision to withdraw are individual characteristics, social and economic conditions and status condition at the beginning of the career, but also the quality of teaching and organizational aspects of the institutions. In the last decade, social networking websites have become a nearly indispensable tool for the new generation of students to be more strongly connected with colleagues and with activities pertaining to university life [3–6]. The most popular social network sites are Twitter and Facebook which are both used as educational tools [7–9]. Although they seem to be unusual platforms for academia, young people (especially those from 16 to 24 years of age), who have grown up with social media, spend two or more hours on these media platforms each day. Universities in ever more competitive contexts have grasped the importance of staying connected with students not only on an academic level, but also on a personal level, and they invest time and resources engaging them with the aim to keep

* Corresponding author.
E-mail addresses: lucio.masserini@unipi.it (L. Masserini), matilde.bini@unier.it (M. Bini).

https://doi.org/10.1016/j.seps.2020.100865
Received 15 November 2019; Received in revised form 3 April 2020; Accepted 6 May 2020
Available online 20 May 2020
0038-0121/

them 'on-board'.

Some recent studies [3–6] have the highlighted the advantages that students receive from joining a social media group rather than not participating, such as the greater perception of belonging to the university and increasing the peer relationships with class or staff [10], and they have conceptualized student engagement in two dimensions: campus engagement and class engagement (see also [11]). Today, Facebook and Twitter have become part of students' lives and now almost all students use these social media [11].

Despite the massive use of Facebook amongst university students and the considerable number of institutions using social media to reach and connect with students [12], very little empirical evidence is available about the use of Facebook on student academic performance and success. To date, some of these studies have shown the role of social media as a means to strengthen engagement and integration amongst students [13].

The aim of this study was to provide some empirical evidence from the Italian university system by examining whether students' decision to drop out within the first year can be affected by the participation in groups or Facebook pages run by other students, considered as a 'treatment' condition for the purpose of this study. The analysis was performed using data from a sample survey carried out by interviewing 1879 freshmen enrolled at a major university located in central Italy in 2016. However, given the non-random treatment assignment, students participating in groups or Facebook pages may differ from those who do not participate. In particular, the choice of joining groups or Facebook pages may reflect different students' characteristics, such as enthusiasm for their college experience and social interaction with peers, but also their level of preparation, academic motivation, living conditions and student engagement, amongst other factors. To the extent that all or part of these factors simultaneously affect selection into treatment and the students' dropout decision, the estimation of the treatment effect may be seriously flawed, as well as the implications arising from the study. To address the problem of potential selection bias, propensity score matching (PSM; [14]) under the potential outcomes framework of causality was employed.

The paper is organized as follows. Section 2 presents a brief literature review, Section 3 illustrates methods and materials, Section 4 describes the results and, finally, Section 5 is devoted to some concluding remarks.

2. A brief literature review

In the higher education literature, several studies based on surveys of various aspects of students' careers with particular reference to the problem of non-completion rate revealed that there are different determinants that lead to interruption of their studies, especially within the first year. For example [15], identified the social sphere and personal and economic issues as the most important dimensions of persistence. Even earlier [16,17], using a path analysis model, identified family background, personal characteristics, pre-schooling and student interactions with colleagues and teachers as factors. Family background and occupation opportunities after graduation were found by Refs. [18]. Other determinants can affect the decision on whether to continue to study, such as university characteristics, teaching quality, the individual's personal characteristics [18], citizenship and income [19]. Other studies [20–22] that refer to the UK context identified parents' social class, type of school and prior educational achievement as causes of withdrawal. The list of factors or potential causes which have been associated with university dropout and to the related concepts of retention and persistence also included interaction with peers [23], the individual's prior performance relative to that of peers [24], local unemployment rate [18,25], student engagement [26], parental educational background [27], student age and marital status [18], student high school average grade [28], measures of student qualification and motivation [15] and academic life conditions. Other authors focused on causal links amongst student background, educational and institutional commitment and academic and social integration. They concentrated on the impact of specific factors on retention, such as student ethnicity and gender [29,30], classroom-based learning experiences [31,32], institutional support services [33], intention to leave [34], academic and social integration [35], pre-collegiate academic preparation [36] and students' perceptions of their own abilities [37].

Social media sites such as Facebook provide students with a tool to present themselves in an online profile, collect new friends, meet new people, post comments and share pictures, opinions and habits, as well as giving them the ability to see each other's pages and profiles. Already by 2010, Roblyer [38] highlighted that Facebook had more than 350 million subscribers around the world and had become the largest site of different communities of users in contexts such as societies, education, business and universities.

By 2012, the number of worldwide registered users of Facebook reached one billion. Now, with 2.41 billion monthly users, Facebook is by far the largest social network worldwide. Recent studies [3,6,9] have indicated that Facebook is the most popular social medium in universities, used as an educational tool not only by students, but also by other stakeholders such as teachers, researchers and non-academic staff (e.g. Refs. [7,8,39]).

Although the use of social media such as Facebook is high, some studies initially done on social media [40–43] highlighted that the use of these platforms was different amongst gender, ethnic and socioeconomic conditions. Hargittai [44] indicated that this is particularly the case in some American and Latino countries, where students whose parents have a college degree are more likely to use Facebook than students whose parents do not have a college degree. Other reasons [45], such as cultural resistance, pedagogical issues or institutional constraints or the type of teaching discipline, can affect attitudes towards digital media and their expectations. Other studies revealed that different attitudes amongst students and teachers, in particular, pedagogical issues [46], or extrinsic factors (e.g. time, training and support), rather than intrinsic factors (e.g. beliefs, motivation and confidence), were the main barriers to faculty using these tools more frequently in education [47].

The number of settings in which Facebook is used is significant and ranges from communications of events, content of courses or programmes of instruction and library promotion in order to increase their presence in online courses with links to services [38]. One mission of Facebook is to increasingly strengthen student engagement, particularly in relationships and communications student–student and staff–student (e.g. Refs. [7,8,39,48,49]) and to increase the sense of belonging to a classroom community (e.g. Refs. [7,39,49–51]). Other studies [52–55] have revealed that the use of Facebook improved some aspects of students' careers, such as their grades, motivation, self-esteem, intention to persist and satisfaction.

Although there exists a recent and increasing number of scientific studies that deal with how Facebook has affected the social lives and relationships of individuals and institutions, as well as student engagement and sense of belonging to institutions, very little empirical evidence is available about the use of Facebook (or any other social network) on success and academic performance, the important aspects connected with the dropout problem. The causal relationship amongst the aforementioned aspects, the role of Facebook as a treatment that could affect aspects students' engagement and their overall academic experience and their decision about whether to withdraw from their university was the primary motivation for this study.

3. Method and materials

The study was carried out with the proposal of a method based on a PSM under the potential outcomes framework of causality, described in Subsection 3.1. It used data collected from a survey conducted by interviewing a simple random sample of freshmen enrolled at a major university located in central Italy in 2016. The description of the survey and variables used in the model are described in Subsections 3.2 and 3.3.

3.1. Statistical approach

To measure the impact of participation in groups or Facebook pages created and managed by university students with the aim of sharing information on degree courses or any material useful for studying and preparing for exams on dropout within the first year, the Average Treatment Effect for the Treated (ATET; [56]) was estimated. For this purpose, two variables were of primary interest: a treatment indicator variable, T_i, that takes the value 1 for those students participating in such groups or Facebook pages (also referred to as 'treated') and the value 0 for students who do not participate (or 'controls'), and an observed outcome variable, Y_i, which represents the dropout decision within the first year and takes the value 1 if the student drops out and the value 0 if the student does not drop out. Under this evaluation framework, the ATET is defined as the expected difference in the outcome variable between the groups of treated and controls amongst students who participated in such groups or Facebook pages ($T_i = 1$):

$$ATET = E(Y_{i1}|T_i = 1) - E(Y_{i0}|T_i = 1),$$

where $E(Y_{i1}|T_i = 1)$ is the mean value of dropout actually observed for those students participating in a group or a Facebook page, and $E(Y_{i0}|T_i = 1)$ is the counterfactual, i.e. the hypothetical mean value of dropout that would have been observed if the same treated students were not members of a group or a Facebook page. Clearly, the counterfactual is not observable since it is impossible to observe both outcomes for the same student at the same time. This is known as the 'fundamental problem of causal inference' [57], and because of this, a proper substitute for the counterfactual must be chosen in order to estimate the ATET. If the students' decision to participate is random, the ATET could be easily estimated by taking the mean outcome of non-participants as an approximation of the counterfactual and then comparing the mean value of dropout of students who participate (treated) with that of students who do not participate (controls). However, given that treatment 'assignment' is not random, the groups of treated students may differ systematically from controls, not only in terms of individual characteristics but also for their life experience and university career, amongst other factors. To the extent that all or some of these factors simultaneously affect the students' choice to participate in groups or Facebook pages (selection into treatment) and the decision to drop out of university, the estimation of treatment effect may be seriously biased, as well as the implications arising from the study. To overcome the fundamental evaluation problem and correct for potential selection bias, a PSM [14] was employed. In particular, PSM is based on the idea that selection bias can be reduced when the outcome variable is compared after matching treated and control students that are as similar as possible with respect to a set of pretreatment (or confounding) observed covariates, in essence approximating the condition of randomized trials [58]. However, since matching becomes impracticable as the number of pretreatment covariates increases (the 'curse of dimensionality'), the treated and control students are matched based on their propensity score (PS), which is an index that summarizes all pretreatment covariates for each student in a single variable.

The PS for student i is defined as the conditional probability of receiving a treatment (i.e. participating in such groups or Facebook pages), given a vector of pretreatment covariates $\mathbf{X}_i$[14]: $PS(\mathbf{X}_i) = Pr(T_i = 1|\mathbf{X}_i)$. Given that there are two treatment conditions, $T_i \in (0, 1)$, the PS is estimated using a logistic regression model. Since PS is a probability function, its score ranges from 0 to 1. If the $PS(\mathbf{X}_i)$ is known, then the ATET can be rewritten and estimated as follows [59]:

$$ATET_{PSM} = E[E[Y_{i1}|T_i = 1, PS(\mathbf{X}_i)] - [E(Y_{i0}|T_i = 0), PS(\mathbf{X}_i)]|T_i = 1]$$

PSM is based on two assumptions: conditional independence and common support [60]. In order to estimate unbiased treatment effects under PS, the following two assumptions are required [14]. The first assumption is conditional independence given the PS (also known as selection on observable):

$$(Y_{i0},\ Y_{i1}) \perp PS(\mathbf{X}_i),$$

which states that potential outcomes are independent of treatment assignment given the $PS(\mathbf{X}_i)$ and implies that selection into treatment is only based on observable characteristics. In practice, it requires that the covariates that simultaneously affect the choice of participating in groups or Facebook pages and the students' dropout decision are observed and included in the vector $\mathbf{X}$, so that, after controlling for $\mathbf{X}$, treatment assignment is 'as good as random' or 'strictly ignorable'. The second assumption is the common support condition (or overlap):

$$0 < PS(\mathbf{X}_i) < 1,$$

which states that each student must have a positive, but less than one probability of participating in the treatment and require some overlapping of the estimated PSs between treated and controls. It ensures that the treated and control groups share a common support region of PSs, i.e. 'treated' and 'untreated' students must be found for each range of values in the PS in order to obtain sufficient matches. In particular, for ATET, it is sufficient to ensure the existence of potential matches in the control group [61]. To match treated and controls on the PS, three different matching algorithms were employed: k-nearest-neighbour matching, caliper matching and kernel matching. For a detailed description of these matching algorithms, see Caliendo and Kopeinig [59].

3.2. The data

This study uses data from a survey carried out by interviewing a simple random sample of 1879 freshmen enrolled at a major university located in central Italy in 2016. The questionnaire used for the interview was divided into six sections concerning several aspects of the students' experience during the first university year: (A) Factors affecting the enrolment decision; (B) Student status during the first year; (C) University life and study-related experiences; (D) Life habits (including accommodation); (E) High school career; and (F) Social and academic integration, including use of social networks and social media. Participants were interviewed at the University's Computer Assisted Telephone Interviewing (CATI) laboratory by selected students who received training prior to conducting telephone surveys.

3.3. Variables

The outcome variable under study was student dropout decision within the first university year. It is a binary variable and took the value 1 if the student dropped out and the value 0 if the student did not drop out. Specifically, the value 1 included students who submitted a conventional waiver of studies and students who did not enrol in the second year who had not submitted a conventional waiver.

A rather large set of covariates was also available as pretreatment variables for the PS estimation. For greater clarity, these covariates were grouped according to the thematic areas of the questionnaire: (A) Factors affecting the enrolment decision: Decision to enrol influenced by the family environment, Choice of the university due to the presence of friends and Choice of the university due to a particular degree course; (B) Student status during first year: Occasional commuter (resident off-site, goes to the university only to take exams), Daily commuter (resident off-site, goes regularly to the university to attend lessons), Weekly commuter (resident off-site, but lives in the university city during the week), On-site (lives in the city where the university is located with his/her family) and Started working during the first year of studies; (C) University life and study-related experiences: Enrolment in the university changed the previous life habits, Attend classes during the first year (at least 75% of lessons), Do not attend classes, Attend professors' reception (at least once), Benefit from the tutoring service, Study regularly in the department's study rooms, and Study difficulties due to a greater commitment required compared to high schools; (D) Life habits (including accommodation): Live in institute together with other students and Live in apartment alone or with other students; and (E) High school career: Have friends already enrolled in the same degree course, Used to studying with other students, Have classmates to exchange lecture notes/information, Have contacts with classmates outside the university, Participation in extra-curricular activities (i.e. seminars and insights on specific topics), Participation in meetings and events organized by students outside of the study activity (i.e. film club, aperitifs, parties and organized tours). In addition to these, information on the students' enrolment department and gender were also available. Table A1 in the Appendix shows the descriptive statistics in terms of relative frequencies for each covariate for the whole sample and for the treated and control groups, as well as the *p* value for the difference between them. A comparison of the treated and control groups before matching revealed that 20 out of 52 covariates showed a significant difference. This meant that the students participating and those not participating in a group or a Facebook page presented very different baseline characteristics and that treatment assignment could not be considered ignorable. Therefore, under these conditions, evaluating treatment effects through the simple comparison of dropout rate between the two groups of students would yield biased estimates

4. Results

All PSM analyses were performed using the Stata's *psmatch2* program [62] and *mhbounds* program [63].

4.1. Estimation of PSM

Following the suggestions of [64,65], the PSM was estimated including only true confounders and predictors of the outcome in the vector of covariates (**X**). In fact, incorporation of true confounders helped to decrease the bias and the variance of the treatment effect estimates. By contrast, predictors of the outcome that were unrelated to treatment assignment reduced the variance of the treatment effect estimates but did not contribute to reduction of bias [66]. On the basis of these recommendations, the selection of covariates was made following the different relationships amongst the treatment, outcome and respective predictors. In particular, the factors influencing the outcome (dropout) were selected based on the existing literature, limited to those available in our dataset, such as the type of school, prior educational achievement of student by high school average grade and gender. Also, such covariates came from the same source (i.e. the same questionnaire), and this made the PS estimates more trustworthy [67].

Table A2 in the Appendix reports the parameter estimates of the logistic regression model, taking some of the answers to the questionnaire and students' characteristics as exogenous explanatory variables. The *p*-value for overall significance was smaller than 0.0001, but McFadden's pseudoR^2 was rather low (0.097). This demonstrated that the chosen set of covariates helped to predict the probability of participating in groups or Facebook pages created and run by other students, even though only a subset of them showed a significant effect. However, as the purpose of PS estimation was to balance the observed distribution of covariates, not to estimate regression parameters or draw inferences about those parameters, all covariates had been maintained in the model, even those that were not significant.

The PS ranged from 0.274 to 0.981 with a mean of 0.826, and this showed that, in general, students were more likely to join groups or Facebook pages created and managed by university students with the aim of sharing information on degree courses or any material useful for studying and preparing for exams, as expected. Results indicated that male students had, on average, a lower probability of participation in groups or Facebook pages ($\beta = -0.260$; $p = 0.003$). This was also the case for those who had enrolled for one or more years after obtaining their high school diploma (variable 'delay in enrolment') ($\beta = -0.361$; $p = 0.001$). Moreover, some differences were also found based on the students' enrolment department: in particular, students from the Department of Physics were less inclined to participate ($\beta = -0.694$; $p = 0.010$), whereas those from the Department of Information Engineering ($\beta = 0.514$; $p = 0.020$) and the Department of Clinical and Experimental Medicine ($\beta = 0.557$; $p = 0.020$) were more inclined. In contrast, students who had started working during the first year had a lower probability of participating ($\beta = -0.233$; $p = 0.013$), as was expected, as engaging in a work activity reduced the possibility of maintaining contacts with the other students and, more generally, with the university. Finally, it is worth noting the positive relationship between participation and the variables related to academic integration and interaction with peers in our study, which were represented by 'choice of the university due to the presence of friends' ($\beta = 0.663$; $p = 0.034$), 'have friends already enrolled in the same degree course' ($\beta = 0.166$; $p = 0.062$), 'have classmates to exchange lecture notes/information' ($\beta = 0.285$; $p = 0.013$) and 'participation in meetings and events organized by students outside of the study activity (i.e. film club, aperitifs, parties and organized tours)' ($\beta = 0.230$; $p = 0.015$). This indicated that social and academic integration were strictly correlated with online or virtual relationships.

4.2. Evaluation of covariate balance and common support

To check for the balance of the distribution of relevant covariates in both the control and treatment groups after matching, standardized percentage bias (SB) [60] and percent of bias reduction (PBR) [14] were calculated for all covariates for each matching algorithm. These are shown in Table 1 as the summary indicators of covariate imbalance, calculated in terms of the mean and median percentage bias of the distribution of the entire vector of covariates. Typically, values higher than 10% represented a meaningful imbalance between the control and treatment groups..

In the first row of Table 1, the values of mean and median percentage bias refer to the unmatched sample, whereas, in the other rows, they refer to the different algorithms and also include the corresponding values of PBR obtained after matching. In particular, the values for the unmatched sample indicated that treatment assignment could not be considered ignorable because the two groups of students presented very different baseline characteristics, with mean SB of 11.0 and median SB of 9.1. In contrast, after matching, the bias was reduced significantly for each algorithm, with a PBR reduction ranging from −46.5 to −56.7 for the nearest-neighbour, from −61.6 to −63.0 for calliper and from −50.7

Table 1
Sample size, mean, median standardized bias (SB) and percent of bias reduction (PBR) across all covariates in the unmatched and matched samples for each matching algorithm.

Matching technique	Sample size	Treated	Controls	Mean SB(%)	Median SB(%)	PBR (%)
Unmatched sample	1655	1373	282	**11.0**	**9.1**	-
Nearest-neighbour	1640	1358	282	4.7	4.1	−46.5
1 nearest-neighbour						
3 nearest-neighbour	1640	1358	282	4.1	3.3	−54.3
5 nearest-neighbour	1640	1358	282	4.0	3.6	−56.7
Radius						
radii = 0.01	1640	1358	282	3.4	3.2	−61.6
radii = 0.02	1640	1358	282	3.3	3.1	−63.0
radii = 0.04	1640	1358	282	3.1	2.3	−63.0
Kernel						
Epanechnikov	1655	1373	282	3.2	2.6	−59.8
Normal	1655	1373	282	3.9	3.2	−50.7
Biweight	1655	1373	282	3.2	2.8	−61.1

to −61.1 for kernel matching. This demonstrated that PSM helped to remove treatment selection bias and that it appeared to be effective in forming two balanced groups of students for the ATET estimation.

Moreover, since a necessary condition for the use of PSM is the existence of a sufficiently broad common support region between treatment and control groups, following suggestions of [68], this was evaluated by visual inspection of the histograms in Fig. 1 which show the distribution in quintiles of the PSs in both groups. The figure shows that there was a good overlap in the PS distributions between the groups, and this ensured that any combination of the students' characteristics observed in the treatment group could also be observed in the control group. This was a favourable condition for the PSM to produce valid estimates and for us to have confidence in the analysis.

4.3. Estimation of ATET

The ATET estimates over the common support which were derived using different matching algorithms are shown in Table 2. In particular, k-nearest-neighbour (with k = 1, k = 3 and k = 5), radius (with radii = 0.01, 0.02 and 0.04) and kernel (using Epanechnikov, normal and biweight kernel functions with a bandwidth of 0.08) were employed. The general pattern of the estimates was rather stable, despite some small differences in the coefficient values. In fact, with the exception of k = 1 nearest-neighbour, all the estimates indicated that students joining groups or Facebook pages had, on average, a lower probability to dropout, compared with those who were not part of such groups. The results also showed that the extent of the difference between the treated and control groups was not negligible, as it varied from 0.081 to 0.113, depending on the matching algorithm.

4.4. Sensitivity analysis

Although the PSM model included a rich set of covariates, it was possible that unobserved factors may have influenced the treatment effect estimates. Thus, in order to evaluate robustness of estimates to possible omitted (or unobserved) confounders, a sensitivity analysis was carried out with the bounding approach proposed by Rosenbaum [58]. The aim of sensitivity analysis was to determine how strong the effect of an omitted confounder should be to alter or undermine an inference about treatment effects, making it insignificant. Table 3 shows, for each matching algorithm, the critical values in terms of the magnitude (indicated with Γ) of an omitted confounder that would be needed in order for the 95% confidence interval of the estimated ATET to include zero. In practice, the value of Γ indicates how strong the influence of an unobserved confounder would need to be in order to nullify the corresponding estimated ATET.

Except for 1 Nearest Neighbour whose estimated ATET was not significant, the results showed that the critical values varied from 1.6 to 1.7 for all matching algorithms and that a confounder would have to be quite influential in order to nullify the estimated effect. Therefore, the results seemed to be quite robust with respect to deviations due to unobserved confounders, and this supported the external validity of our analysis.

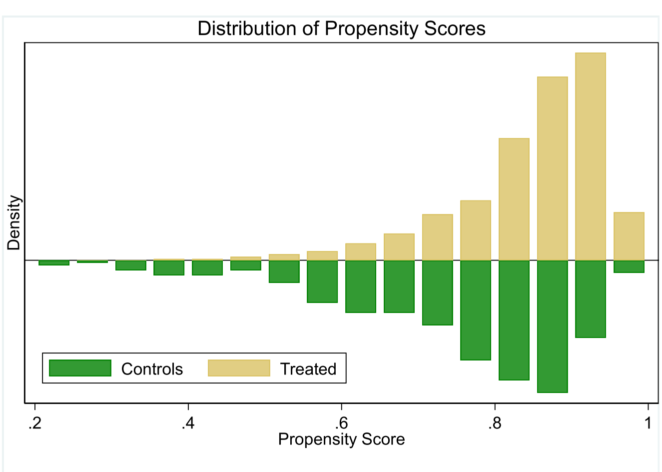

Fig. 1. Density distribution of the propensity scores in the treatment and control groups.

Table 2
Average treatment effect (ATET) of participation in groups or Facebook pages on university dropout within the first university year.

Matching technique	ATET	S.E.	T	Treated	Controls
Nearest-neighbour nearest-neighbour	−0.057	0.046	−1.23	1358	282
3 nearest-neighbour	−0.094	0.041	−2.30	1358	282
5 nearest-neighbour	−0.091	0.040	−2.27	1358	282
Radius					
with radii = 0.01	−0.081	0.039	−2.06	1358	282
with radii = 0.02	−0.080	0.038	−2.10	1358	282
with radii = 0.04	−0.086	0.038	−2.26	1358	282
Kernel					
Epanechnikov (0.08)	−0.097	0.036	−2.65	1373	282
normal (0.08)	−0.113	0.035	−3.26	1373	282
biweight (0.08)	−0.113	0.037	−2.53	1373	282

Table 3
Sensitivity analysis: critical values in order to nullify the corresponding estimated ATET.

Matching technique	Γ
Nearest-neighbour	
1 nearest-neighbour	–
3 nearest-neighbour	1.6
5 nearest-neighbour	1.6
Radius	
with radii = 0.01	1.7
with radii = 0.02	1.6
with radii = 0.04	1.6
Kernel	
Epanechnikov (0.08)	1.7
normal (0.08)	1.7
biweight (0.08)	1.7

5. Discussion

In the international literature, studies that have dealt with the issue of dropout are various, and they have explored this topic with particular reference to the factors that determine the choice to leave university within the first year of enrolment. Over the past decades, issues related to the use of Facebook and its effects in the social and relational life of people, especially of young generations and their academic life, have also been addressed. The universities, for their part, continue to be concerned about the high dropout rate and have therefore activated policies that counteract this problem, amongst others being the implementation of Facebook pages. However, there is poor empirical evidence of the use of Facebook as a factor that affects the choice to leave university. The main purpose of this paper is to contribute to this direction with empirical evidence from a case of an Italian university. Facebook use can indeed be applied in ways that are advantageous to students in respect to their engagement in reference to various factors, such as more investment in the academic experience of the college, greater interaction with faculty, greater involvement in extracurricular activities and more interaction with peers.

In our study, participation in groups or Facebook pages created and managed by university students with the aim of sharing information on degree courses or any material useful for studying and preparing for exams was considered as a treatment condition in order to estimate its potential effect on dropout within the first year. To correct for selection bias in the treatment assignment, the average treatment effect on the treated was estimated using PSM. All the proposed matching techniques in the analysis indicated that university dropout within the first year is lower for students participating in such groups or pages. This could happen because social networks help to strengthen academic integration and student engagement which, in turn, favour retention. Also, sensitivity analysis showed that the results seemed quite robust with respect to deviations from the conditional independence assumption, due to possible unobserved confounders.

Based on this evidence, universities could encourage the creation and maintenance of groups or Facebook (or any other social network) in order to enhance sharing information on degree courses or any material useful for studying and preparing for exams amongst students. Moreover, given that Facebook continues to be more and more popular amongst university students, and given that faculty and non-academic staff are interested in engaging and retaining them, it is also important that the use of Facebook within the academic world should be organised by academic institutions so it could continue to provide benefits and avoid the risk of misuse or abuse of this tool. Although these results cannot be considered conclusive, they can provide some useful insights and fuel the debate on this topic.

CRediT authorship contribution statement

Lucio Masserini: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing. **Matilde Bini:** Conceptualization, Investigation, Formal analysis, Methodology, Supervision, Writing - original draft, Writing - review & editing.

Appendix

Table A1
Descriptive statistics for the whole sample and the treated and control groups.

Variables	Total	Treated	Controls	*p*
Male	0.486	0.470	0.550	0.008
Lyceum	0.600	0.621	0.505	<0.001
Technical Institute	0.067	0.060	0.099	0.010
Vocational	0.044	0.041	0.056	0.210
Linguistic Institute	0.113	0.112	0.115	0.892
Delay in obtaining a high school diploma	0.143	0.125	0.222	<0.001
Delay in enrolment	0.175	0.150	0.289	<0.001
High school diploma grade 75–90	0.394	0.401	0.364	0.219
High school diploma grade 91–100	0.269	0.276	0.238	0.171
University department (Biology as reference)	0.065	0.064	0.070	<0.001
Chemical and Industrial Chemistry	0.020	0.019	0.022	0.778
Civilisation and Forms of Knowledge	0.066	0.061	0.085	0.114
Economics and Management	0.133	0.133	0.134	0.974
Pharmacy	0.037	0.036	0.046	0.401
Philology, Literature and Linguistics	0.095	0.091	0.115	0.252
Physics	0.023	0.018	0.046	0.003
Law	0.084	0.084	0.085	0.955
Computer Science	0.019	0.018	0.021	0.715
Civil and Industrial Engineering	0.127	0.131	0.100	0.126
Energy, Territory and Construction Engineering	0.036	0.036	0.030	0.623
Information Engineering	0.067	0.076	0.030	0.003
Mathematics	0.015	0.016	0.012	0.636
Clinical and Experimental Medicine	0.052	0.055	0.033	0.110
Surgical, Medical and Molecular Pathology	0.015	0.014	0.022	0.759
Translational Research	0.040	0.039	0.043	0.774
Agricultural, Food and Agro-Environmental Sciences	0.028	0.029	0.024	0.662
Geosciences	0.007	0.006	0.012	0.214
Political Sciences	0.057	0.056	0.064	0.585
Veterinary	0.017	0.017	0.018	0.872
Decision to enrol influenced by the family environment	0.019	0.019	0.018	0.934
Choice of the university for the presence of friends	0.024	0.027	0.010	0.076
Choice of the university for a particular degree course	0.058	0.058	0.058	0.984
Started working during the first year of studies	0.248	0.225	0.355	<0.001
Occasional commuter	0.029	0.021	0.064	0.000
Daily commuter	0.499	0.502	0.480	0.461
Weekly commuter	0.053	0.053	0.0.049	0.771
In site	0.281	0.281	0.284	0.911
Enrolment changed previous life habits	0.660	0.676	0.590	0.003
Attend classes during the first year (at least 75%)	0.784	0.806	0.686	<0.001
Do not attend classes	0.049	0.038	0.104	<0.001
Attend professors' reception (at least once)	0.567	0.582	0.501	0.008
Benefit from the tutoring service	0.103	0.110	0.077	0.077
Study regularly in the department's study rooms	0.438	0.452	0377	0.013
Study difficulties as compared to high schools	0.387	0.391	0.370	0.487
Live in institute together with other students	0.045	0.048	0.034	0.276
Live in apartment alone or with other students	0.306	0.305	0.306	0.972
Have friends already enrolled in the same degree course	0.15	0.328	0.252	0.007
Used to studying with other students	0.676	0.661	0.745	0.003

(*continued on next page*)

Table A1 (*continued*)

Variables	Total	Treated	Controls	*p*
Have classmates to exchange lecture notes/information	0.879	0.900	0.780	<0.001
Have contacts with classmates outside the university	0.687	0.715	0.570	<0.001
Participation in extra-curricular activities (i.e. seminars)	0.072	0.077	0.049	0.072
Participation in other activities (i.e. meetings and events)	0.371	0.391	0.281	<0.001

Table A2
Propensity Score estimates by logistic regression.

Variables	Coefficient	*SE*	*p*
Male	−0.260	0.087	0.003
Lyceum	0.109	0.110	0.321
Technical Institute	−0.247	0.172	0.153
Vocational	−0.024	0.207	0.906
Linguistic Institute	0.127	0.160	0.428
Delay in obtaining a high school diploma	−0.146	0.114	0.200
Delay in enrolment	−0.361	0.105	0.001
High school diploma grade 75–90	0.067	0.094	0.472
High school diploma grade 91–100	0.036	0.111	0.747
University department (Biology as reference)			
Chemical and Industrial Chemistry	0.031	0.305	0.919
Civilisation and Forms of Knowledge	0.012	0.209	0.953
Economics and Management	0.251	0.192	0.190
Pharmacy	−0.108	0.242	0.657
Philology, Literature and Linguistics	−0.112	0.190	0.556
Physics	−0.694	0.271	0.010
Law	0.157	0.200	0.432
Computer Science	0.277	0.316	0.381
Civil and Industrial Engineering	0.263	0.193	0.174
Energy, Territory and Construction Engineering	0.178	0.265	0.502
Information Engineering	0.514	0.236	0.029
Mathematics	0.014	0.350	0.969
Clinical and Experimental Medicine	0.557	0.240	0.020
Surgical, Medical and Molecular Pathology	0.162	0.359	0.651
Translational Research	−0.137	0.238	0.565
Agricultural, Food and Agro-Environmental Sciences	0.166	0.271	0.541
Geosciences	−0.275	0.406	0.498
Political Sciences	0.226	0.223	0.310
Veterinary	0.174	0.340	0.608
Decision to enrol influenced by the family environment	−0.040	0.282	0.888
Choice of the university for the presence of friends	0.663	0.313	0.034
Choice of the university for a particular degree course	0.099	0.175	0.572
Started working during the first year of studies	−0.233	0.094	0.013
Occasional commuter	−0.030	0.255	0.905
Daily commuter	0.039	0.129	0.763
Weekly commuter	−0.031	0.209	0.880
In site	−0.053	0.177	0.765
Enrolment changed previous life habits	0.164	0.083	0.050
Attend classes during the first year (at least 75%)	0.141	0.104	0.173
Do not attend classes	0.165	0.210	0.432
Attend professors' reception (at least once)	0.085	0.084	0.313
Benefit from the tutoring service	0.137	0.144	0.343
Study regularly in the department's study rooms	0.085	0.086	0.323
Study difficulties as compared to high schools	−0.045	0.084	0.589
Live in institute together with other students	0.147	0.269	0.584
Live in apartment alone or with other students	−0.133	0.158	0.399
Have friends already enrolled in the same degree course	0.166	0.089	0.062
Used to studying with other students	−0.084	0.092	0.359
Have classmates to exchange lecture notes/information	0.285	0.121	0.019
Have contacts with classmates outside the university	0.106	0.091	0.248
Participation in extra-curricular activities (i.e. seminars)	0.207	0.170	0.225
Participation in other activities (i.e. meetings and events)	0.230	0.094	0.015
Constant	0.352	0.280	0.209

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.seps.2020.100865.

References

[1] Porter OF. Undergraduate completion and persistence at four-year colleges and universities. Washington, DC: National Institute of Colleges and Universities; 1990.

[2] McKenzie K, Schweitzer R. Who succeeds at university? Factors predicting academic performance in first year Australian university students. High Educ Res Dev 2001;20:21–33. https://doi.org/10.1080/07924360120043621.

[3] Aydin S. A review of research on Facebook as an educational environment. Educ Technol Res Dev 2012;60(6):1093–106. https://doi.org/10.1007/s11423-012-9260-7.

[4] Manca S, Ranieri M. Is Facebook still a suitable technology-enhanced learning environment? An updated critical review of the literature from 2012 to 2015. J Comput Assist Learn 2016;32(6):503–28. https://doi.org/10.1111/jcal.12154.
[5] Piotrowski C. Pedagogical applications of social media in business education. J Educ Technol Syst 2015;43(3):257–65. https://doi.org/10.1177/0047239515570575.
[6] Tess PA. The role of social media in higher education classes (real and virtual)—a literature review. Comput Hum Behav 2013;29(5):A60–8. https://doi.org/10.1016/j.chb.2012.12.032.
[7] Barczyk CC. Facebook in higher education courses: an analysis of students' attitudes, community of practice, and classroom community. Int Bus Manag 2013;6(1):1–11. https://doi.org/10.3968/j.ibm.1923842820130601.1165.
[8] Clements JC. Using Facebook to enhance independent student engagement: a case study of first-year undergraduates. High Educ Stud 2015;5(4):131–46. https://doi.org/10.5539/hes.v5n4p131.
[9] Duggan M, Ellison NB, Lampe C, Lenhard A, Madden M. Social media update 2014. 2015. http://www.pewinternet.org/2015/01/09/social-media-update-2014/.
[10] Gunuc S, Kuzu A. Student engagement scale: development, reliability and validity. Assess Eval High Educ 2014;40(4):587–610. https://doi.org/10.1080/02602938.2014.938019.
[11] Junco R. The relationship between frequency of Facebook use, participation in Facebook activities, and student engagement. Comput Educ 2012;58(1):162–71. https://doi.org/10.1016/j.compedu.2011.08.004.
[12] Fagioli L, Rios-Aguilar C, Deil-Amen RJ. Changing the context of student engagement: using Facebook to increase community college student persistence and success. Teach Coll Rec 2015;117(12):1–42. https://www.coe.arizona.edu/sites/coe/files/HED/Changing%20the%20Context%20of%20Engagement_0.pdf.
[13] Junco R. The relationship between frequency of Facebook use, participation in Facebook activities, and student engagement. Comput Educ 2011;58(1):S162–71. https://doi.org/10.1016/j.compedu.2011.08.004.
[14] Rosenbaum PR, Rubin DP. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70(1):41–55.
[15] Kalsner L. Issues in College student retention. High Educ Ext Serv Rev 1991;3(1):1–10. https://eric.ed.gov/?id=ED350894.
[16] Tinto V. Dropout from higher education: a theoretical synthesis of recent research. Rev Educ Res 1975;45:89–125. https://89-125.10.2307/1170024.
[17] Tinto V. Leaving college: rethinking the causes and cures of student attrition. Chicago: University of Chicago Press; 1987.
[18] Smith JP, Naylor RA. Dropping out of university: a statistical analysis of the probability of withdrawal for UK university students. J R Stat Soc Ser A Stat Soc. Series A (Statistics in Society) 2001;164(2):389–405. www.jstor.org/stable/2680559.
[19] Belloc F, Maruotti A, Petrella L. University drop-out: an Italian experience. High Educ 2010;60(2):127–38.
[20] Johnes G. Determinants of students wastage in higher education. Stud High Educ 1990;15:87–99. https://eric.ed.gov/?id=EJ405559.
[21] Noel L, Levitz R. Increasing student retention. San Francisco: Jossey-Bass; 1991©1985.
[22] Fielding A, Belfield C, Thomas H. The consequences of drop-out on the cost-effectiveness of 16-19 colleges. Oxf Rev Educ 1998;24(4):487–511. https://doi.org/10.1080/0305498980240405.
[23] Pascarella ET, Terenzini PT. How college affects students: a third decade of research. second ed. San Francisco, CA: Jossey Bass; 2005. 2005.
[24] Arulampalam W, Naylor R, Smith J. Factors affecting the probability of first year medical student dropout in the UK: a logistic analysis for the intake cohorts of 1980–92. Med Educ 2004;38(5):492–503. https://doi.org/10.1046/j.1365-2929.2004.01815.x.
[25] Hovdhaugen E. Working while studying: the impact of term-time employment on dropout rates. J Educ Work 2015;28(6):631–51. https://10.1080/13639080.2013.869311.
[26] Kuh GD. Assessing what really matters to students learning: inside the national survey of student engagement. Change 2009;33(3):10–66.
[27] Cingano F, Cipollone P. University drop-out. the case of Italy. Temi di discussione (Economic working papers) 626. Bank of Italy. Economic Research Department; 2007.
[28] Murtaugh PA, Burns LD, Schuster J. Predicting the retention of university students. Res High Educ 1999;40(3):355–71.
[29] Grandy J. Persistence in science of high-ability minority students. J High Educ 1998;69(6):589–620.
[30] Leppel K. Similarities and differences in the college persistence of men and women. Rev High Educ 2005;25(4):433–50.
[31] Braxton JM, Milem JF, Sullivan AS. The influence of active learning on the college student departure process. J High Educ 2002;71(5):569–90.
[32] Tinto V. Classrooms as communities: exploring the educational character of student persistence. J High Educ 1997;68(6):599–623. https://doi.org/10.2307/2959965.
[33] Lau LK. Institutional factors affecting student retention. Education 2003;124(1):126–36.
[34] Okun MA, Benin M, Brandt-Williams A. Staying in college: moderators of the relation between intention and institutional departure. J High Educ 1996;67:577–96.
[35] Beil C, Reisen CA, Zea MC. A longitudinal study of the effects of academic and social integration and commitment on retention. NASPA J 1999;37(1):376–85.
[36] Cambiano RL, Denny GS, De Vore JB. College student retention at a midwestern university: a six-year study. J Coll Admiss 2000;166:22–9.
[37] Stinebrickner T, Stinebrickner R. Learning about academic ability and the college dropout decision. J Labor Econ 2012;30(4):707–48. 74810.1086/666525.
[38] Roblyer MD, McDaniel M, Webb M, Herman J, Witty JV. Findings on Facebook in higher education: a comparison of college faculty and student uses and perceptions of social networking sites. Internet High Educ 2010;13(3):134–40. https://doi.org/10.1016/j.iheduc.2010.03.002.
[39] Hurt NE, Moss G, Bradley C, Larson L, Lovelace M, Prevost L, Riley N, Domizi D, Camus MS. The 'Facebook' effect: college students' perceptions of online discussions in the age of social networking. Int J Scholarsh Teach Learn 2012;6(2):1–24. https://doi.org/10.20429/ijsotl.2012.060210.
[40] Cooper J, Weaver KD. Gender and computers: understanding the digital divide. Mahwah, NJ: Lawrence Erlbaum Associates; 2003.
[41] DiMaggio P, Hargittai E, Celeste C, Shafer S. Digital inequality: from unequal access to differentiated use. In: Neckerman K, editor. Social inequality. New York: Russell Sage Foundation; 2004. p. 355–400.
[42] Hargittai E. The digital reproduction of inequality. In: Grusky D, editor. Social stratification. Boulder, CO: Westview Press; 2008. p. 936–44. 2008b.
[43] Junco R, Merson D, Salter DW. The effect of gender, ethnicity, and income on college students' use of communication technologies. Cyberpsychol, Behav Soc Netw 2010;13(6):37–53.
[44] Hargittai E. Whose space? Differences among users and non-users of social network sites. J Comput Mediat Commun 2008;13(1):276–97.
[45] Greenhow C, Askari E. Learning and teaching with social network sites: a decade of research in K-12 related education. Educ Inf Technol 2017;22(2):623–45. https://doi.org/10.1007/s10639-015-9446-9.
[46] Ajjan H, Hartshorne R. Investigating faculty decisions to adopt Web 2.0 technologies: theory and empirical tests. Internet High Educ 2008;11:71–80. https://doi.org/10.1016/j.iheduc.2008.05.002.
[47] Rogers-Estable M. Web 2.0 use in higher education. Eur J Open Dist E Learn 2014;17(2):129–41. https://doi.org/10.2478/eurodl-2014-0024.
[48] Lam L. An innovative research on the usage of Facebook in the higher education context of Hong Kong. Elec J e-Learn 2012;10(4):378–86.
[49] McCarthy J. Blended learning environments: using social networking sites to enhance the first year experience. Australas J Educ Technol 2010;26(6):729–40. https://doi.org/10.14742/ajet.1039.
[50] Hung H-T, Yuen SC-Y. Educational use of social networking technology in higher education. Teach High Educ 2010;5(6):703–14. https://doi.org/10.1080/13562517.2010.507307.
[51] Kabilan MK, Ahmad N, Abidin MJZ. Facebook: an online environment for learning of English in institutions of higher education? Internet High Educ 2010;13(4):179–87. https://doi.org/10.1016/j.iheduc.2010.07.003.
[52] Freeman TM, Anderman LH, Jensen JM. Sense of belonging in college freshmen at the classroom and campus levels. J Exp Educ 2007;75(3):203–20. https://doi.org/10.3200/JEXE.75.3.203-220.
[53] Hausmann LRM, Schofield JW, Woods RL. Sense of belonging as a predictor of intentions to persist among African American and white first-year college students. Res High Educ 2007;48(7):803–39. https://doi.org/10.1007/s11162-007-9052-9.
[54] Kahu ER. Framing student engagement in higher education. Stud High Educ 2013;38(5):758–73. https://doi.org/10.1080/03075079.2011.598505.
[55] Trowler V, Trowler P. Student engagement evidence summary. Department of Educational Research, University of Lancaster; 2010. 2010.
[56] Imbens G. Nonparametric estimation of average treatment effects under exogeneity: a review. Rev Econ Stat 2004;86(1):1–29.
[57] Holland PW. Statistics and causal inference. J Am Stat Assoc 1986;81(396):945–60.
[58] Rosenbaum PR. Observational studies. New York, NY: Springer; 2002.
[59] Caliendo M, Kopeinig S. Some practical guidance for the implementation of propensity score matching. J Econ Surv 2008;22(1):31–72.
[60] Rosenbaum PR, Rubin DP. Constructing a control group using multivariate matched sampling models that incorporate the propensity score. Am Statistician 1985;39(1):33–8.
[61] Bryson A, Dorsett R, Purdon S. The use of propensity score matching in the evaluation of active labour market policies. 2002. Department for Work and Pensions Working Paper No. 4.
[62] Leuven E, Sianesi B. PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. Statistical software components S432001. Boston College Department of Economics; 2003.
[63] Becker S, Caliendo M. Sensitivity analysis for average treatment effects. STATA J 2007;7(1):71–83.
[64] Dehejia RH, Wahba S. Propensity score matching methods for nonexperimental causal studies. Rev Econ Stat 2002;84(1):151–61.
[65] Heckman J, Ichimura H, Smith J, Todd P. Characterizing selection bias using experimental data. Econometrica 1998;66(5):1017–98. https://doi.org/10.2307/2999630.
[66] Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avron J, Sturmer T. Variable selection for propensity score model. Am J Epidemiol 2006;163(12):1149–56.
[67] Heckman J, LaLonde R, Smith J. The economics and econometrics of active labor market programs. In: Ashenfelter O, Card D, editors. Handbook of labor economics, vol. III. Amsterdam: Elsevier; 1999. p. 1865–2097.
[68] Lechner M. A note on the common support problem in applied evaluation studies. Univ. of St. Gallen Economics, Disc. Paper. 2001 Nov;1.

Lucio Masserini is associate professor in Economic Statistics at the Department of Economics and Management at the University of Pisa. He holds his Ph.D. in Applied Statistics

at the University of Florence in 2011. His research interests are in policies evaluation in the field of higher education and university transition to the labour market and also in quality of services, satisfaction analysis and market research. As regards statistical methods, his interests are in structural equations models, models for longitudinal data, latent growth models, latent variables, generalized mixed linear models, multilevel analysis and methods for the evaluation of treatment effects.

Matilde Bini Ph.D. in Applied Statistics (1995), University of Florence, she is now professor of Economic Statistics at the Department of Human Sciences of European University of Rome. Her research areas of interest are in macroeconomics and business analyses, labour markets, evaluation of higher education processes and public services. In statistical methodology, she works on Generalized Linear Latent and Mixed Models (GLLAMMs), structural equation models, multivariate methods, longitudinal models. Robust approach of the forward search for diagnostic analyses for outlier detection and data quality, in surveys and censuses of economic and social populations. She was visiting Student at the Department of Statistics of the University of Florida (USA), visiting Professor at School for Advanced Studies di Institutions, Markets, Technologies of Lucca, and at Department of Estadística, Econometría, Investigación operativa, Organización de Empresas y Economía Aplicada of the University of Cordoba (Spain). She is member of various national and international Scientific Societies of Statistics.