Endogenous Shocks in Social Networks: Exam Failures and Friends' Future Performance

Marchenko, Maria

DOI:
10.57938/c193101d-e440-496c-8036-4be54f481689

Published: 01/01/2019

Document Version
Publisher's PDF, also known as Version of record

Department of Economics
Working Paper No. 292

Endogenous Shocks in Social Networks: Exam Failures and Friends' Future Performance

Maria Marchenko

August 2019

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

Endogenous Shocks in Social Networks: Exam Failures and Friends' Future Performance

Maria Marchenko *

June 18, 2019

Abstract

Exam failures of the students in a specific network may influence not only the future performance of the student but also all students from their friendship networks, affecting the overall cohort's performance. Therefore, it is crucial to understand how the whole network responses to failure. The difficulty of such analysis is incorporated in the probability of the failures being highly endogenous. In this paper, I am applying the novel identification and estimation approach to deal with such endogeneity. I am exploring the dynamic data on the students' networks in HSE, Nizhniy Novgorod. The results suggest that, on average, the exam failure of the friend have a negative effect on future performance.

*JEL classification codes: C21, C49, I21
Keywords: Peer effects; exam failures; dropouts; dynamic networks
Please check the latest version at sites.google.com/site/mariavmarchenko/examfailures.pdf
WU Vienna, Austria. Email: maria.marchenko@wu.ac.at.
This research has benefited from discussions with Andreas Dzemski, Maria Yudkevich, Alex Teytelboym, Andrea Weber, Enno Mammen, Jan Nimczik, Stephen Kastoryano, seminar participants in Mannheim, and participants of RSSIA 2015.
The financial support from the Government of the Russian Federation within the framework of the Basic Research Program at the National Research University Higher School of Economics and within the framework of the implementation of the 5-100 Programme Roadmap of the National Research University Higher School of Economics is acknowledged.

1 Introduction

The peer effect on behaviour and performance is widely studied in the literature across many fields, and its importance is well understood. However, spillover effects of shocking events, such as exam failures, through the network are rarely analyzed. Such events occurring to one or several network members will not only influence themselves but will potentially have an effect on the future outcomes of their peers. Network links, hence, can be viewed as the channels transferring the shock from one person to the others. I look at the dynamic network of university students linked by friendship ties to analyse the possible connection between the exam failures and the future educational outcomes of peers. In case of existence of such spillover effects of the failures, they may have bigger consequences for the university performanve as well.

Students who have to retake exams, or potentially having to drop out of university completely, may feel under intense pressure and this can result in lower productivity (see, for example Baumeister (1984) or Apesteguia and Palacios-Huerta (2010)). Determining how these shocks influence the behaviour of students and their friends can help to better understand the dynamics of network performance. In addition, understanding more about the influence on all students' behaviour of such shocks may also help universities to adjust their strategy in establishing a threshold for passing the classes. Very selective university may relax the threshold for passing the courses, if the failure negatively affects too many students. A large proportion of university dropouts is the result of the class passing policies created by the universities - policies which can be controlled. In some institutions of higher education, most of the dropouts are directly affected by the failures of the exams.

I use data from the Higher School of Economics, where three exam failures during the same term will lead to the expulsion of the student[1]. However, in some institutions of higher education, students who fail exams may remain in the study program and repeat

[1]Each academic year consists of 2 terms, each includes 2 modules.

the failed courses under individual study plan, but the probability of leaving the university before graduation for such students will be higher than for the students without failed exam. Dropouts are viewed as an important problem for the universities since they create sunk costs. For example, costs of the university dropouts in Germany were estimated at \$11.5 billion in 2007, Stifterverband (2007). In Australia they are \$1.36 billion, according to the report of Adams and Banks (2010). Therefore, understanding the influence the failed exam may have on future student performance may allow for possible university level policy improvements. This could reduce sunk costs and adjust the performance for an optimum outcome.

Following standard argument of peer effect literature (Manski (1993)), other things equal, two students with the same abilities influence their peers' performance in a similar way. These two students have equal chances of failing their exams. However, for these similar students, the failure will be determined by a unobserved random component (error term). If one of the students passes the exam and the other fails it, they should affect the behaviour and performance of the peers differently. The failure of the exam is endogenous in the university context, since the probability of the failure may also depend on the network's performance. The logic is similar to standard peer effect in education suggesting the comovement of outcomes (Jackson (2010)). The student with worse performing friends are more likely to fail the exam than the student with better performing friends. The endogeneity does not allow identification of the effect of the failure on the friends' future performance with the tools used in the peer effect literature. To overcome this problem, I use an approach suggested in Marchenko (2018)[2]. It uses the difference between failing and passing the exam for similar students being determined by unobserved random component. The predicted probability of failing the exam for two similar students, according to the peer effect argument, should be the same. Therefore, the above mentioned unobserved random component is the one that

[2]Please check the latest version at sites.google.com/site/mariavmarchenko/endshocks.pdf

defines the effect of the failure on the friends' future performance and, moreover, it is exogenous, which eliminates the problem of endogeneity of the failure.

I address the question of failures in the university by studying the sample of students from one cohort at the National Research University - Higher School of Economics, Nizhniy Novgorod campus, a highly selective university in Russia. The data consist of two waves and include self-reported friendship links to the classmates as well as the information about all the grades of the students and some additional individual information. I consider the exam failures during the first year of studies as something that will be one of the determinants of the changes in the friends' grades in the second year. I assune, that the network reported at the end of the first year influences the probability of experiencing the exam failure during the first year. Additionally, these links will transfer the shock to affect the performance of the friends in the next period. The students' networks change between two surveys, and the new network from the second year influences the performance in the second period. Moreover, the new links are crucial for comparing the outcomes with and without influence of the failure. The network is defined by the self-reported links from the two waves of the survey at the end of the study year.

Although I do not examine the direct peer effect in this paper, I use the general framework of peer effects literature and its methodological fundamentals of the model introduced by Manski (1993). The direct peer effect is usually considered as instantaneous, whereas the shocking event can have a deferred effect. It will have an influence on the network only in the period after the shock happens and is revealed to the peers. The instantaneous and dynamic component of the network's influence on individual outcomes should be distinguished: the former is due to the direct peer effect, the latter is related to the spillovers from shocking events.

In general, the shock's influence is ambiguous. Although the unexpected shock may serve as a wake-up call to some students and could motivate them to be more focused on their studies, for others, the negative influence may dominate. First of all, the connections between students can be extremely close. The student may reduce the amount of time spent on personal studies due to the shared activities outside of the university with the friend who failed. Alternatively, the stress experienced after observing the friends failing may be so high, that it will negatively affect future performance.One or another direction may dominate depending on characteristics of the sample and institutional environment. Therefore, it is important to empirically examine the effect of the exam failures in order to determine the direction of the effect for the particular setting. The paper finds the negative effect of the exam failures on the future outcomes of the peers, which suggests that the stress or tightness of the connections effects are prevailing in the setting under consideration.

I focus on first-year failures only. In the first year, students are more likely to fail due to lack of ability or because of difficulties adjusting to the new environment. The first exams may appear to be too difficult for some of the students despite the student having gained entrance to the university in the first place. Therefore, the effect of the exam failure of the friends relates to the misconception of peers' abilities.

This paper relies on the model and identification results proposed by Marchenko (2018). The identification requires, firstly, the existence of friends of friends that are not connected to the student. Hence, the friends of friends do not affect the student directly, but via the common friend only. It is not necessary that each student has such link structure, but it is enough that part of the network satisfy this requirement. Secondly, the identification demands longitudinal network variation with the existence of some network variation not caused by the exam failure. Changes in the network allow for a comparison of the influence of old and new peer group on the outcome. The presence of new friends and absence of old ones creates variation in the peer group characteristics, moreover, only old links are transferring the failure, and this helps to identify the effect of the failure. However, it is important that some changes in the network are driven solely by a natural adjustment in the social environment, i.e. friends reveal different interests. Certainly, some of the links might be broken due to exam retakes and friends dropouts. However, the natural variation of the network is a valid assumption for the students' network setting, which I am proving later in the paper. The links formed in the first year are highly likely to be revised due to the gradual unveiling of the friends' personal characteristics. Of course, this natural variation suggests the endogenous link formation during the second period, which does not allow the correct identification of the second year peer effects, but does not affect the identification of the shock. This is formally proved in Marchenko (2018).

I discuss the model without link formation process and, therefore, it analyses only the direct effect of the shock, although the shock might also affect the changes in performance indirectly by influencing the network.

The paper is organized as follows. Section 2 discusses the related literature. Section 3 describes the institutional environment of the education system in Russia and the data used. Section 4 discusses the proposed model, states the identifying assumptions and proposes the estimation method for the model. Section 5 discusses the network characteristics and identifying assumptions. Section 6 provides the estimation results and evidence of the influence of dropouts and retakes on peers. Section 7 concludes.

2 Related literature

I contribute to the literature by studying changes in peer behaviour and achievement in response to an individual shock. I look at the exam failures, which might result in the dropouts from the university, and provide an evidence of the shock transferring through the network.

The influence of social connections on the student's performance, primarily on their academic achievements, is being widely studied in the literature. Coleman (1966) suggests that the social environment of the students is influencing their individual achievements. Manski (1993) introduced the three effects playing the role in determining the individual outcomes in presence of social connections: exogenous effect, endogenous effect, and correlated effect. The endogenous effect explains how individual outcomes will be affected by the average outcomes of their peer group and it is the most important effect to identify and estimate as it may have policy implications. The identification of such effect, in the case of group interactions, requires an additional source of exogenous variation. One strand of empirical literature uses the exogenous group formation as the source of exogenous variation. For example, Lyle (2007), Lyle (2009) and Carrell et al., 2009 find peer effect by looking at the military institutions framework. The groups in the US Military academy are formed administratively and the syllabi are predetermined, so the students are interacting closely inside of their group. In the university context, the endogenous group formation due to the free choice of the study courses is more common, however, there are several examples, where the compulsory courses and exogenous group formation allow to study the peer effect for the whole group. For example, De Giorgi et al. (2010) study the sample of students at Bocconi University, who take nine compulsory courses during the first three semesters and who are randomly assigned to the study groups. The exogenous group formation helps to find the peer effect in the choice of major. Androushchak et al. (2013) examine the influence of the ability of students on the achievements of classmates at the university under consideration.

When study groups are formed endogenously, the random assignment of dorm mates may help to obtain the necessary exogenous variation. For example, Sacerdote (2001) looks at the sample of roommates at Dartmouth, who are randomly assigned at the beginning of their studies. The estimation of the endogenous peer effect as an effect of an average group performance has been criticized. The whole group is not likely to affect the student equally. Additional assumptions on the structure or the ranking inside the peer group or exact links are preferable for better estimation of peer effect, although social network data is not always available.

Use of social network data requires other identifying assumptions that restrict the network. It can be more fruitful in comparison to the group interaction since more information is available about the network structure. Bramoullé et al. (2009) prove the identification of the peer effect in social networks under rather mild assumptions. The network should include intransitive triads, i.e. two members of the network connected via the third person. They apply the identification results to the Add Health database of high school students in the USA. This identification result used broadly in network effects analysis. Poldin et al. (2016) apply this result to a cohort of third-year students, similar to the dataset used in this paper.

The effect of the shock on the future outcome of the network is less studied, however, such questions are gaining more popularity. Angelucci and De Giorgi (2009) discusses the indirect effect of cash transfers, Graham (2008) analyses the changes of school performance in response to the class size changes using group interactions. Dieye et al. (2014) explore scholarship programs in Colombia based on the network data. Comola and Prina (2014) also use the network data to explore the effect of a shock and is the first to use the network observed in dynamics, developing the dynamic peer effect model, similar to the one proposed in Marchenko (2018) and used in this paper. However, all focus on the randomized treatment as a shock, therefore, they are not facing the endogeneity of the shock problem. To the best of my knowledge, I am the first to analyse the indirect effect of endogenous shocks through the network.

Unlike the academic achievements of the students, the academic failures are not studied that widely in the framework of the social environment. Tinto (1975) asserts that the institutional and social environment of a university is believed to affects the academic failures of the students. However, recent economic literature analysing the dropouts is limited, for example, Li et al. (2013), Goux et al. (2014) and some others. Dropouts and failures are more widely discussed from the point of view of psychology or sociology, see Rumberger (2011). I fill an existing gap in the literature, studying the spread of the effect of the exam failures further in the network.

3 Data and Descriptive analysis

The data in this paper a from a survey of one cohort of students in the National Research University Higher School of Economics (HSE further in the text), Nizhny Novgorod campus. I first discuss shortly characteristics of the Russian higher education system and of the HSE.

3.1 The system of higher education in Russia and specifics of the sampled university.

Individuals with completed full vocational education or completed professional education of non-university level are eligible to enter the university.Most of the places in the state universities are financed by the government: around 65%[3], but it differs among institutions. For example, the university analysed in this paper provided 340 state-financed places out of total 431 in 2012[4]. The tuition fee varies from institution to institution, in our example, it varies between 130000 and 165000 Rubles per year, which equals to 28-36 times the minimum monthly wage or 18-23 times the minimum cost of living in Russia.

Students are accepted to the universities depending on the scores of the obligatory standardized examination, Unified School Examination, conducted at the end of the last school year. Each high school graduate has to take exams in several subjects: Mathematics and Russian are mandatory to graduate from the school, the other subjects are chosen by the graduates depending on their preferences and the requirements of the universities they are aiming to apply to. For example, economic department of NRU-HSE requires the USE results in Social Studies (a mixture of basic knowledge about different

[3]According to the Monitoring of education markets and organizations (MEMO), NRU HSE. Report in Russian.

[4]The main dataset uses 2012 cohort of students, details are described in the next subsection

aspects of society: philosophy, sociology, social psychology, law, political science) and Foreign language in addition to the mandatory results in Mathematics and Russian. A second way to enter some universities is via regional and national level Olympiads. These Olympiads are subject-specific and considered to be more sophisticated than the school exams, so they are designed to attract more talented students. The university under consideration accepts the winners and prize-takers of these competitions without exams if the major of the Olympiads corresponds to the university department (Economic Olympiads for economics department, Entrepreneurship Olympiads for management department etc.) or automatically given the highest score for the other subjects. However, those students are still required to take the USE and have the scores not lower than the required minimum (65 out of 100 in 2015, significantly lower than the requirement to be accepted). The share of students entering universities via the Olympiads is around 5-6% overall in Russia, but it is much higher for the university under consideration, around 40%, because of the selective status of the university. Therefore, in general, the group of students entering HSE is more or less homogeneous and consists of the high-achievers. Even though Nizhniy Novgorod branch of HSE is less selective than the main Moscow branch, the level of the admitted students is still very high. The list of all accepted students is publicly available in the university itself as well as on the website.

Usually, universities in Russia have an exogenous group formation. The students are randomly split into groups of 20-30 people before the beginning of the studies. Students rarely know their their new classmates before the group is formed, hence, almost all links are newly formed in the study groups. These groups stay mainly intact for the first three years. Several groups or even all students attend lectures together, whereas each group has separate tutorials. Changes to the group structure may occur if many students leave the university and the group becomes too small. Most of the universities have by now adopted the Bologna Process model of 4 years for Bachelor's degree and 1-2 years of Master's degree. In most universities each academic year has 2 terms with exams periods after each. But HSE has 4 terms per year, with some exams or pass-fail exams after the 1st and the 3rd term and with most exams after the 2nd and the 4th term. The student is not allowed to fail more than two exams per half-year (1+2 or 3+4 term) and retakes are conducted only after the 2nd and 4th exam periods of each year. All results of all students are publicly available near the students' office in the university and online so that everybody can follow their own performance, compare to the peers, and the tuition students can understand, whether they are eligible for the tuition discount.

3.2 Data description

The dataset includes the administrative and network information for the 2012 cohort of HSE, Nizhny Novgorod. Students from Economics, Management, Law and Computer Science are surveyed at the end of each study year, but before the final round of examinations. The first year survey has 320 participants, the second - 296. Students were asked to indicate their friends in the same cohort. The setup has long intervals between the survey waves. This allows capturing more persistent trends of network dynamics, although the failures of examinations during the first three terms of the year before the network is observed may have a spillover effect on the last term's examination grades.

The data also provide information from administrative records on all exam results, retakes and dropouts, as well as some personal data: gender, high school examination results, type of living (dormitory or not, dorm mates for those who live in the dormitory), parental education, some indicators of willingness to succeed or efforts (time spent on homework, time spent online on social networks, indicator of having a job parallel to studies).

The typical problems of self-reported data are present in the dataset. There are several observations with partially missing information on the network links. These entries need to be handled with care since they might suggest either the students without friends, indicating the antisocial behaviour, or the students that just skipped the questions, while answering the questionnaire. 13 students have no friends links, however, two of those provide an information about connections in the help networks, which could indicate the antisocial behaviour of the students. There is no information on particular friends for four more students, who just said they are friends with a lot of students, or even with all students.

Sampling is of a slight concern as well. The survey has 320 observations out of 432 students that entered the four departments of the university in 2012, that gives approximately 75-80% of the full population of students (Table 1). Some of the students have indicated the link to somebody outside of the sample, which can lead to overestimation of the importance of the observed links. However, the survey was conducted on several occasions, during lecture periods, so those, who did not answer the survey, are likely to attend the university only infrequently, and hence to have less influence on the other students. Survey was administered at the end of the year before the last term's exam, some students not in the sample had dropped out earlier.

Insert Table 1 here

Table 1 also demonstrates an inability of the dataset to catch all the information about the dropouts (only 40% are present in the sample) and their small amount in the network, because interviews are conducted at the end of the year, and many students are no longer continue their studies at that moment. Hence, the econometric analysis of dropouts is implausible, and forces to study exam retakes instead.

Inability to observe the dropouts causes the existence of links not captured by the survey. The friends' dropout from the university potentially influences the future performance of the student in the same way, as the retakes, and exclusion of those links from the analysis may lead to the overestimating of the influence of the retakes. However, many students are failing at the very beginning of the studies, when the newly established links are less likely to be very tight. The dropouts later in the course of studies are more likely to be predicted by the friends than the retakes, therefore, the actual shock might be less significant. It is still possible that the final results are slightly biased due to the missing links, but the bias are not likely to distort the results very much.

The survey design on friendship networks changed over the waves of the survey. The first wave asked for no more than seven friends and has seven lines for the names, which was ignored by approximately 2% of the sample, the second wave did not put any restriction on the number of friends, although it has seven lines as well. This lead to almost 50% of the students in the first period reporting exactly seven friends, whereas only 10,3% of the students in the second period indicated the same number of friends, and seven friends are the maximum of the second wave. The distribution of the number of the friends is presented in Table A.1[5] of Appendix. The average and median number of connections is 6 in the first year, whereas it is 4 in the second wave. It is likely, that in the first wave some of the students had to restrict themselves to exactly seven names, whereas some felt obliged to include more people than they are actually tightly connected to, which may cause underestimation of the importance of some links and overestimation of the others. Lower average number of friends in the second period may be caused by particularities of the survey construction as well as by the real trends in the network development.

3.3 Descriptive analysis

People often tend to connect based on similarities in their observed and unobserved characteristics. Table 2 summarizes the findings on the affinity of the peers in the network. Most friends are coming from the same group, more than 84% and almost all

[5] A.1 also have information on the distribution of number of friends from the similar survey with no restriction on number of friends, conducted on the next year cohort. However, this survey is not suitable for my analysis, and is only used for the comparison.

friends are from the same department.

Females are more likely to connect to peers of the same gender, whereas males have more diverse networks. Gender difference also exists in the probability of connecting to the dormitory mates: males are more likely to connect. The share of the friends with the same living conditions is, however, decreasing with time, suggesting that some other characteristics matter more for creating and sustaining the links.

Future plans on average seem not to matter a lot for the link formation: friends with the same plans for the future education are about 50% of the peers. This share could probably be higher, if the students were asked about there plans later in the course of their studies, and not during the first year. However, given the student's plans to pursue a Master's degree, her peers are as well more oriented on continuing the studies after the Undergraduate level.

Insert Table 2 here

Table 3 summarizes the findings on the number and shares of friends with retakes in both waves. More than 1/3 of all links in the first wave are links to students with retakes (37%). The share of the links to students with retakes in the first period in the total amount of second wave links is slightly smaller: 33%. It might be caused by the intention of students to improve their peer group and connect to peers with higher outcomes. The average amount of the friends with retakes in the first period is 1.83 while it is lower for the second period: only 1.25. The average amount of peers with exam retakes for the subsample of all students that have at least one peer with retake is higher than the average of the full sample and is equal to 2.5. For the same students in the second wave, the average number of peers who had exam retakes in the first period is now much lower: 1.55. It can be suspected that the decrease in this value may be partially explained by the readjustments of the network towards better connections. Moreover, for the same subsample, the average number of peers with retakes in the second period is even lower: 1.37. However, the difference between two periods is also determined by the decreased average number of friends, as well as dropouts of the students with the worst performance.

Some of those, who didn't have any friends with retakes in the first period, connected to new peers that had the retakes in the second period, the average number of such friends is only 0.35 though, but the average number of friends with retakes in the next period is 0.57. So the changes in the network are leading to the improvements as well as worsening of the new peer group.

Insert Table 3 here

Table 4 highlights that students in the studied framework tend to connect to peers, having higher average grades than the students themselves, for the full sample as well as for the samples with and without retake friends. Students, who do not connect to peers with retakes, are performing better than those, whose friends are having retakes. This suggests the existence of peer effects, since the better students tend to connect with the peers with similar performance. The improvements in the performance in the future are not significant, with the changes in the performance of the students without peers' retakes being slightly higher.

Insert Table 4 here

It is not possible to distinguish between the predicted and unexpected components of retakes by simply looking at the data. Therefore, the deeper econometric analysis is needed to make conclusions about the existence and the magnitude of the effect of unpredicted shock.

4 Model

Since a big share of the student's failure can be explained by observed components of the model, the failure should be treated as endogenous. To deal with arising identification issues, I am using the two-step model and estimation procedure proposed in Marchenko (2018) to estimate the effect of friends' failure on one's future performance. It uses the peer effect approach for the first period to capture the predictable component of the probability of the failure and takes the remained unpredicted part to estimate the effect of the shock on the future performance. The first step is as follows:

$$P(retake_i) = \alpha + \beta \sum_{j \neq i} G^1_{ij} y^1_j + \gamma X^1_i + \delta \sum_{j \neq i} G^1_{ij} X^1_j + \xi_i + \nu_i, \qquad \mathbb{E}[\nu_i | X^1] = 0 \quad (1)$$

where y^1_i and y^2_i are outcome variables of student i in the first wave and the second wave correspondingly. I consider average grade as an outcome variable in the empirical part. G^1_{ij} and G^2_{ij} are two adjacency matrices for the first and the second waves correspondingly, weighted by the number of links, and their entries have the value of $1/n_i$ if the link from student i to student j exists. Note that these matrices are not necessarily symmetrical, since the social network can be both directed (as in the empirical example) or undirected.

X_i is a vector of individual characteristics that should be controlled for, such as gender, city of origin, living conditions, some socioeconomic family characteristics, mandatory unified high-school exam results, etc.

ξ_i - student-level unobserved fixed characteristics, which may influence students' performance and choice of connections. It consists of the common for individual's connections unobservable component and individual's own unobserved fixed characteristics.

The unobserved individual characteristics also reflect homophily of the individuals, which may influence both link formation and network outcomes. In the case of interactions in the relatively big network, such as students' network local differences, proposed by Bramoullé et al. (2009), may be used to deal with the presence of correlated effects, together with the dynamic structure of the data. Both models with and without correlated effects will be considered throughout the analysis.

If assumption of no correlated effects is made, 1 is transformed as follows:

$$P(retake_i) = \alpha + \beta \sum_{j \neq i} G^1_{ij} y^1_j + \gamma X^1_i + \delta \sum_{j \neq i} G^1_{ij} X^1_j + \nu_i \tag{1a}$$

The second stage takes the average of the residuals of the equation (1) or (1a) (depending on the assumptions about presence of correlated effects): $UR_i = \sum_{j \neq i} G^1_{ij} \hat{\nu}_j$, which is used as an unexpected component of the shock in the following equation:

$$\Delta y_i = (\alpha_2 - \alpha_1) + \beta_2 \sum_{j \neq i} G^2_{ij} y^2_j - \beta_1 \sum_{j \neq i} G^1_{ij} y^1_j + \tilde{\boldsymbol{\delta}} UR_i + \gamma_2 X^2_i - \gamma_1 X^1_i +$$

$$+ \delta_2 \sum_{j \neq i} G^2_{ij} X^2_j - \delta_1 \sum_{j \neq i} G^1_{ij} X^1_j + \Delta\epsilon_i \tag{2}$$

This equation is obtained as a difference between the one period linear-in-mean peer effect on the level of GPA. The detailed discussion of the model is provided in Marchenko (2018).

$\tilde{\boldsymbol{\delta}}$ is the desired effect. It captures the influence of the unpredicted component of the friends' retakes on the changes in one's own performance.

Model in differences gives a better interpretation of the studied effect, additional to the elimination of individual fixed effect. It estimates the changes of student's performance in response to the failure of the friends additional to the changes of performance in comparison to the classmates, obtained by the single-period model.

Note that the coefficients for the endogenous peer effect and exogenous characteristics are considered to be different in two periods: β_2 and β_1 and δ_2 and δ_1. Students may experience different magnitude of the effects depending on how advanced they are in their studies, how well they are adjusted to the university environment, etc. Moreover, this also allows to take into account the changes in the network, since the students are experiencing the influence of two different peer groups in two periods.

The setting of the paper suggests the possible existence of the correlated effects that appear due to the similar individual characteristics within a subgroup, as most of the connections are formed inside of the same department, or even inside of the same exogenously formed study group. I conduct my analysis for both models with and without correlated effects. All the necessary modifications for the former are applied as proposed in Marchenko (2018).

I am applying the 2SLS estimation procedure of Marchenko (2018) in both of the stages with the friends' of the friends characteristics of both periods as instruments.

5 Network characteristics

Marchenko (2018) proves that the identification of the effect of friends' failures on the future performance is achieved under following assumptions:

1a) existence of intransitive triads in the network at all points in time, i.e. the existence of a set of three individuals i, j, k such that i is influenced by j, j is influenced by k, but i is not influenced by k.

1b) for the model with correlated effects there exist paths of length 3 in the network at all points in time, i.e. the existence of a set of four individuals i, j, k, l such that i is influenced by j, j is influenced by k, k is influenced by l, but i is not influenced by k and l directly, and j is not influenced by l.

2) existence of the exogenous changes of the friendship network, i.e. the changes not caused by the failure of the friend.

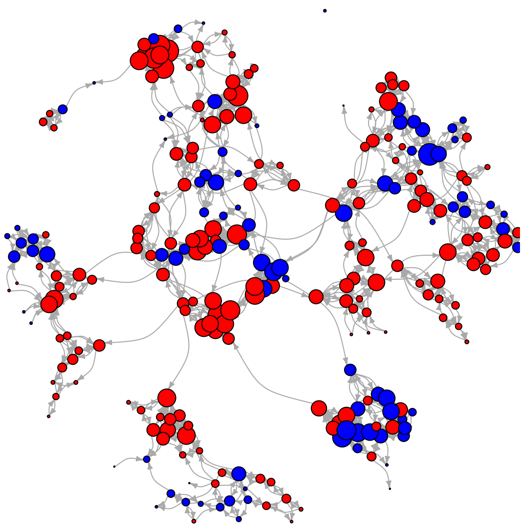
All of the assumptions are valid for most networks, in particular, for the sample analysed in this paper, which is discussed later in this section.

5.1 Network stability

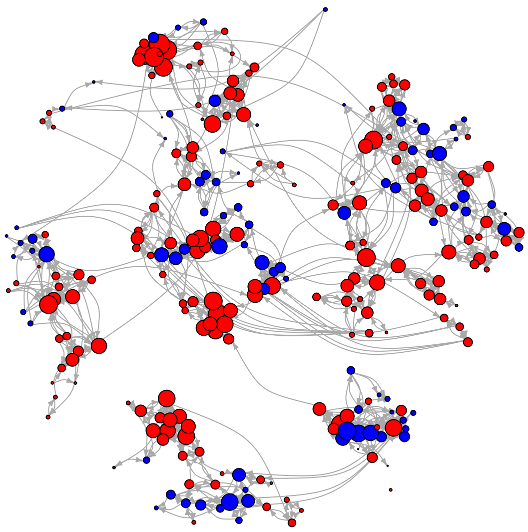
First, I discuss the variation in the network structure, on which the identification of the second step relies heavily. It is important that some changes in the network are exogenous, which is guaranteed by the natural processes of adjusting the friends circle: students are likely to learn more about their classmates with time, and the friendships, created during the first year, are often unstable.

Figure 1 visualizes the whole networks for the first wave (left) and for the second wave (right). Red nodes are females, blue - males, the size of the nodes is proportional to the overall degree of the node. It can be observed from this figure that two networks differ. For example, two clusters in the bottom part of the graph are not connected in the first wave, whereas there are several edges between them in the second wave.

More formal evidence about the variability of the network is presented in Table 5.



(a) Wave 1.



(b) Wave 2.

Figure 1: Networks.

Quite a lot of variation can be observed: around 11-12% of the students reported exactly the same set of friends. However, the share of completely new individual networks varies with gender. Females have only 5% of completely new set of friends. Hence, females tend to be more persistent in forming and retaining the links.

Insert Table 5 here

Table 6 provides more evidence of the network variation: only around 16% of the links survived after the first period, and around 78% of the links formed in the second period are new.

Insert Table 6 here

5.2 Transitivity

Now I provide evidence of the existence of intransitive triads.
Table 6 describes several characteristics of the networks in the sample. The transitivity is measured by the shares total amount of connected triangles in the whole graph. So in more than 50% of all possible sets of three students, at least, one link is missing.

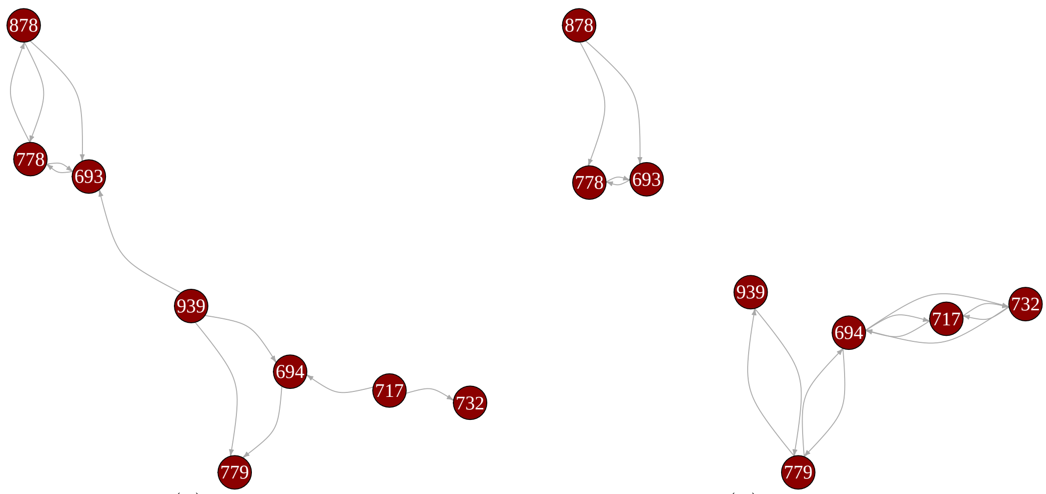
Figure2 shows the subgraph of the network to demonstrate the existence of intransitive triads in both of the samples. For example, in wave 1 the following triad is intransitive: $717 \rightarrow 694$, $694 \rightarrow 779$, but $717 \nrightarrow 779$. Other examples of intransitive triads are: 939→ 693→ 778, 693→778→ 878 in the first wave and 939→ 779→ 694, 779→ 694→ 717 in the second wave, and some more.

The characteristics of the networks (see Table6) also clearly suggest that the network is directed and cannot be assumed to be indirected since only around 60% of the links are reciprocal. Also, the networks are sparse with the density of the links around 1.5%.

6 Results

6.1 Main specification

I use the following variables for the main specification of the model:



(a) Wave 1. (b) Wave 2.

Figure 2: Subgraph of the network

Outcome: average weighted grade of the student in the corresponding period. The grades are summed up weighted by the amount of the credits assigned to the particular course.

Retakes: indicator of at least one retake in the first period.

Initial ability, measured as the sum of mandatory Unified State Examinations (mathematics and Russian) plus the sum of cross-products between these USE results and a dummy of winning any relevant Olympiads.

Controls: **time-invariant**, such as gender, socio-economic background like a dummy of parental higher education, a dummy of having a single parent before entering the university and dummy for siblings; and a set of dummies for three departments with law department serving as a base.

Controls: **time-varying**, such as tuition, which is mostly time-invariant, but some rare students change the type of tuition, working status (dummy for not working versus any type of job) and living conditions (dormitory versus everything else).

Descriptive statistics for these variables is provided in Table 7. It can be observed that the average changes in the time-variant variables are rather modest, as well as the changes in the performance. However, the average grade has higher standard deviation and spread in the second period.

Insert Table 7 here

Table 8 summarizes some of the findings of the estimation of the model without correlated effects. Note that the sample size is smaller than was discussed in the data description, due to the absence of some students in one of the waves. And it is critical to have the information in both waves for each of the students to estimate the effect.

Insert Table 8 here

It can be observed that for the full sample the estimator of the effect of the unpredicted component of retakes is negative but in most specifications insignificant. The magnitude of the effect in specification (5) suggests that if a friend of some student had a retake during the first year, which this student couldn't predict at all, the difference between the average grade in year 2 and the average grade in year 1 of the student will be on average 0.05-0.39 lower, than in the case the student expected the retake of a friend, depending on the total number of friends. For example, the median student on average improves her grades in the second period relative to the first by 0.24, which is 2.4% of the maximum grade. The presence of unpredicted retake of the friend, other things equal, may leave the average grade in the second year at the same level or even decrease it up to 1.5% of maximum grade, changing the direction of the dynamics and, moreover, putting the student on average 5-25 positions lower in the overall students' rating, falling lower with less friends.

Note, that there is a highly significant difference between the economics and other departments for most of the specifications. On average, students of economics department have -0.5 lower difference of grades, which suggest the overall lower grades of the economics department in the second year. This evidence indicates the necessity of using the model with correlated effects or treating the departments separately by splitting the full sample.

Discussing the results for those, who had their own retakes, versus those, who did not is the other possible way to improve the estimation results.

The further analysis is given in the next subsections, where I present the estimation results in the subsamples, of the model with correlated effects, as well as the estimation with a possibly improved network. However, it is worth pointing out, that the sample size for the main specification is 250 students, which may be not sufficiently big to capture the desired effect, and the results of the estimation in the subsamples should be treated with even more care, since with the lower sample size the asymptotic properties of the proposed estimator may suffer.

6.2 Connection to one's own retake

I first report the results for the subsamples of students with and without own retakes. It can be suggested that students who had their own retake may, in general, be connected to worse peers. Therefore, having friends with retakes might lower the performance even further, whereas the friends' retakes are more likely to have either no effect or even positive influence for the better students.

Insert Table 9 here

Table 9 gives hints that the outcomes are influenced differently by the peers in case of presence of one's own retake and in the case when the student passed all the exams from the first attempt. First of all, unexpected retake has an insignificant and negative effect of higher magnitude in case of own retake than without own retakes. So, when students in the network have retakes together, they will less likely improve in the future. It may be partially explained by the worse peer group, and partially by the fact that fewer friends are able to help to catch up with the courses after retakes. It can also be observed that the endogenous effect changes the sign, from negative to positive, and is more significant for students with own retakes, which may suggest that the students, especially the ones with their own retake tend to seek for the better peers in the future. However, the data does not provide evidence that the willingness to connect to better peers is coming from the discussed shock, therefore, the changes may be considered as a natural learning process.

6.3 Effects in different departments

In this subsection, I discuss the results for subsamples of different departments. I present the results for two departments: economics and management. The economics department showed significantly different results in comparison to the others in the main specification, and the management department is quite similar to the economics in the curriculum and direction of study.

Insert Table 10 here

As can be seen from Table 10, the discussed effect is surprisingly different for two departments. While specification (1) for the economic department have the negative effect of the unexpected retake, the same effect in the specifications for management is positive. However, estimators are not significant. Both subsamples have a small number of observations, which can cause the low significance of the effect of the interest, and the results should be treated with caution. It is possible to eliminate the differences between the departments and estimate the full sample, by exploring the model with correlated effects.

6.4 Estimation in presence of correlated effects

In this Subsection, I would like to discuss the results of the estimation proposed in Section 2.4.2. Simple estimation in the presence of correlated effects might lead to the biased results. Next table presents the summary of results, judging from which I can then compare the two specifications: with and without correlated effects.

Insert Table 11 here

Controlling for correlated effects leads to more significant and persistent value of negative effect of unexpected retakes than in the main specification. The magnitude of the effect in specification (5) suggests that if a friend has a retake during the first year, which the student couldn't predict at all, this will make the difference between the average grade of year two and the average grade of year one for this student on average 0.46 lower, if the student has only 1 friend, and approximately 0.065 lower, if the student has 7 friends. The maximum of the grades is 10 so that the person lose almost 5% of the maximum grade when the network includes friends with retakes.

6.5 Additional analysis

6.5.1 Improving network

As it was mentioned before, students were asked to name up to 7 friends from their cohort, although some named more than 7. However, it is reasonable to assume that all named friends are not equal for the person. I introduce two possible ways to account for better friends so that the quality of the network can be improved.

First, I assume that the friends named among the first are more important than the others, since they were remembered earlier, and the best friends can't be named last. I reduced the network, only taking up to three named first students. I conducted analysis for both models with and without accounting for correlated effects. The suggested improvement of the network didn't, however, increased the significance of the results[6]. The effect of an unpredicted component of friends' exam retake is not significantly different from zero. Therefore, it might be reasonable to conclude that the unexpected negative or positive performance of the whole network of friends is more important for the future performance of students than the performance of only best friends.

Second, I observe that about 60% of the network is reciprocal, so I conduct similar analysis limiting the network to only reciprocal connections. This again does not bring any improvement in terms of the significance of the studied effect. It seems that the students' performance is shaped not only by their mutual friends, but although by those, who don't consider them as friends, but are considered as friends by the students. These students may be viewed as a sort of role models, and therefore, are important to be taken into account.

Thus, the initial full network is able to capture the effect of unexpected shock better than the versions of the network, considered initially as possible improvements.

6.5.2 Important classes

The further analysis divides the subjects, studied by the students in the sample, into two parts: more important and less important. All subjects have the corresponding amount of ECTS credits, from 0 to 8 with average around 2.5. For the analysis, I set the threshold of 4 ECTS points. However, some subjects have several exams, for example, Mathematical Analysis, and the weight of some of the exam in the series can be lower than 4, but, at least, one exam has ECTS higher than 4. In these cases, I am including all the exams of the series in the sample of important exams. This restricts the set of the students with retakes to 2/3 of the initial set.

Table 12 provides the results of the analysis in the new setting for the model without correlated effects.

[6]The detailed results are presented in Tables A.3, A.4 in Appendix

Insert Table 12 here

It can be observed that a lot of the results resemble the results for the model with all retakes, however, the effect of the unpredicted retake is more significant when only important classes are taken into consideration. The sign of the estimator remains negative but it gains much more significance, suggesting the different effect that different classes may have on the future performance of the network. The results also suggest the higher magnitude than in the initial model. Now, the friend's unexpected retake of the important class may make the difference between average grades in two periods bigger and reduce the average grade of the second year additionally by up to 0.5, which equals to 5% of the maximum grade.

This result is expectable. For example, the new set of retakes does not include the class of Discrete Mathematics in the Economics department but includes Mathematical Analysis. These two classes differ not only in the amount of ECTS but also in the length and importance for the further classes. Mathematical Analysis is studied throughout the whole length of the first year, whereas Discrete Mathematics only for one term. Moreover, the former introduces a lot of methods used later in the core classes of the higher years, such as Micro or Macro, while the latter might be considered to contribute less in future studies. The full list of classes, which were retaken at least once and the subset of more important classes are presented in Table A.5 of Appendix.

The significance of a dummy of the Economics department suggests that the model with correlated effects may be used, as in the model with the full set of retakes. Surprisingly, the estimator of the effect of the unexpected retake in the model with correlated effects loses the significance once I restrict the set of the retakes.

6.6 Model validity

6.6.1 Predictability

The first step of the model predicts the probability of having the retake for each student. This prediction is made by fellow students. The model is, therefore, operating under the assumption that the students can predict the probability of the friends' retakes the same way the econometricians can. However, this assumption is not necessarily true. On the one hand, the students may now more about their friends' everyday life, personal characteristics and psychological mood than the econometricians. On the other hand, the students are likely not to have all the information about all the friends' friends, and their characteristics.

I check the alternative model specification, assuming no knowledge about the friends' friends on the first step. Only student's own characteristics are used to predict the probability of the retakes. The residuals obtained from the main model and from the simplified model have a similar distribution, however, they tend to differ for many of the observations, which leads to insignificant results on the second step. If one believe in the existence of discussed effect of the shock, this additional analysis suggests that the information about the connection is crucial for the analysis, hence, the initial model should be used.

6.6.2 Comparison on the threshold

In many cases, the difference between the students, who fail the exam and who pass the exam with the minimal possible grade is almost negligible. The students on both sides of the passing threshold often have the similar level of abilities and knowledge, and the failure of the exam at the end could depend on simple luck. The sample of students, who were close to failing the exam, can provide some additional evidence of the retakes' importance, as shocking events.

I repeat the main analysis, substituting the probability of obtaining the retake with the probability of getting 4 on the first step. If only the performance of the friends mattered, the effect of unpredicted part of the probability of getting 4 on the changes in the friends' future performance would be similar to the effect discussed in the paper. However, none of the specifications reveal the significant effect. The mistake in predicting the friends' exam passing probability does not influence the future performance, whereas the unpredicted retakes' probability on average decreases the friends' future performance. Hence, the true abilities of the friends, revealed in the grades, matter little in comparison to the fact of unexpected failure.

7 Conclusion

The influence of social connections on the performance of the peers is studied in many fields and is of particular importance in the economics of education. Most of the existing literature, however, studies the direct peer effect on performance, that is the instantaneous influence of the peers' performance. This paper chooses a different approach and looks at the deferred effect of the peers' outcomes, represented by the changes of the performance in response to the failure of the friends one period before. I discuss the spread of the unpredicted shock across the network of friends in the university environment using the newly introduced in Marchenko (2018) dynamic peer effect model in the presence of endogenous shock.

I take an advantage of both the structure of the network data and its dynamic nature. These features allow to guarantee the identification of the effect of the friends' exam failures. First, the presence of intransitive triads allows using the friends of the friends characteristics (or friends of the friends of the friends in model with correlated effects) as instruments for the friends' outcomes to deal with the possible simultaneity of the peers' influence. Variation of the network data, assuming that some of it is happening not due to the exam failures, allow the model to capture the changes in the network effect that is happening due to the failured.

Failured exams are important in determining the future of a student. Even when the student is allowed to stay in the university after failing the exam, accumulating the failures may highly likely result in dropping out of the university. However, a failure may influence not only the students with a failure but also the whole network of friends, as proven in the paper. In most cases, the effect is not large, but still should not be ignored by universities and researchers. When the threshold of failing the exam is too high, some students, viewed by their friends as high-achievers, are likely to fail too. This anticipation mistake leads to the decrease of the average grades of the whole friendship network. The evidence provided in this paper suggests that the unpredicted exam retakes of the friends will have a negative effect on the changes of the future performance of students. This effect is especially prominent when only the more important and relevant exams are included into the consideration.

The results of the paper rely on the particular institutional environment, and cannot be extended to the more general university framework without additional analyses of each particular case. Nevertheless, they should serve as an indication of the importance of taking the social network into account, when discussing the consequences of individual exam failures and establishing the thresholds for passing the examination.

References

Ammermueller, A. and J.-S. Pischke (2009). "Peer effects in European primary schools: Evidence from the progress in international reading literacy study". In: *Journal of Labor Economics* 27.3, pp. 315–348.

Androushchak, Gregory, Oleg Poldin, and Maria Yudkevich (2013). "Role of peers in student academic achievement in exogenously formed university groups". In: *Educational Studies* 39.5, pp. 568–581.

Angelucci, Manuela and Giacomo De Giorgi (2009). “Indirect effects of an aid program: how do cash transfers affect ineligibles’ consumption?” In: *The American Economic Review* 99.1, pp. 486–508.

Angrist, J.D. and K. Lang (2004). “Does school integration generate peer effects? Evidence from Boston’s Metco Program”. In: *American Economic Review*, pp. 1613–1634.

Apesteguia, Jose and Ignacio Palacios-Huerta (2010). “Psychological pressure in competitive environments: Evidence from a randomized natural experiment”. In: *American Economic Review* 100.5, pp. 2548–64.

Aral, S., L. Muchnik, and A. Sundararajan (2009). “Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks”. In: *Proceedings of the National Academy of Sciences* 106.51, pp. 21544–21549.

AUSTRALIA: High drop-out rates cost $1.4 billion (2010). URL: `http://www.universityworldnews.com/article.php?story=20101022203542738`.

Battin-Pearson, S., M.D. Newcomb, R.D. Abbott, K.G. Hill, R.F. Catalano, and J.D. Hawkins (2000). “Predictors of early high school dropout: A test of five theories.” In: *Journal of educational psychology* 92.3, pp. 568–582.

Baumeister, Roy F (1984). “Choking under pressure: self-consciousness and paradoxical effects of incentives on skillful performance.” In: *Journal of Personality and Social Psychology* 46.3, p. 610.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009). “Identification of peer effects through social networks”. In: *Journal of econometrics* 150.1, pp. 41–55.

Buchmann, Claudia and Ben Dalton (2002). “Interpersonal influences and educational aspirations in 12 countries: The importance of institutional context”. In: *Sociology of education* 75, pp. 99–122.

Calvó-Armengol, A., E. Patacchini, and Y. Zenou (2009). “Peer effects and social networks in education”. In: *The Review of Economic Studies* 76.4, pp. 1239–1267.

Carrell, S.E., R.L. Fullerton, and J.E. West (2009). "Does your cohort matter? Measuring peer effects in college achievement". In: *Journal of Labor Economics* 27.3, pp. 439–464.

Chandrasekhar, A. and R. Lewis (2011). "Econometrics of sampled networks". In: *Unpublished manuscript, MIT.[422].*

Coleman, James S et al. (1966). "Equality of educational opportunity." In:

Coleman, J.S. (1990). *Foundations of social theory.* Harvard university press.

Comola, M. and S. Prina (2014). "Do interventions change the network? A dynamic peer effect model accounting for network changes". In: *Working paper.*

De Giorgi, Giacomo, Michele Pellizzari, and Silvia Redaelli (2010). "Identification of social interactions through partially overlapping peer groups". In: *American Economic Journal: Applied Economics*, pp. 241–275.

Dieye, Rokhaya, Habiba Djebbari, and Felipe Barrera-Osorio (2014). "Accounting for peer effects in treatment response". In:

Dokuka, Sofia, Diliara Valeeva, and Maria Yudkevich (2015). "The Diffusion of Academic Achievements: Social Selection and Influence in Student Networks". In: *Higher School of Economics Research Paper No. WP BRP* 65.

Elhorst, J. Paul (2010). "Dynamic panels with endogenous interaction effects when T is small". In: *Regional Science and Urban Economics* 40.5, pp. 272 –282.

French, D.C. and J. Conrad (2001). "School dropout as predicted by peer rejection and antisocial behavior". In: *Journal of Research on adolescence* 11.3, pp. 225–244.

Gardner, Michael (2007). *GERMANY: Heavy cost of student drop-outs.* URL: `http://www.universityworldnews.com/article.php?story=20071025102357719`.

Garnier, H.E., J.A. Stein, and J.K. Jacobs (1997). "The process of dropping out of high school: A 19-year perspective". In: *American Educational Research Journal* 34.2, pp. 395–419.

Gaviria, A. and S. Raphael (2001). "School-based peer effects and juvenile behavior". In: *Review of Economics and Statistics* 83.2, pp. 257–268.

Goldsmith-Pinkham, P. and Guido W Imbens (2013). "Social networks and the identification of peer effects". In: *Journal of Business & Economic Statistics* 31.3, pp. 253–264.

Goux, D., M. Gurgand, and E. Maurin (2014). "Adjusting your dreams? The effect of school and peers on dropout behaviour". In: *IZA Discussion Paper No. 7948*.

Graham, Bryan S (2008). "Identifying social interactions through conditional variance restrictions". In: *Econometrica* 76.3, pp. 643–660.

Hanushek, E.A., J.F. Kain, J.M. Markman, and S.G. Rivkin (2003). "Does peer ability affect student achievement?" In: *Journal of applied econometrics* 18.5, pp. 527–544.

Hinke, Stephanie von, George Leckie, and Cheti Nicoletti (2019). "The Use of Instrumental Variables in Peer Effects Models". In: *Oxford Bulletin of Economics and Statistics*.

Hoxby, C. (2000). *Peer effects in the classroom: Learning from gender and race variation*. Tech. rep. National Bureau of Economic Research.

Indicators of economics of education, URL: `https://memo.hse.ru/ind`.

Jackson, Matthew O. (2010). "An overview of social networks and economic applications". In: *The handbook of social economics* 1, pp. 511–585.

Janosz, M., M. LeBlanc, B. Boulerice, and R. E. Tremblay (1997). "Disentangling the weight of school dropout predictors: A test on two longitudinal samples". In: *Journal of youth and adolescence* 26.6, pp. 733–762.

Kelejian, Harry H and Ingmar R Prucha (1998). "A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances". In: *The Journal of Real Estate Finance and Economics* 17.1, pp. 99–121.

Lee, Lung-fei (2003). "Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances". In: *Econometric Reviews* 22.4, pp. 307–335.

Li, Q., W. Zang, and L. An (2013). "Peer effects and school dropout in rural China". In: *China Economic Review* 27, pp. 238–248.

Lomi, A., Tom A.B. Snijders, Christian E.G. Steglich, and V.J. Torl (2011). "Why are some more peer than others? Evidence from a longitudinal study of social networks and individual academic performance". In: *Social Science Research* 40.6, pp. 1506 – 1520.

Lyle, David S (2007). "Estimating and interpreting peer and role model effects from randomly assigned social groups at West Point". In: *The Review of Economics and Statistics* 89.2, pp. 289–299.

– (2009). "The effects of peer group heterogeneity on the production of human capital at West Point". In: *American Economic Journal: Applied Economics* 1.4, pp. 69–84.

Manski, C.F. (1993). "Identification of endogenous social effects: The reflection problem". In: *The review of economic studies* 60.3, pp. 531–542.

Marchenko, M. (2018). "Dealing with endogenous shocks in dynamic network". In: *unpublished manuscript.*

McPherson, M., L. Smith-Lovin, and J. M. Cook (2001). "Birds of a feather: Homophily in social networks". In: *Annual review of sociology*, pp. 415–444.

Metco program official webpage. URL: http://www.doe.mass.edu/metco/.

Newcomb, M.D. (1996a). "Adolescence: Pathologizing a normal process." In: *The counseling Psychologist* 24, pp. 482–490.

– (1996b). "Pseudomaturity among adolescents: Construct validation, sex differences, and associations in adulthood." In: *Journal of Drug Issues* 26, pp. 477–504.

Nicoletti, Cheti and Birgitta Rabe (2019). "Sibling spillover effects in school achievement". In: *Journal of Applied Econometrics* 34.4, pp. 482–501.

Paula, Áureo de (2015). "Econometrics of Network Models". In: *cemmap Working paper* CWP52/15.

Poldin, Oleg, Diliara Valeeva, and Maria Yudkevich (2014). "Friendship and study assistance ties of university students". In: *Working Papers in Sociology, Higher School of Economics* 37.

– (2015). "Choice of specialization: do peers matter?" In: *Applied Economics* 47.44, pp. 4728–4740.

– (2016). "Which Peers Matter: How Social Ties Affect Peer-group Effects". In: *Research in Higher Education* 57.4, pp. 448–468.

Rumberger, R.W. (1983). "Dropping out of high school: The influence of race, sex, and family background". In: *American Educational Research Journal*.

– (2011). *Dropping out*. Harvard University Press.

Sacerdote, B. (2011). "Peer effects in education: How might they work, how big are they and how much do we know thus far?" In: *Handbook of the Economics of Education* 3, pp. 249–277.

Sacerdote, Bruce (2001). "Peer Effects with Random Assignment: Results for Dartmouth Roommates". In: *Quartetly Journal of Economics* 116.2, pp. 681–704.

Shugal, N. (2010). "Potoki obuchayushchikhsya v rossiyskoy sisteme obrazovaniya [Student flows in Russian education system]". In: *Voprosy obrazovaniya / Educational Studies* 4, pp. 122–148.

Smith, J.P. and R.A. Naylor (2001). "Dropping Out of University: A Statistical Analysis of the Probability of Withdrawal for UK University Students". In: *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 164.2, pp. 389–405.

Socio-Economic situation in Russia (2014). URL: `http://www.gks.ru/bgd/regl/b14_01/Main.htm`.

Stinebrickner, T. and R. Stinebrickner (2013). *Academic performance and college dropout: Using longitudinal expectations data to estimate a learning model*. Tech. rep. National Bureau of Economic Research.

Tinto, V. (1975). "Dropout from higher education: A theoretical synthesis of recent research". In: *Review of educational research*, pp. 89–125.

Triandis, Harry C (1989). "The self and social behavior in differing cultural contexts." In: *Psychological review* 96.3, pp. 506–520.

Tables

Table 1: Comparison of survey participants and population of students

		Sample	All students	Share
I, year 1	Size	320	432	74.07%
	Retake of at least one exam	157	203	77.34%
	Dropouts	16	40	40 %
I, year 2	Size	296	393	75.32%
	Retake of at least one exam	148	190	77.89%
	Dropouts	24	39	62.54%

Note: interviews were conducted at the end of each year of studies

Table 2: Characteristics of reported network links by sex

Variables	1st wave			2nd wave		
	Male	Female	All	Male	Female	All
Average size	4.53	5.19	4.96	3.57*	4.18*	3.93*
Average size including out of sample links	5.22	5.78	5.58	4.29*	4.96*	4.69*
Study group/department relation (% of network partners)						
Same study group	84.17	87.23	86.76	87.21	89.89	88.78
Same department	98.54	99.21	98.99	97.54	99.39	98.95
Individual characteristics of network partners(% of network partners)						
Same gender	64.05	81.97	76.18			
Same working status	62.43	70.33	67.78	50.74	60.95	56.41
Same education of mother	61.75	66.84	65.19	-	-	-
Same education of father	56.45	50.08	52.14	-	-	-
Same living conditions	57.59	46.71	50.23	50.97	39.61	43.33
Same living conditions (dorm/not)	84.14	76.23	78.79	74.27	70.55	72.16
Future plans (% of network partners)						
Same plans for Master	54.44	57.37	56.41	-	-	-
Same plans for Doctorate	47.18	47.32	47.27	-	-	-
Subsample of planning to do Master:						
Share of friends planning to do Master	68.34	72.42	74.46	-	-	-

*the network data in the 2nd wave is truncated at 7 friends

Table 3: Distribution of friends with retakes

Links formed in wave	1	2	2
Retakes in wave	1	1	2
Share of retakes links in all links	36.99%	32.99%	29.15%
Average number of friends with retakes	1.83	1.25	1.15
Subsample of students with at least one friend with retake in wave 1			
Average number of friends with retakes	2.5	1.55	1.37
Subsample of students without friends with retakes in wave 1			
Average number of friends with retakes	0	0.35	0.57

Table 4: Average grades in samples and subsamples

	Full sample	With retakes of friends	No retakes of friends
Average grade	7.04 (0.99)	6.98 (0.96)	7.37 (0.98)
Average grade of friends	7.18 (0.65)	7.03 (0.63)	7.68 (0.49)
Sample size	320	234	86
Average grade next period	7.13 (1.14)	7.02 (1.15)	7.44 (1.07)
Sample size	297	217	80

Note: the highest grade is 10, the lowest is 1, the lowest passing grade is 4.

Table 5: Overlap of network partners between wave 1 and wave 2 (in % of total number of individual networks)

Network statistics	Full sample	Male	Female
Complete overlap	11.49	11.21	11.89
No new links	24.66	22.43	26.49
Partial overlap	65.20	46.73	77.30
Complete turnover	12.16	24.30	5.41
Observations	296	107	185

Note: Percentages of 1st, 3rd, and 4th rows do not add up to 100%, because there are new observations in the 2nd wave, for which we do not observe the network in the 1st wave

Table 6: Some network characteristics

Network statistics	Definition	1 year	2 year
Average indegree	Average number of ingoing ties	4.96 (2.73)	3.93 (2.53)
Average outdegree	Average number of outgoing ties	4.96 (2.01)	3.93 (2.2)
Density	Proportion of existing ties in the network	0.015	0.014
Reciprocity	Proportion of ties which are reciprocated	0.639	0.636
Transitivity	The ratio of the triangles and the connected triples in the graph	0.454	0.443
Share of the links that remained from the 1st wave in total amount of links of the 2nd wave		-	22.61%
Share of the links that remained from the 1st wave in total amount of links of the 1st wave		16.57%	-

Table 7: Descriptive statistics

Variable	Mean	St.Dev.	Min	Max
Average grade, wave 1	7.20	0.94	4.58	9.35
Average grade, wave 2	7.23	1.13	4.50	9.86
Retakes (dummy)	0.33	0.47	0	1
Retakes (number)	0.684	1.25	0	6
Ability	183.6	70.09	106	355
Gender (female)	0.67	0.47	0	1
Tuition, wave 1 (private=1)	0.18	0.38	0	1
Tuition, wave 2 (private=1)	0.184	0.39	0	1
Economics department	0.328	0.47	0	1
Management department	0.272	0.45	0	1
Computer Science department	0.26	0.44	0	1
Working status, wave1 (not working=1)	0.804	0.39	0	1
Working status, wave2 (not working=1)	0.74	0.44	0	1
Higher Education of mother	0.796	0.4	0	1
Higher Education of father	0.624	0.49	0	1
Single parent family	0.2	0.40	0	1
Family with more than 1 kid	0.54	0.50	0	1
Living conditions, wave 1 (dormitory=1)	0.16	0.37	0	1
Living conditions, wave 2 (dormitory=1)	0.172	0.38	0	1

Table 8: Estimation of main specification, dependent variable: change of the average student grade

Variable	(1)	(2)	(3)	(4)	(5)
Constant	-0.1521		-0.1840		-0.0482
Unexpected Retake	-0.2638	-0.2143	-0.3077$^{\bullet}$	-0.2064	-0.3907*
Endogenous effect, period 1	-0.0307	-0.0425	-0.0317	0.0908*	0.0614*
Endogenous effect, period 2	0.0205	0.0085	0.0218	0.0419	0.0306
Time-variant own exogenous characteristics					
Tuition, w1	0.0208			0.0102	
Tuition, w2	-0.0912			-0.1518	
Working status, w1	-0.0664	-0.0719	-0.0716		
Working status, w2	0.1381$^{\bullet}$	0.1147*	0.1346$^{\bullet}$		
Living in dorm, w1					0.1061
Living in dorm, w2					0.1651
Friends' exogenous characteristics					
Economics, w1	0.2417	0.1568	0.2692	-0.0732	
Economics, w2	-0.4681**	-0.4367**	-0.4513**	-0.5893***	
Management, w1	0.5409*		0.5712*		
Management, w2	0.1790		0.1996		
Working status, w1				-0.7352*	-0.5420**
Working status, w2				-0.0497	-0.1903
HE of father, w1		0.4010$^{\bullet}$			
HE of father, w2		0.0006			
Sample size	250	250	250	250	250
BIC	-216.68	-225.24	-226.51	-225.79	-196.71

*** - p-value < 0.01, ** - p-value < 0.05, * - p-value < 0.1, $^{\bullet}$ - p-value < 0.15

Table 9: Estimation in subsamples of students with and without their own retake

Variable	(1), yes	(2), yes	(1), no	(2), no
Constant	0.2602	0.1027	-0.1001	-0.3734*
Unexpected Retake	-0.2092	-0.1788	0.0246	0.0586
Endogenous effect, period 1	-0.0237	-0.0539	0.0352	-0.0262
Endogenous effect, period 2	0.0756**	0.0671$^\bullet$	0.0429	0.0577$^\bullet$
Time-variant own exogenous characteristics				
Tuition, w1	-0.1612		0.0032	
Tuition, w2	-0.1834		-0.2685	
Working status, w1		-0.1175		-0.0713
Working status, w2		0.1379		0.1192
Friends' exogenous characteristics				
Economics, w1	-0.1964	-0.0474	0.0616	0.1081
Economics, w2	-0.9973***	-0.9131***	-0.3932	-0.5344$^\bullet$
Management, w1				0.3098
Management, w2				-0.0338
Working status, w1			-0.5632**	
Working status, w2			-0.2305	
HE of father, w1	0.6306$^\bullet$	0.3753		
HE of father, w2	-0.3627	-0.4972$^\bullet$		
Dummy siblings, w1		0.7689***		
Dummy siblings, w2		0.2113		
Sample size	83	83	167	167
BIC	-336.10	-348.14	-288.34	-290.82

*** - p-value < 0.01, ** - p-value < 0.05, * - p-value < 0.1, $^\bullet$ - p-value < 0.15

Table 10: Estimation in subsamples of students in economics and management departments

Variable	(1), Econ.	(2), Econ.	(3), Man.	(4), Man.
Constant	-0.5228*	-0.2265	0.5614*	0.7032**
Unexpected Retake	-0.4375	-0.4794	0.4043	0.3943
Endogenous effect, period 1	-0.0426	-0.0150	0.2884***	0.3927***
Endogenous effect, period 2	0.0334	0.0544	0.1635**	0.2164**
Time-variant own exogenous characteristics				
Tuition, w1		0.2538		-0.4889
Tuition, w2		-0.0479		-0.7778*
Working status, w1	-0.0888		0.0054	
Working status, w2	0.2119		0.0880	
Friends' exogenous characteristics				
Ability, w1			-0.0056*	-0.0062**
Ability, w2			-0.0049**	-0.0048**
Gender, w1		1.0102*		
Gender, w2		0.2889		
Working status, w1		-0.6228		-0.6353*
Working status, w2		-0.5707		-0.4209
HE of mother, w1			-0.4308	-0.7535*
HE of mother, w2			-0.3144	-0.5867*
Dormitory, w1	-1.5248***	-0.8094•		
Dormitory, w1	-0.9558**	-0.8145*		
Dummy siblings, w1	0.4005			
Dummy siblings, w2	-0.3945			
Sample size	82	82	68	68
BIC	-305.45	-300.61	-456.68	-471.57

*** - p-value < 0.01, ** - p-value < 0.05, * - p-value < 0.1, • - p-value < 0.15

Table 11: Estimation of specification with correlated effects

Variable	(1)	(2)	(3)	(4)	(5)
Unexpected Retake	-0.4144$^\bullet$	-0.3899$^\bullet$	-0.3817$^\bullet$	-0.3817$^\bullet$	-0.4616*
Endogenous effect, period 1	-0.0361	-0.0526	-0.0378	-0.0379	0.0186
Endogenous effect, period 2	0.0143	0.0016	0.0544	0.0544	0.0461
Time-variant own exogenous characteristics					
Tuition, w1		0.0834	0.0411		
Tuition, w2		-0.1011	-0.1292		
Working status, w1	-0.0382	0.0266		0.0411	0.0590
Working status, w2	0.1077	0.1355$^\bullet$		-0.1292	0.0991
Living conditions, w1	-0.1323				
Living conditions, w2	0.2102				
Friends' exogenous characteristics					
HE of mother, w1	-0.6547	-0.5532	-0.5785	-0.5785	-0.6717
HE of mother, w2	-0.2011	-0.1541	-0.3396	-0.3396	-0.3875
HE of father, w1	0.5325	0.5789	0.4663	0.4663	
HE of father, w2	-0.0167	0.0378	-0.0831	0.3817	
Sample size	250	250	250	250	250
BIC	-183.89	-185.56	-195.56	-192.25	-197.99

*** - p-value < 0.01, ** - p-value < 0.05, * - p-value < 0.1, $^\bullet$ - p-value < 0.15

Table 12: Estimation with retakes in courses with ECTS 4 and higher

Variable	(1)	(2)	(3)	(4)	(5)
Constant		-0.2176	-0.1670		-0.2005
Unexpected Retake	-0.4912**	-0.5484***	-0.5158**	-0.4907**	-0.5564**
Endogenous effect, period 1	0.1072*	-0.0211	-0.0160	0.1076*	-0.0158
Endogenous effect, period 2	0.0378	0.0279	0.0284	0.0401	0.0307
Time-variant own exogenous characteristics					
Tuition, w1	0.0417		0.0430	0.0530	
Tuition, w2	-0.0861		-0.0575	-0.0830	
Working status, w1		-0.0568	-0.0570		-0.0616
Working status, w2		0.1488*	0.1494*		0.1469*
Living conditions, w1	-0.0222	-0.2673			
Living conditions, w2	0.0312	-0.1808			
Friends' exogenous characteristics					
Economics, w1	-0.1032	0.1652	0.1257	-0.1129	0.1387
Economics, w2	-0.6177***	-0.5279**	-0.5466**	-0.6240***	-0.5411**
Management, w1		0.4322	-0.4049		0.4159
Management, w2		0.1209	0.1045		0.1127
Working status, w1	-0.8120*			-0.8186**	
Working status, w2	-0.0074			-0.0102	
Sample size	250	250	250	250	250
BIC	-215.86	-220.76	-220.81	-226.98	-230.99

*** - p-value < 0.01, ** - p-value < 0.05, * - p-value < 0.1, • - p-value < 0.15

Appendix

Additional tables

Table A.1: Distribution of the number of friends in samples

# of friends	Long study, wave 1		Long study, wave 2		Short study	
0	17	5.29%	26	8.12 %	9	4.39 %
1	1	0.31%	14	4.37%	3	1.46 %
2	5	1.56%	34	10.62%	4	1.95 %
3	28	8.72%	39	12.19%	17	8.29 %
4	32	9.97%	56	17.5%	28	13.66 %
5	39	12.15%	55	17.19%	34	16.59 %
6	41	12.77%	39	12.19%	34	16.59 %
7	150	46.73%	33	10.31%	28	13.66 %
8	2	0.62%	0	0.00%	21	10.24 %
9	1	0.31%	0	0.00%	14	6.83 %
10	3	0.93%	0	0.00%	4	1.95 %
11	0	0.00%	0	0.00%	3	1.46 %
12	0	0.00%	0	0.00%	1	0.49 %
13	1	0.31%	0	0.00%	3	1.46 %
14	0	0.00%	0	0.00%	2	0.98 %

Table A.2: Unified State Exams statistics

Subject	Number of participated	Average grade
Mathematics	305	59.87
Russian	305	79.85
Biology	2	71.5
Chemistry	1	80
Computer Science	49	76.96
Economics	27	32.52
Foreign Language	272	70.64
Geography	4	67
History	78	70.94
Law	20	69.4
Literature	20	69.35
Orientalism	2	75
Physics	49	58.45
Social Studies	269	71.01

Note: the highest grade is 100.

Table A.3: Results for the model with reciprocal links and the model with best friends, no correlated effects

Variable	Recipr., (1)	Recipr., (2)	Best, (3)	Best, (4)
Constant		-0.2318$^{\bullet}$		-0.3127**
Unexpected Retake	0.1320	-0.0097	0.0469	0.0768
Endogenous effect, period 1	0.0180	0.0111	0.0231	-0.0467**
Endogenous effect, period 2	0.0480*	0.0407*	0.0818***	0.0215
Time-variant own exogenous characteristics				
Tuition, w1	0.0317		-0.0771	
Tuition, w2	-0.1547		-0.2518	
Working status, w1		-0.0909		-0.1235
Working status, w2		0.1483*		0.1510*
Friends' exogenous characteristics				
Economics, w1	0.0778	-0.0214	0.0082	0.1953
Economics, w2	-0.4701**	-0.5337**	-0.4869**	-0.4096**
Computer Science, w1		-0.4977**		
Computer Science, w2		-0.3899$^{\bullet}$		
Working status, w1	-0.2272		-0.1881	
Working status, w2	-0.1623		-0.4340**	
Siblings, w1				0.2611*
Siblings, w2				-0.0080
Sample size	250	250	250	250
BIC	-224.67	-221.77	-221.41	-218.07

*** - p-value $<$ 0.01, ** - p-value $<$ 0.05, * - p-value $<$ 0.1, $^{\bullet}$ - p-value $<$ 0.15

Table A.4: Results for the model with reciprocal links and the model with best friends, with correlated effects

Variable	Recipr.,(1)	Recipr.,(2)	Best,(3)	Best,(4)
Unexpected Retake	-0.1212	-0.0913	0.0406	-0.1365
Endogenous effect, period 1	-0.0081	-0.1188	0.0235	-0.0404
Endogenous effect, period 2	0.0498	0.0016	0.0811	-0.0262
Time-variant own exogenous characteristics				
Tuition, w1		0.0394		0.1946
Tuition, w2		-0.1933		-0.0608
Working status, w1	0.0296		-0.0248	
Working status, w2	0.0493		0.1437	
Friends' exogenous characteristics				
Abilities, w1		-0.0004		
Abilities, w2		-0.0029		
Tuition, w1	-0.5923			0.2977
Tuition, w2	-0.4462*			0.1171
Economics, w1			1.0059	
Economics, w2			-0.2755	
HE of mother, w1				-0.3578
HE of mother, w2				-0.2662
Single Parent, w1	-0.0083			
Single Parent, w2	-0.1985			
Siblings, w1			0.2174	
Siblings, w2			0.1179	
Sample size	250	250	250	250
BIC	-191.02	-192.71	-189.98	-183.06

*** - p-value < 0.01, ** - p-value < 0.05, * - p-value < 0.1, • - p-value < 0.15

Table A.5: List of classes with retakes in the sample

Class	Department	Total No. of retakes	ECTS 4 and higher
Algebra	Computer Science	21	No
Architecture of Computer Systems	Computer Science	2	No
Architecture of ECM	Computer Science	1	No
Basics of computer technology and programming	Computer Science	3	Yes
Discrete Mathematic	Computer Science	8	No
Discrete Mathematic	Economics	2	No
Economic Theory and Institutional Analysis	Management	28 (in 2 terms)	Yes
Economic Theory and Institutional Analysis	Computer Science	12	No
Economic Theory Basics	Economics	27 (in 3 terms)	Yes
Economics	Computer Science	3	No
English and other languages	All departments	9	No
Geometry and Algebra	Computer Science	8	Yes
History of economic thoughts	Economics	1	Yes
History of foreign state and law	Law	2	Yes
Introduction to software engineering	Computer Science	3	Yes
Judicial power and law enforcement	Law	1	No
Life safety	All departments	3	No
Linear Algebra	Economics	28	No
Mathematical Analysis	Computer Science	68 (in 2 terms)	Yes
Mathematical Analysis	Economics	12	Yes
Mathematics	Management	31 (in 2 terms)	Yes
Methods of financial and economic computations	Economics	1	No
Microeconomics	Computer Science	18 (in 2 terms)	Yes
Philosophy	Management	6	Yes
Roman Law	Law	1	No
Socio-Economic Statistics	Economics	2	No
Sociology	Management	1	Yes
Theoretical basics of computer technology	Computer Science	9 (in 2 terms)	No
Theory of state and law	Law	4	Yes