NATIONAL RESEARCH UNIVERSITY HIGHER SCHOOL OF ECONOMICS
INSTITUTE OF EDUCATION

Ivan Smirnov

Students' differentiation by academic performance on a social networking site

Summary of the PhD thesis

Translated from Russian

Academic supervisor:
Prof. Katerina Polivanova
Doctor of Science

Moscow — 2018

The thesis was prepared at the National Research University Higher School of Economics.

Publications:

1. Smirnov I., Sivak E., Kozmina Y. In Search of Lost Profiles: The Reliability of VKontakte Data and Its Importance for Educational Research // Educational Studies. 2016. №4. C. 106-122.
2. Polivanova K., Smirnov I. What's in My Profile: VKontakte Data as a Tool for Studying the Interests of Modern Teenagers. // Educational Studies. 2017. №2. C. 134-151.
3. Smirnov I., Thurner S. Formation of homophily in academic performance: Students change their friends rather than performance. // PloS one. Vol. 12. №. 8, e0183473.
URL: https://doi.org/10.1371/journal.pone.0183473
4. Smirnov I. Predicting PISA scores from students' digital traces // Proceedings of the Twelfth International Conference on Web and Social Media. American Association for Artificial Intelligence (AAAI) Press, 2018. P. 360-364

Selected Conferences:

— RAHER — VII International Conference of Higher Education Researchers, Moscow, Russia, 2016.

— XVIII April International Academic Conference on Economic and Social Development, Moscow, Russia, 2017.

— NetSci-X – International Conference on Network Science, Tel Aviv, Israel, 2017.

— IC2S2 — International Conference on Computational Social Science, Cologne, Germany, 2017.

— SocInfo — International Conference on Social Informatics, Oxford, United Kingdom, 2017.

— RAHER — VIII International Conference of Higher Education Researchers, Moscow, Russia, 2017.

— XIX April International Academic Conference on Economic and Social Development, Moscow, Russia, 2018.

— ICWSM — International AAAI Conference on Web and Social Media, Stanford, USA, 2018.

INTRODUCTION

Scientific tools have always played an important role in the scientific progress. The invention of the telescope opened a new vast world of celestial bodies. Observations of the planetary motion laid the foundation of the classical mechanics and gave the rise to physics as a science in the modern sense. The invention of the microscope opened a new vast world of microorganisms and laid the foundation of the modern biology and medicine. Today, we observe comparable changes in the social science [1–3]. In place of optical instruments are methods of collecting and analyzing digital traces, vast amounts of data made available by the recent development of information technology.

Digital traces allows researchers to see what was previously hidden from them. For instance, educational researchers have long been studying social relationships between students [4–9]. However, for a long time, the only method for collecting the data on students' social ties was survey. Surveys only allow reconstructing the social network of a small closed group of people. As a result, researchers well understand the social networks of students within one school but know almost nothing about interschool social ties. In contrast, Facebook has information about friendship ties of two billion people[1]. Information about social ties need not necessarily be explicitly indicated, as in the case of friendship on a social networking site. It can effectively be reconstructed from phone calls logs [10] or from information about physical proximity from mobile sensors [11]. The direct access to information about human behavior also resolves the problem of censoredness and social acceptability characteristic of survey answers [12].

Another advantage of digital traces is a significant reduction in the cost of research. The largest international study on education PISA covers about 500,000 students [13]. Its organization requires a large amount of resources and coordinated effort from dozens of teams in different countries. At the national level, studies with several thousand participants are considered as large. The thesis could be successfully completed with hundreds or even dozens of respondents. Data from social networking sites allows individual researchers to work with much larger samples up to tens of millions of participants [14–16]. Data need not necessarily be accessible directly. It was previously shown that various socio-demographic characteristics such as ethnicity, gender or income level could be effectively predicted from tweets [17], profile images [18], user posts [19], or photos of neighborhoods [20].

Traditionally, social sciences make an emphasis on theories and their ability to explain human behavior while ignoring their predictive power which is often much inferior to natural

[1] https://newsroom.fb.com/company-info/

sciences [21, 22]. This significantly limits the possibility of practical application of such knowledge. One of the reasons for the weak predictive power lies in the fact that social systems and the laws that govern them are too complex to be described by simple variables [23]. For instance, a variable with a few discrete values cannot describe the wide variety of personal characteristics and human behavior even in theory. Today, thanks to big data, one could construct new variables with much greater predictive power. For instance, it has been shown that information about the likes on a social networking site allows to predict various personal traits from sexual preferences to intelligence and parents' divorce [24]. While individual likes do not have much predictive power, together they form a unique digital trace that contains a significant amount of information about an individual.

Thus, the use of digital traces for studying the human behavior is one of the most promising and actively developing areas in the social science. This approach has a great potentiation in the education science as well. In the Russian context, the most promising source of data is the social networking site VK, that is already been used in educational research [25–27]. VK is of particular interest not only because it provides an access to a vast amount of data but also because of its role in the lives of modern students.

Today, young people are almost always online [28, 29] and spend most of their online time on SNSs [30]. Recently SNSs have become the main source of traffic to other resources of the Internet [31]. In Russia, the largest SNS VKontakte (VK) is named as the main source of information about the country and the world by 70.3% of young people, more than any other source of information [32]. This SNS is also more trusted by young people than traditional media [32]. Thus, VKontakte largely determines the information environment of modern Russian students. That is why it is particularly important to study the nature of this environment and to understand whether it is different for different groups of students.

This work is devoted to the development of methods for mining and analyzing VK data in the context of educational research. These methods are applied to study differentiation of students by academic performance. The need to develop the methods is due to the fact that little is known about the reliability of the data and the potential biases of VK samples and that analysis of big multidimensional data requires special approaches. The differentiation of students on a social networking site is of particular importance in the context of digital inequality [33–35].

Theoretical framework

The authors of the concept of digital inequality note that one generalization that emerges from research on inequality in access to information about cultural resources is the differentiation principle. This principle means that at the beginning — when access to a particular resource is limited — the use of this resource remains relatively undifferentiated. When the resource becomes more accessible, its use becomes more differentiated. It is explained by the fact that more-privileged groups find a way to use the resource to extract more from it than less-privileged groups [33].

The typical example here is education. When access to secondary education became almost universal, most children from the middle and upper classes started to attend universities. At that point, university degrees became the main differentiating characteristic. When higher education became widespread, two groups of universities emerged: mass universities and selective universities, and the main differentiating characteristic shifted from having any higher education degree to graduation from the prestigious university [33].

The same principle applies to digital resources. In theory, the global network provides all students with equal access to virtually unlimited knowledge. However, as early as the mid-1990s, researchers became concerned about the digital divide between those who have or do not have access to computers (later to the Internet) [36]. When the Internet became ubiquitous, the concern was about the second digital gap related to the gap in the level of information literacy and the ability to use all the opportunities that are provided by the Internet. Today, the main focus is shifted to a broader concept of digital inequality, which is understood as the difference in the Internet use (or the use of any other digital technology) between groups of people who have formally equal access to it [33].

The existing research [37, 38] suggests that there are significant differences in the Internet use between users with different levels of formal education. More educated people often use the Internet for capital-enhancing activities (such as job seeking, seeking for medical or professional advice, and acquiring new knowledge) while lower educated people use it predominantly for entertainment [37, 38]. It was also found that students with low socioeconomic status enroll less often in online courses and if they are enrolled then they successfully complete it less often than their peers [39, 40].

In most cases, the comparison is made between broad socio-demographic groups (such characteristics as gender, level of education, etc. are analyzed [41]). The information about the use of the Internet is obtained from surveys. In our work, we suggest using much more detailed data namely the information about the academic performance of students and detailed information about their actual behavior on a social networking site.

Aim and objectives of the study

The aim of the study is to develop methods for mining and analyzing VK data and to apply these methods to understand the nature of the relationship between students' behavior on the SNS and their academic achievements. To achieve this goal, the following objectives were set:

— To develop a methodology for extracting reliable data from VKontakte and combining it with educational data; identify possible biases in the sample of VKontakte users.

— To study the structure of students' online ties and its dependence on the structure of educational organizations; to analyze the structure of online ties between students of different educational organizations and its dependence on the geographical distance between them.

— To determine whether there is a relationship between students' online ties and their academic performance, and to study the evolution of this relationship over time.

— To identify and analyze the relationship between the online interests of students and their academic performance.

Methodology and design of the study

For the purposes of the study, we developed a software program that performs queries to the VKontakte API and receives a list of all users of a certain age who indicated that they are studying at a given educational organization. The program matches and merges information from user profiles with information about students from the educational organization. After the merge, the data is anonymized and stored for later research.

A direct matching by first and last name allows identifying only a small proportion of the students. That is why we additionally created an extensive dictionary of various forms of the same name (for example, Ivan-Vanya-Vanyusha), and also developed an algorithm for matching names that are written in Latin and Cyrillic. The program searches not only among users who indicated that they are studying in a given educational organization but also among their friends on the social network.

In the case when information about students was not available, we excluded from the sample those users who did not have friends from their own educational organization. This approach allows one to efficiently filter out profiles with inaccurate information.

We use this method to collect four unique data sets that combine educational data with data on the online behavior of students.

— “School”, a complete sample of students from a Moscow school from the 5th to 11th grades (N = 766)

— “University”, a complete sample of students from HSE University (N = 15 757)

— “City”, a complete sample of users who indicated that they are studying or have graduated from one of Saint Petersburg’s schools (N = 1 742 392)

— “Country”, a representative Russian sample of students who were born in 1995–1997 (N = 4 893)

The latest sample includes participants of a longitudinal panel study “Trajectories in education and career” who agreed to provide their personal information for research purposes [42].

To study the relationship between the structure of friendship ties and the structure of educational organizations, we compute modularity Q for partitions of the entire network corresponding to the partition of a school into grades and buildings and the partition of a university into years of studies, educational programs, and campuses. To ensure that observed values of Q cannot be explained by random noise, we use a permutation test. We fixed the structure of a network and randomly permute the attributes of nodes (i.e. school grades, educational programs, etc.). We then computed Q_{rand} for such a random network. Comparison of 10 000 generated Q_{rand} values with Q allows us to compute the significance level for the observed results. To study the relationship between the structure of friendship ties with geographical distances, we analyze the relationship between the probability of the friendship tie between two schools and the distance between them.

To investigate the detailed evolution of social ties within an educational organization, we use information about the interaction between students instead of information about their friendship ties. We split the period of the study into 3-month intervals and then for each of the intervals we built a network of social ties. We assume that the tie between two users exists if one of them gives at least one “Like” to another on VKontakte during this period. We then compute the Pearson correlation coefficient between the academic performance of students and the average academic performance of students connected to them. This approach allows us not only to determine the level of differentiation of social ties by academic performance but also to track its evolution with time.

The level of social ties differentiation on a city scale was estimated by computing a correlation coefficient between the average USE scores of graduates of a school and the average USE scores of graduates from schools with which this school is connected on the SNS. Two schools are considered connected on SNSs if there is at least one friendship tie between their students on such sites. To ensure that any observed effect is not driven by geographical locations of schools, we use a random graph model that preserves the relationship between the probability of a friendship tie between schools and geographical distance between them.

We analyze the interests of VKontakte users using information about their subscriptions to various public pages (groups) on the SNS. We compute the average GPA for subscribers to different pages along with their average age and proportion of females among them. To show that there is a correlation between interests and academic performance, we build a model that can predict academic performance from information about subscriptions. We identify the main components containing information about students’ interests and then use them in a linear regression model to predict academic performance.

Main results

Reliability of VKontakte data

Our approach to extracting information about the online behavior of students and combining it with educational data [43] allows us to significantly increase the total coverage of students in comparison to the direct matching of profiles (Table 1). We identify 88% of students from the “School” sample, 93% of students from the “University” sample and 82% of students from the “Country” sample. We do not find any significant differences between identified and not identified students from the “School” sample in academic performance, age, or gender (Tables 2, 3). Similar results are obtained for the “University” sample [44].

Table 1. Percentage of students whose VK profiles were identified using the proposed methods.

		Dictionary of first names with alternative forms	
		No	Yes
Friends list	No	18%	27%
	Yes	57%	88%

Table 2. Percentages of identified VK users who did not indicate their school and/or used alternative forms of their names, by age (grade).

	Grade						
	5th	6th	7th	8th	9th	10th	11th
Percentage of students identified	85%	89%	88%	90%	88%	91%	85%
Percentage of students who did not indicate their school	64%	72%	69%	74%	70%	58%	72%
Percentage of students who used alternative forms of their names	39%	36%	29%	33%	33%	31%	38%

Table 3. Groups of school students differing in the way of presenting their personal data on VK, by gender and academic performance.

	Girls ratio	GPA
Identified on VK	46%	3.80
Not identified on VK	48%	3.79
Those who did not indicate their school	48%	3.77
Those who used alternative forms of their names	50%	3.79

For the “Country” sample, the proportion of identified users does not depend on city size (Table 4), with the only exception of rural-type settlements where the proportion of females is significantly lower.

Information about friendship ties allows to effectively filter out fake profiles. By one of our estimates, the number of fake profiles could be reduced from 66% to 8% on the “School” sample if users who do not have friends from this school are excluded.

Table 4. Percentages of identified VK users for different city sizes.

	Females	Males
Rural-type settlement	68%	83%
Urban-type settlement	86%	93%
Cities with less than 50 000 people	87%	90%
Cities with 50 000 – 100 000 people	81%	85%
Cities with 100 000 – 450 000 people	81%	88%
Cities with 450 000 – 680 000 people	83%	86%
Cities with more than 680 000 people	79%	84%
Saint Petersburg	83%	87%
Moscow	78%	73%

Structure of social ties in the digital space

We find that the structure of social ties on VK reproduces the structure of educational organizations, including the division of a school into grades, Q = 0.47, (Figure 1a) and buildings, Q = 0.35, (Figure 1б), and the division of a university into campuses, Q = 0.32, years of studies, Q = 0.58, and educational programs, Q = 0.68 (Figure 2). In all cases, the observed values of Q are statistically significant with p-values < 10^{-4} (permutation test).

Figure 1. Structure of school social ties. Nodes correspond to students and links to friendships on VK. Different colors correspond to grades from 5th to 11th (a) and school buildings (б).

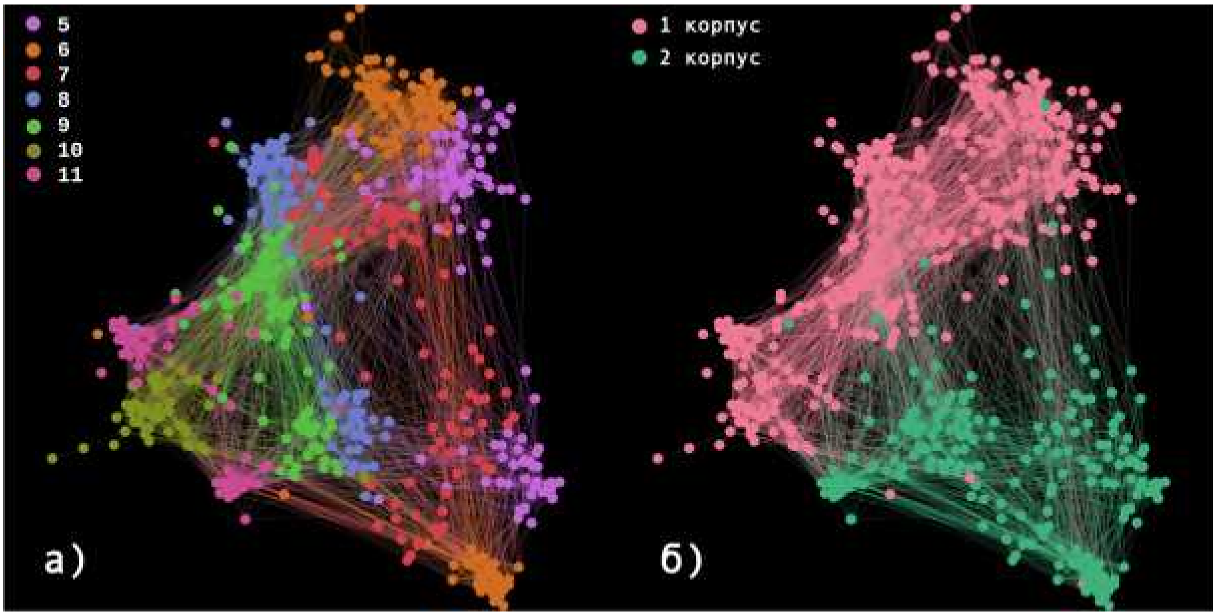


Figure 2. Structure of university social ties. Different colors correspond to years of studies. Visible clusters inside each year correspond to different educational programs.

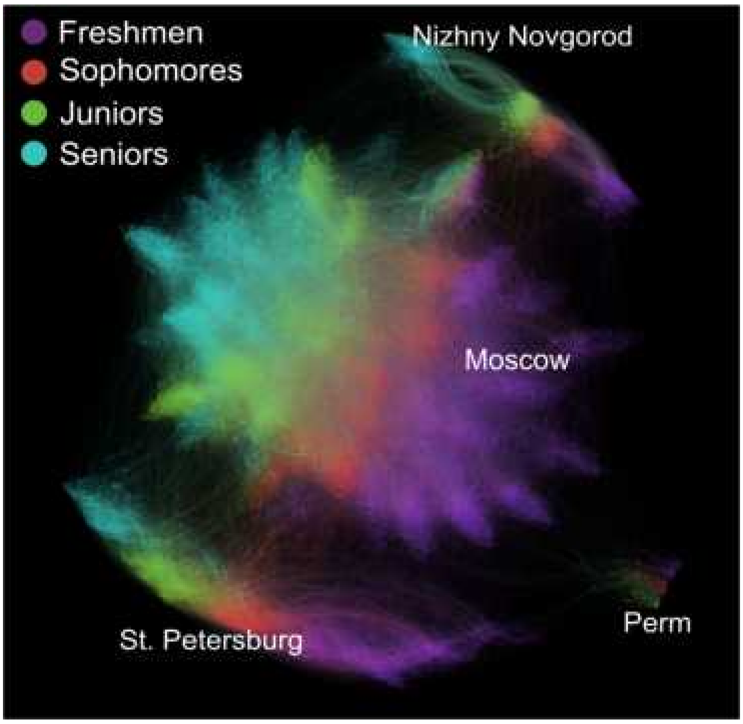
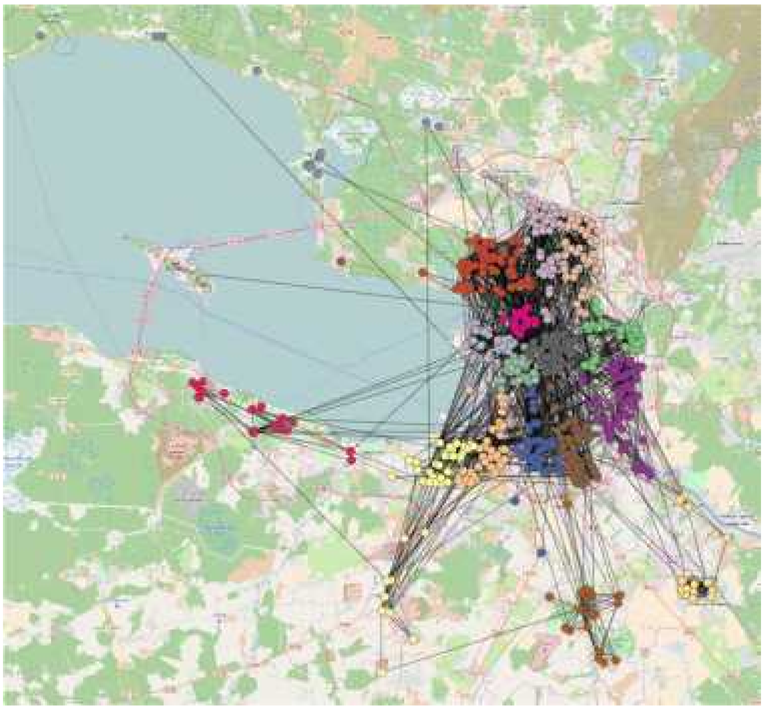
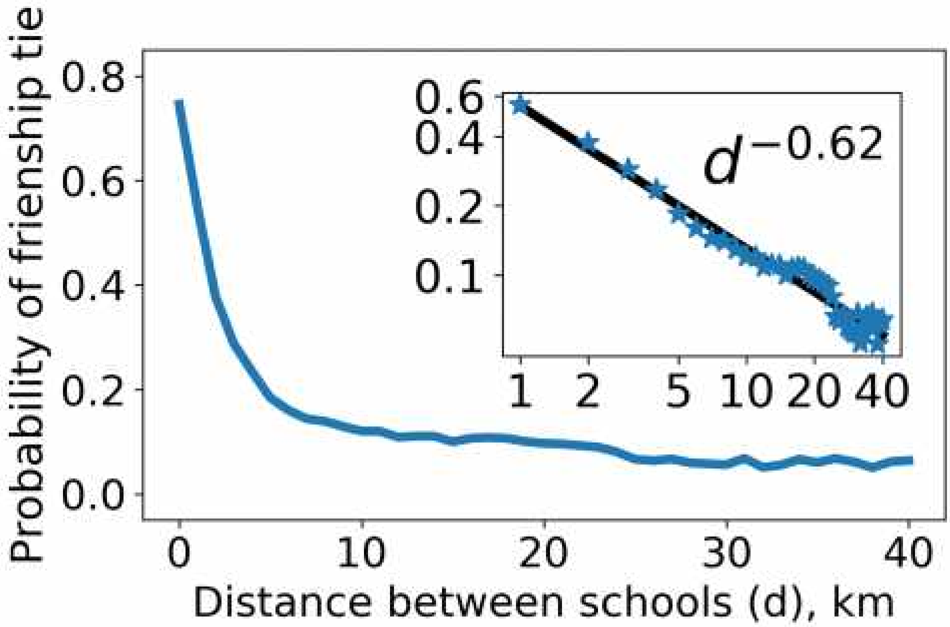


Figure 3. Structure of social ties between schools of Saint Petersburg. Different colors correspond to different administrative districts.



We also study the structure of friendship ties on a city scale (Figure 3). We show that the probability of a friendship tie between two schools decreases with geographical distance following the power law (Figure 4).

Figure 4. The relationship between the probability of a friendship tie between schools and geographical distance between them.



Differentiation of social ties in the digital space

We show that online social ties of students from an educational organization are differentiated by academic performance, namely that students with similar academic performance are interacting more frequently online. We also find that the level of this differentiation increases with time (Figure 5). We show that this increase cannot be explained by changes in academic performance but rather is explained by rewiring of social ties. Less-similar students break social ties with higher probability and more-similar students create new ties with higher probability. These results are presented in [44].

We also study the social ties of students on a city scale. We show that the probability of a friendship tie on VK between students from different schools is higher for the schools with similar academic performance. These results hold true regardless of the geographical distance between schools (Figure 6). Hence, schools are segregated in the digital space despite the absence of geographical segregation.

Figure 5. Increase in correlation (Homophily Index) between GPA of students and the average GPA of their friends for a school (a) and a university (b).

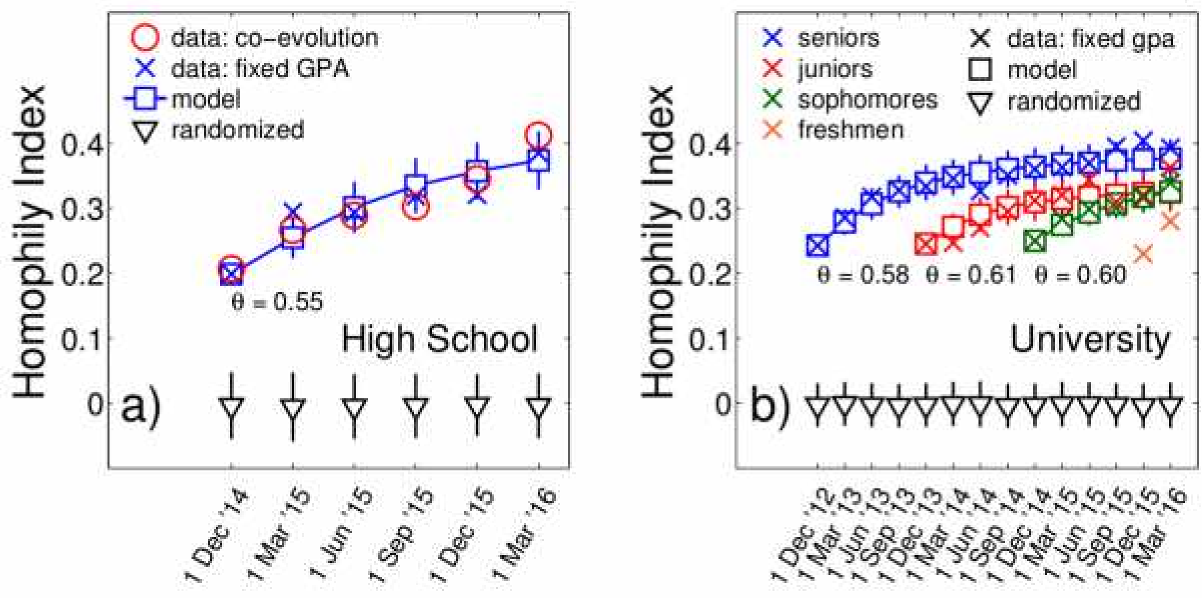
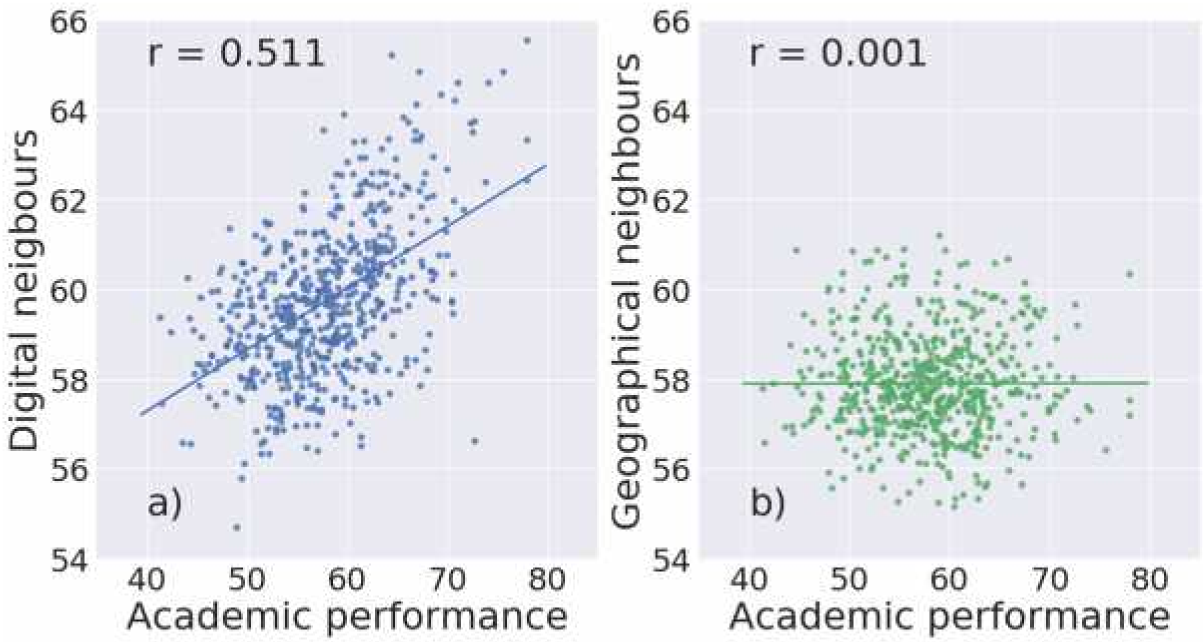


Figure 6. Correlation between the average USE scores of schools and their closest neighbors in the digital (a) and physical (b) space.



Differentiation of interests in the digital space

We find that students' interests are correlated with their gender (for instance, boys prefer public pages related to football and computer games), age (for instance, older students are interested in graduation examinations) and also with their academic performance [2]. Low-performing students are subscribed to such pages as "Love Horoscope" and "Unorthodox Horoscope" while high-performing students prefer such pages as "Interesting facts" and "The best poems of great poets" (Figure 7).

We also show that online interests could explain as much as 25% of the variation in academic performance of students (Figure 8). This is comparable to the percentage of variation that can be explained by the socioeconomic status of students. The gap in educational outcomes of subscribers to different groups (for instance "World Art and Culture" and "Love Horoscope") could be equivalent to two years of formal schooling (Table 5).

Figure 7. Students' interests map.

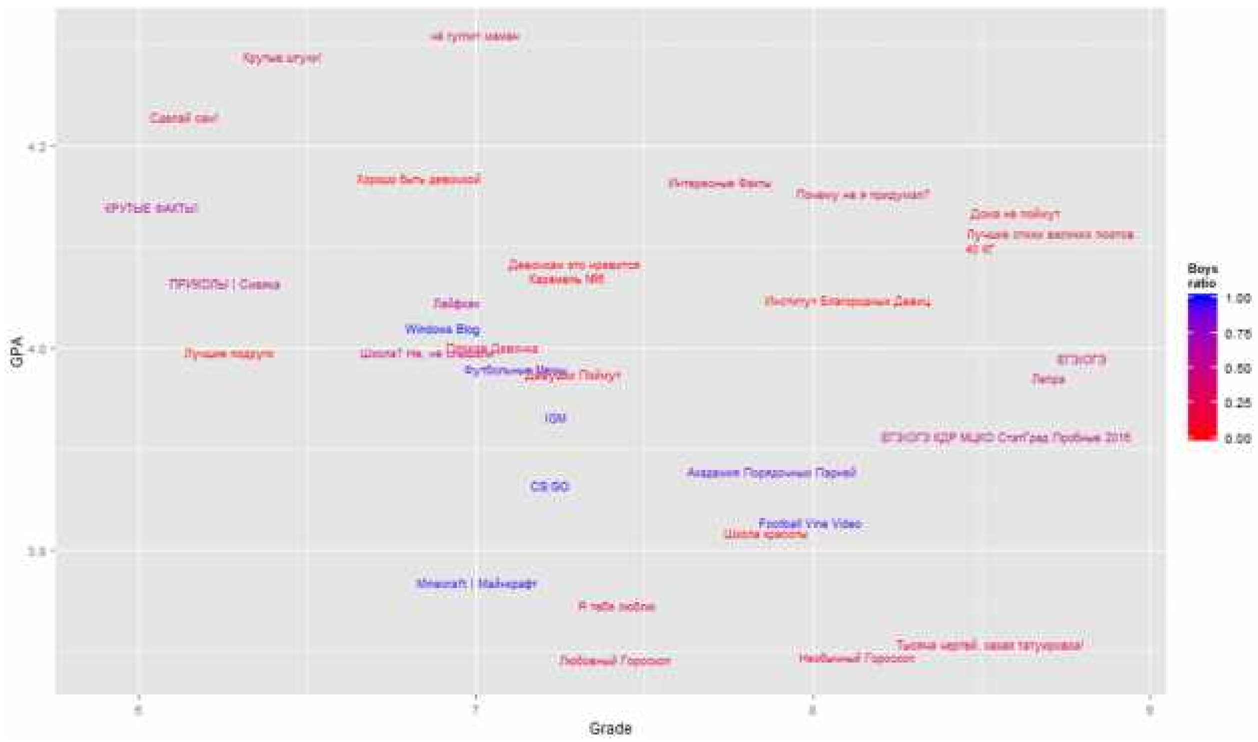


Figure 8. Pearson correlation coefficient between predicted and real PISA scores as a function of the number of components used in the linear regression.

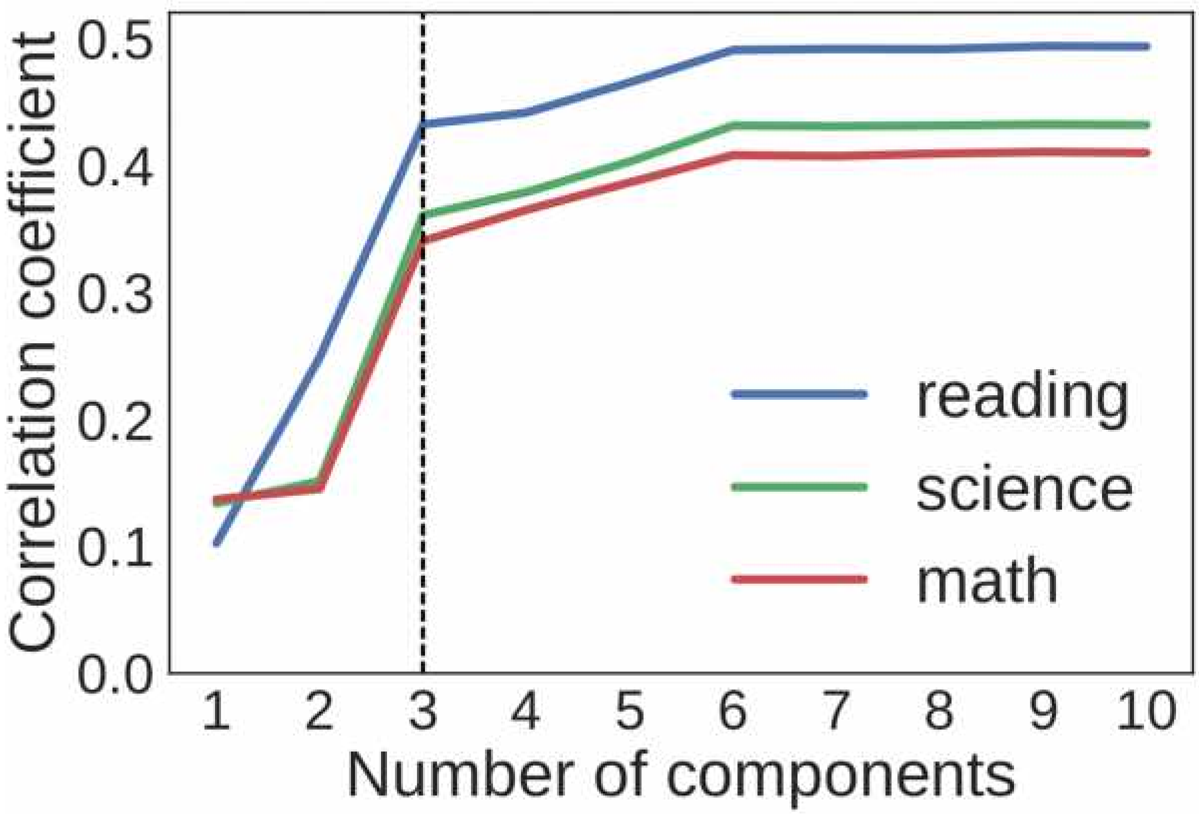


Table 5. Names of public pages that contribute most to the academic component of users' interests. Names are translated from Russian. Mean values of subscribers' scores with standard errors (in parentheses) are provided for each of three PISA subjects.

	Mathematics	Reading	Science
Positive contribution			
WAC (World Arts and Culture)	538 (4.6)	530 (4.5)	532 (4.3)
Science	521 (4.2)	502 (4.1)	516 (3.8)
Best poems of great poets	509 (4.0)	507 (4.0)	508 (3.9)
Science and Technology	507 (4.1)	479 (4.3)	504 (4.0)
Five Best Movies	505 (3.9)	492 (3.9)	503 (3.7)
Negative contribution			
F*CK	473 (3.3)	449 (3.4)	472 (3.2)
Killing humor	471 (5.1)	447 (5.1)	471 (4.7)
Cool Gags	467 (4.9)	444 (5.1)	465 (4.9)
Unorthodox Horoscope	462 (5.1)	450 (5.3)	460 (5.0)
Love Horoscope	450 (5.3)	442 (5.8)	453 (5.2)

These results could be summarized as the following:

— the method we propose allows to extract reliable information from VK and to combine it with educational data; resulting data could be used to study social ties of students and their interests;

— the structure of online friendship reproduces the structure of educational organizations; the social proximity in the digital space is closely related to the geographical proximity; the probability of a friendship tie between students from different schools declines with geographical distance following power law;

— social ties of students are differentiated by academic performance in the digital space; students with similar performance create ties with higher probability and students with dissimilar performance break ties with higher probability; the students from similar performing schools are more often connected on a social networking site regardless the geographical distance between schools;

— students' interests are differentiated by academic performance in the digital space; the gap in educational outcomes of subscribers to different public pages could be equivalent to two years of formal schooling.

Conclusion

Ethical considerations

For the purposes of this work, we use only publicly available information from the social networking site. The VK team confirmed to us that this data can be used for research purposes. The matching of VK profiles with information about students was done automatically, after matching the data was anonymized and later used for analysis only in this anonymized form. The procedure was approved by the IRB of Higher Schools of Economics.

It is important to note that new sources of data not only opens up new opportunities for researchers but also raise new ethical questions. For instance, the notion of informed consent requires special attention. By accepting terms of service users of social networking sites agree that information about them could be accessed by third parties and used for a variety of unspecified purposes. However, it is not clear if such consent could be considered as informed. Especially considering the fact that terms of services are rarely read and if read then users may still not fully understand all the consequences of their consent. For example, it was shown that digital traces allows to effectively predict information that users did not disclose and may prefer not to be known by others. Despite the appearance of the first ethical guidelines this field is still largely a grey zone and requires additional attention from the research community.

Scientific novelty and significance of the results

We have conducted the first large-scale study that combines detailed information about the behavior of students on VK with educational data. We introduced methods that could increase the reliability of VK data and provided estimates for sample biases. We introduced a novel approach to studying the evolution of students’ social ties that did not require to conduct expensive longitudinal studies. We showed how publicly available information from SNSs could be used to infer information about students’ interests and that this information could have a large predictive power in respect to various students’ characteristics including their age, gender, and academic performance. These results are important for further educational research because our methods could be adopted by other researchers and have already been used in various works [46–48].

For the first time, the structure of students’ friendship ties have been studied on a city scale and the relationship between inter-school friendship and geographical distance has been revealed. We also studied the evolution of social ties of students within educational institutions and showed that the differentiation of these ties by academic performance increased with time. We explained the mechanisms behind this phenomenon with a simple model. We showed that there was a differentiation of students’ online interests by academic performance and, for the first time, provided an estimate of the gap in educational outcomes between subscribers to various public pages.

We showed that social ties of students and their online interests had a large predictive power in respect to academic performance. The variables that were constructed by us explained as much variation in educational outcomes (for individual students and for whole schools as well) as the socioeconomic status measured by traditional indexes such as the index of economic social and cultural status (ESCS) used by PISA. This allows one to use constructed variables for operationalization of social and cultural capital of students (at least for its digital dimension). Traditional indexes include such variables as parents’ level of education and number of books at home. Such variables have a low resolution (e.g. parents’ level of education) or disputable face validity in the modern world (e.g. a number of books at home). Another advantage of our approach is that it shifts focus from family characteristics to characteristics of students themselves.

It is important to note that our results do not give an answer to the question of whether the observed differentiation leads to amplification or reproduction of inequality. We also do not study any effects that families might have on the observed differentiation. Further research might seek answers to these important questions.

Our results may have implications for the practice of education due to the important role that SNSs play in the life of modern students. The main component of the SNS is a newsfeed that is constructed from information posted by friends and from subscriptions to various public pages. We showed that both of these sources are differentiated by academic performance. It means that everyday digital flow of information is fundamentally different for students with varying academic performance. This fact should be taken into account by teachers in their everyday practice. The digital environment of students is out of the control of traditional pedagogical tools; however, teachers could influence it. For instance, one of the main features of modern SNSs is recommendation algorithms that suggest new information based on the history of users’ behavior. Our research showed that this might lead to information bubbles of horoscopes and cool gags for low-performing students. Teachers, however, might ask their students to find some educational information on SNSs or elsewhere on the Internet. This student activity will lead to digital traces that later could be used by recommendation algorithms to show students new information that might be more enriching than information that would be shown without intervention. Development of similar pedagogical interventions along with studying of their potential impact may be a logical continuation of our work.

REFERENCES

1. Watts D.J. A twenty-first century science // Nature. 2007. Vol. 445. №. 7127. P. 489–489.

2. Lazer D., Pentland A., Adamic L., Aral S., Barabási A., Brewer D., Christakis N., Contractor N., Fowler J., Gutmann M., Jebara T., King G., Macy M., Roy D., Alstyne M. V. Computational social science // Science. 2009. Vol. 323. №. 5915. P. 721–723.

3. McFarland D. A., Lewis K., Goldberg A. Sociology in the era of big data: The ascent of forensic social science // The American Sociologist. 2016. Vol. 47. №. 1. P. 12–35.

4. Kandel D. B. Homophily, selection, and socialization in adolescent friendships // American journal of Sociology. 1978. Vol. 84. №. 2. P. 427-436.

5. Tuma N. B., Hallinan M. T. The effects of sex, race, and achievement on schoolchildren's friendships // Social Forces. 1979. Vol. 57. №. 4. P. 1265-1285.

6. Shrum W., Cheek Jr. N. H., MacD S. Friendship in school: Gender and racial homophily // Sociology of Education. 1988. P. 227-239.

7. Flashman J. Academic achievement and its impact on friend dynamics // Sociology of education. 2012. Vol. 85. №. 1. P. 61-80.

8. Currarini S., Jackson M. O., Pin P. Identifying the roles of race-based choice and chance in high school friendship network formation // Proceedings of the National Academy of Sciences. 2010. Vol. 107. №. 11. P. 4857-4861.

9. Lomi A., Snijders T. A., Steglich C. E., Torló V. J. Why are some more peer than others? Evidence from a longitudinal study of social networks and individual academic performance // Social Science Research. 2011. Vol. 40. №. 6. P. 1506-1520.

10. Eagle N., Pentland A. S., Lazer D. Inferring friendship network structure by using mobile phone data // Proceedings of the national academy of sciences. 2009. Vol. 106. №. 36. P. 15274-15278.

11. Mastrandrea R., Fournet J., Barrat A. Contact patterns in a high school: a comparison between data collected using wearable sensors, contact diaries and friendship surveys // PloS one, 2015. Vol. 10. №. 9, e0136497. URL: https://doi.org/10.1371/journal.pone.0136497 (дата обращения: 07.06.2018).

12. Stephens-Davidowitz S. Everybody lies: Big data, new data, and what the internet can tell us about who we really are. – Dey Street Books, 2017.

13. OECD (2016), PISA 2015 Results (Volume I): Excellence and Equity in Education, PISA, OECD Publishing, Paris, URL: https://doi.org/10.1787/9789264266490-en (дата обращения: 07.06.2018).

14. Bond R. M., Fariss C. J., Jones J. J., Kramer A. D., Marlow C., Settle J. E., Fowler J. H. A 61-million-person experiment in social influence and political mobilization // Nature. 2012. Vol. 489. №. 7415. P. 295–298.

15. Kramer A. D., Guillory J. E., Hancock J. T. Experimental evidence of massive-scale emotional contagion through social networks // Proceedings of the National Academy of Sciences. 2014. Vol. 111. №. 24. P. 8788–8790.

16. Hobbs W. R., Burke M., Christakis N. A., Fowler J. H. Online social integration is associated with reduced mortality risk // Proceedings of the National Academy of Sciences. 2016. Vol. 113. №. 46. P. 12980-12984.

17. Preoţiuc-Pietro D., Volkova S., Lampos V., Bachrach Y., Aletras N. Studying user income through language, behaviour and affect in social media // PloS one. 2015. Vol. 10. № 9, e0138717. URL: https://doi.org/10.1371/journal.pone.0138717 (дата обращения: 07.06.2018).

18. An J., Weber I. # greysanatomy vs.# yankees: Demographics and hashtag use on twitter. Proceedings of the Tenth International Conference on Web and Social Media. - Cologne, Germany: AAAI Press, 2016.

19. Rao D., Paul M., Fink C., Yarowsky D., Oates T., Coppersmith G. Hierarchical bayesian models for latent attribute detection in social networks. In Proceedings of the International Conference on Weblogs and Social Media. 2011.

20. Gebru T., Krause J., Wang Y., Chen D., Deng J., Aiden E. L., Fei-Fei L. Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States // Proceedings of the National Academy of Sciences. 2017. Vol. 114. №. 50. P. 13108–13113.

21. Watts D. J. Common sense and sociological explanations // American Journal of Sociology. 2014. Vol. 120. №. 2. P. 313-351.

22. Hofman J. M., Sharma A., Watts D. J. Prediction and explanation in social systems // Science. 2017. Vol. 355. №. 6324. P. 486–488.

23. Von Hayek F. A. The pretence of knowledge // The American Economic Review. 1989. Vol. 79. №. 6. P. 3–7.

24. Kosinski M., Stillwell D., Graepel T. Private traits and attributes are predictable from digital records of human behavior // Proceedings of the National Academy of Sciences. 2013. Vol. 110. №. 15. P. 5802–5805.

25. Krasilnikov A., Semenova M. Do Social Networks Help to Improve Student Academic Performance? The Case of Vk.com and Russian Students // Economics Bulletin. 2014. Vol. 34. №. 2. P. 718–733.

26. Докука С. В., Валеева Д. Р., Юдкевич М. М. Коэволюция социальных сетей и академических достижений студентов // Вопросы образования. 2015. №3. С. 44–65.

27. Alexandrov D., Karepin V., Musabirov I. Educational migration from Russia to China: social network data. In Proceedings of the 8th ACM Conference on Web Science. 2016. URL: doi:10.1145/2908131.2908192 (дата обращения: 07.06.2018).

28. Королева Д. О. Всегда онлайн: использование мобильных технологий и социальных сетей современными подростками дома и в школе // Вопросы образования. 2016. № 1. С. 205-224.

29. Королева Д. О. Исследование повседневности современных подростков: присутствие в социальных сетях как неотъемлемая составляющая общения // Современная зарубежная психология. 2016. Т. 5. № 2. С. 55-61.

30. Google (2017). Новое поколение интернет-пользователей: исследование привычек и поведения российской молодежи онлайн. URL: https://www.thinkwithgoogle.com/intl/ru-ru/research-study/novoe-pokolenie-internet-polzovatelei-issledovanie-privychek-i-povedeniia-rossiiskoi-molodezhi-onlain/ (дата обращения: 07.06.2018).

31. Shareaholic. Social Media Traffic Report. 2014. URL: https://blog.shareaholic.com/social-media-traffic-trends-01-2015/ (дата обращения: 07.06.2018).

32. Касамара В. А. Ценностные ориентация российского студенчества. 2017. URL: https://ioe.hse.ru/seminar1617 (дата обращения: 07.06.2018).

33. Dimaggio P., Hargittai E., Celeste C., Shafer S. Digital inequality: From unequal access to differentiated use. Social Inequality. – New York: Russell Sage Foundation, 2004. P. 355-400.

34. DiMaggio P., Hargittai E. "From the 'Digital Divide' to 'Digital Inequality': Studying Internet Use as Penetration Increases" Working Papers 47. Princeton University, Woodrow Wilson School of Public and International Affairs, Center for Arts and Cultural Policy Studies. 2001. 4(1), 4-2.

35. DiMaggio P., Hargittai E., Neuman W. R., Robinson J. P. Social implications of the Internet // Annual review of sociology. 2001. Vol. 27. №. 1. P. 307-336.

36. Poole, G. A. A new gulf in American education, the digital divide. // New York Times. 1996. 29.01.1996. D3.

37. Hargittai E., Hinnant A. Digital inequality differences in young adults' use of the Internet // Communication Research. 2008. Vol. 35. №. 5. P. 602-621.

38. Волченко О. В. Динамика цифрового неравенства в России // Мониторинг общественного мнения: экономические и социальные перемены. 2016. № 5. С. 163-182

39. Hansen J. D., Reich J. Democratizing education? Examining access and usage patterns in massive open online courses // Science. 2015. Vol. 350. №. 6265. P. 1245-1248.

40. Kizilcec R. F., Davis G. M., Cohen G. L. Towards Equal Opportunities in MOOCs: Affirmation Reduces Gender & Social-Class Achievement Gaps in China. In Proceedings of the Fourth ACM Conference on Learning @ Scale. Cambridge, MA: 2017. P. 121-130.

41. Redmiles E. M. Net Benefits: Digital Inequities in Social Capital, Privacy Preservation, and Digital Parenting Practices of U.S. Social Media Users. In Proceedings of the Twelfth International AAAI Conference on Web and Social Media. 2018. P. 270-279.

42. Бессуднов А.Р., Куракин Д.Ю., Малик В.М., Янбарисова Д.М. Бюллетень Российского лонгитюдного панельного исследования образовательных и трудовых траекторий. Выпуск 1. Национальная панель: первая волна (2011-12) // Научные доклады Института образования WP 02/IE/2014.

43. Смирнов И. Б., Сивак Е. В., Козьмина Я. Я. В поисках утраченных профилей: достоверность данных ВКонтакте и их значение для исследований образования // Вопросы образования. 2016. №4. С. 106-122.

44. Smirnov I., Thurner S. Formation of homophily in academic performance: Students change their friends rather than performance. // PloS one. Vol. 12. №. 8, e0183473.

45. Поливанова К. Н., Смирнов И. Б. Что в профиле тебе моем: данные «ВКонтакте» как инструмент изучения интересов современных подростков // Вопросы образования. 2017. №2. С. 134-151.

46. Smirnov I The Digital Flynn Effect: Complexity of Posts on Social Media Increases over Time. // SocInfo 2017. Lecture Notes in Computer Science. 2017. Vol. 10540. P. 24-30.

47. Elizaveta Sivak, Ivan Smirnov. Gender Bias in Sharenting: Both Men and Women Mention Sons More Often Than Daughters on Social Media. 2018. arXiv:1803.10329.

48. Кашарин, М. Ю. Связь между участием студентов в студенческих организациях и их социальным капиталом. 2017. Магистерская диссертация. URL: https://www.hse.ru/edu/vkr/206749166 (дата обращения: 07.06.2018).