THE CORRELATION BETWEEN THE SOCIAL NETWORK ACTIVITY AND ACADEMIC PERFORMANCE OF TECHNICAL UNIVERSITY STUDENTS: A CASE STUDY OF *VKontakte*

V. V. Sych

Bauman Moscow State Technical University, Moscow, Russia
Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia

✉ sycheslavovich@gmail.com

Abstract. The academic performance of students at Bauman Moscow State Technical University and its correlation to their activity in the *VKontakte* social network are considered. Machine learning methods are used to identify distinct performance paths reflecting the dynamics of educational achievement. The subscription lists of students are analyzed to identify marker communities characterizing the predominance of specific performance categories. Graph-theoretic clustering is applied to reveal structural groups of student interests. For each path, the stochastic vectors of interest shares across the community clusters are constructed and then used to identify the clusters having a statistically significant relation to particular performance paths. The results indicate a correlation between the digital behavior and academic outcomes of students, which contributes to the development of performance prediction models considering student interests in a social network.

Keywords: academic performance, *VKontakte*, machine learning, graph theory, performance prediction.

INTRODUCTION

Within the digitalization of the educational process, the data generated by students on social networks are becoming an important source of information for analyzing their academic and social activity. The possibility of differentiating students by their academic performance level based on their activity on *VKontakte* (VK), a Russian online social media and social networking service, was earlier demonstrated in several research works [1]. The dependence of career achievements on participation in professional social networks was also investigated [2], and academic success was considered to spread through friendly and educational links [3]. An inverted U-shaped relationship between the intensity of social network use and academic performance was identified [4]. As also established, mental health can act as a mediator in this relationship [5].

The vast majority of students have VK pages with a significant amount of information reflecting their interests and social links. Although social networks are rarely used to search for serious information and are more often launched “for leisure,” the long-term structure of subscriptions reflects the stable preferences and personal orientations of users. Therefore, their digital footprint can be associated with educational indicators and used as a reliable predictor of academic performance. This research is intended to identify the correlation between the academic performance and behavior of students in the digital space using machine learning methods and graph theory.

The topicality of this research is due to the need to develop effective tools for academic achievement monitoring and early identification of students at risk. Traditional performance assessment methods based on limited samples and surveys often neglect the multifaceted aspects of student behavior outside educational institutions. Analyzing digital footprints on social networks provides a more complete picture of students with their interests, social links, and even the interests of their friends, supplementing the existing assessment methods.

The aim of this research is to analyze the correlation between the academic performance paths and digital interests of students. Its uniqueness lies in the use of an extensive database: among 76 000 students of the Bauman Moscow State Technical University in 2014–2023, real VK pages were successfully matched for 48 000 students. The data include both academic performance indicators (end-of-semester examination grades) and information about subscriptions and friendships; therefore, a comprehensive digital portrait of each student can be created. Figure 1 shows an example of a friendship graph based on VK data.

The methodological foundation of this research includes clustering, standardization, and principal component analysis (PCA) algorithms to identify typical academic performance paths, as well as graph theory methods to analyze community structures. The research is based on VK data and university electronic records, anonymized after matching.

1. DATA DESCRIPTION

The analysis involves two main sources of information. The first source is an academic database obtained from the Electronic University system, covering end-of-semester examination grades for 2014–2023. For each student, grades for each semester are recorded, and the average grade for each of the four academic years can be calculated accordingly. The second source is VK data, collected using the VK API. For each user, the lists of community subscriptions and friendship information are extracted to analyze the interests and network environment of students. The profiles subscribed to university-related communities and all their friends were downloaded. A similar approach to data mining was used in [6].

The matching of academic data with VK profiles was carried out iteratively. At the first stage, the "core" of students with a high data fit degree was formed, where the comparison was carried out via the strict matching of full names and dates of birth, supplemented by the Levenshtein distance [7] calculated for non-matching name forms.

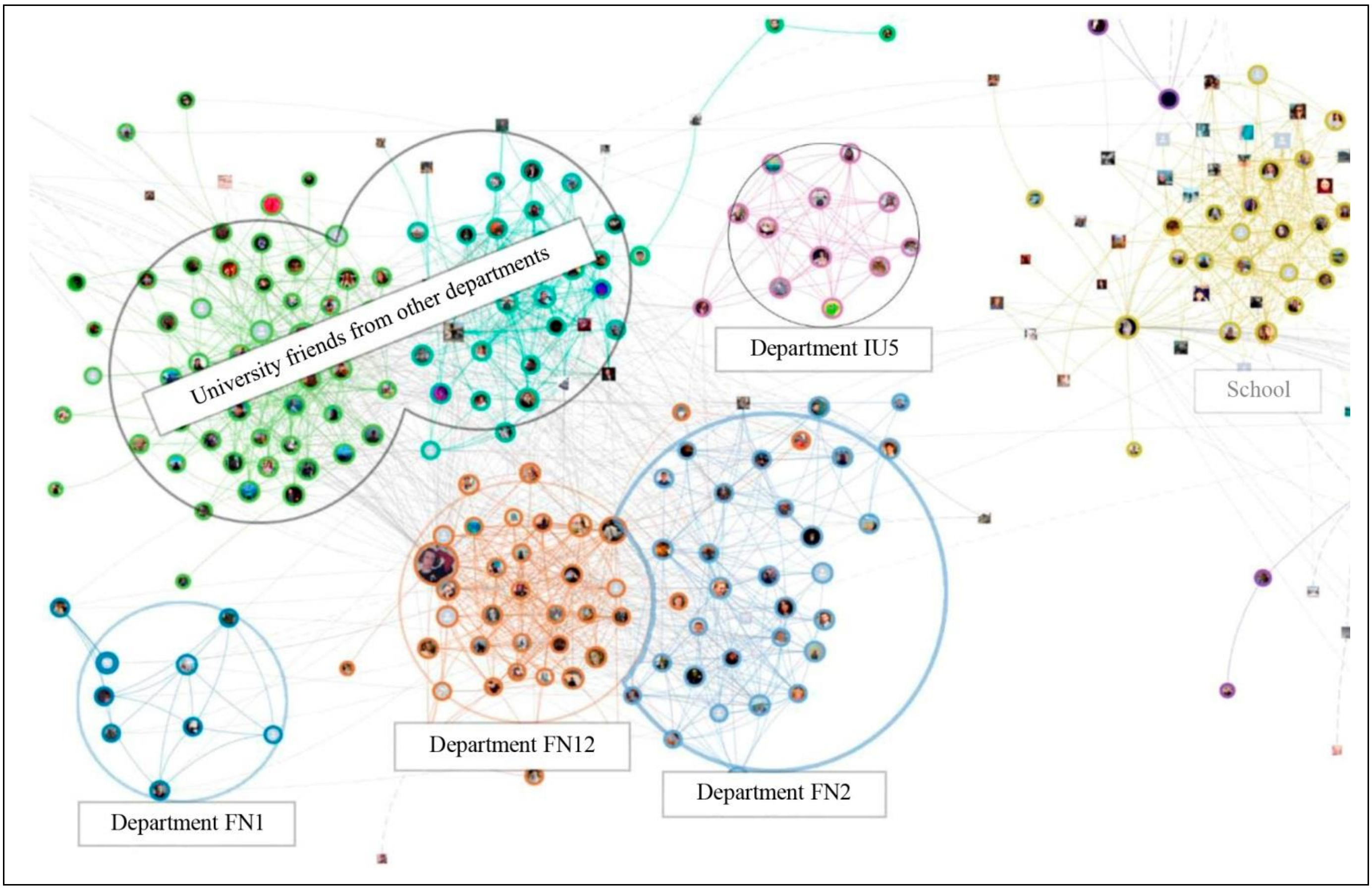

Fig. 1. The author's friendship graph based on VK links as an example (large clusters are labeled).

At the second stage, additional profiles were identified using friendship information: if a candidate matched only the basic parameters (e.g., name and year of birth) but was associated with the identified students from a particular study group, he/she was assigned to the same group with high probability. This process was based on the well-known principle: "tell me who your friends are, and I'll tell you who you are" (friendships as an additional marker of reliability). Following this approach, we managed to include students without enough initial matches for unambiguous identification.

The iterative repetition of the process (a core of 5000 students obtained at the first iteration, about 22 000 at the second, and 46 000 at the third) is based on the principle that the more friend profiles have already been confirmed for a student, the higher the probability of correctly matching his/her profile with academic data will be. For example, in cases where many users have the same name and the same birth date, an additional feature in the form of friendships with students with confirmed profiles significantly increases the reliability of identification. This approach considerably enlarges the sample and ensures its high representativeness. After the matching procedure, the data were anonymized and used for analysis only in this form. Similar methods for the extended matching of digital footprints with educational data demonstrated high effectiveness when investigating the academic performance of high school and college students [8, 9].

2. ACADEMIC PERFORMANCE ANALYSIS

To analyze academic performance paths, we selected students who had successfully completed four bachelor's or specialist program years without academic debt; the corresponding sample contains 15 495 males and 6285 females. A cluster approach to the segmentation of academic outcomes was earlier applied to assess the relationship between time spent on social networks and the grades of students specialized in biology [10] and analyze the use of social media on mobile devices in an international sample [11], confirming its suitability for identifying latent academic performance paths. For each student i, we calculated the average year grade vector

$$x_i = (x_{i,1},\, x_{i,2},\, x_{i,3},\, x_{i,4}) \in \mathbb{R}^4,$$

where $x_{i,c}$ denotes the average grade for academic year c (c = 1, 2, 3, 4). Next, the matrix $X \in R^{N\times 4}$ was standardized as follows:

$$z_{i,c} = \frac{x_{i,c} - \mu_c}{\sigma_c},$$

where μ_c and σ_c stand for the sample mean and standard deviation for academic year c, respectively.

We employed the k-means algorithm to partition students by academic performance paths. The optimal number k of clusters was selected based on the *silhouette score*, a measure of intra-cluster compactness (tightness) and inter-cluster separation proposed by P. Rousseeuw [12]. The maximum value of this measure (about 0.43) was achieved at $k = 2$, but this partition does not reflect the diversity of academic performance paths. Increasing k from 6 to 16 stabilized the silhouette score in the range of 0.21–0.27, indicating moderate but stable clustering quality.

To confirm the optimal number of clusters, we applied the *Gap Statistic* method proposed by R. Tibshirani et al. [13]. This method allows comparing the clustering quality of the original data with that of randomly generated data with the uniform distribution in the same range. The higher the value Gap(k) of this statistic is, the more distinct structure the data will have. (In other words, the clusters identified will be more significant than random ones.) For $k = 6$ and $k = 8$, the values of Gap(k) were close: 0.994 ± 0.006 and 0.911 ± 0.003, respectively. The slight decrease in the gap statistic value when passing from six to eight clusters is within the standard error, forming the so-called "plateau." This indicates that the detail of the clusters does not significantly deteriorate when increasing their number from six to eight.

Thus, despite the moderate values of the quality metrics, the choice of $k = 8$ is a reasonable trade-off between the statistical stability of the results and the sufficient detail of academic performance paths for the subsequent analysis of their correlation with the digital interests of students.

After the partition, each student was assigned a cluster number, in ascending order of the overall average academic performance (the simple arithmetic mean of all grades for eight semesters, see Table 1). Note the decreasing trend in the proportion of males when passing from clusters with low academic performance to those with high academic performance.

The academic performance paths are shown in Figs. 2 and 3.

The final partition is characterized by the following paths:

– **Path 0.** Stably low academic performance.

– **Path 1.** Low grades in the initial academic years, followed by an improvement.

– **Path 2.** Initially good outcomes, followed by a decline.

– **Path 3.** Low academic performance with noticeable growth.

– **Path 4.** Stably good grades with a slight dip.

– **Path 5.** Moderate grades transitioning to stably good outcomes.

Table 1

Statistics by path clusters

Cluster	The number of students	Average grade	The share of males, %
0	1910	3.48	90.0
1	2510	3.81	85.7
2	2120	3.92	82.8
3	2870	4.19	76.7
4	1950	4.21	71.6
5	2200	4.43	69.7
6	3500	4.57	52.9
7	4720	4.86	56.5

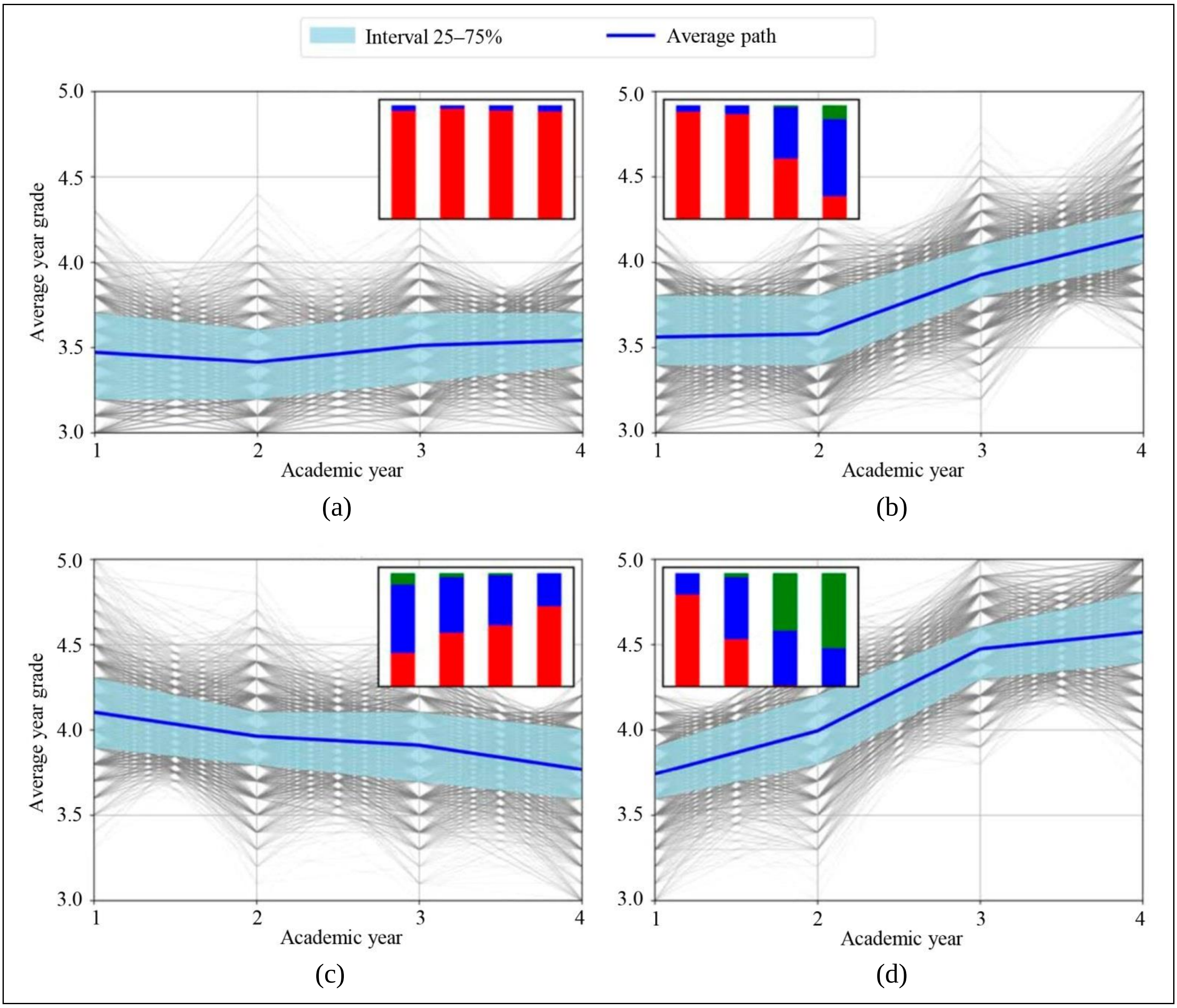

Fig. 2. Path clusters: (a) cluster 0, (b) cluster 1, (c) cluster 2, and (d) cluster 3.

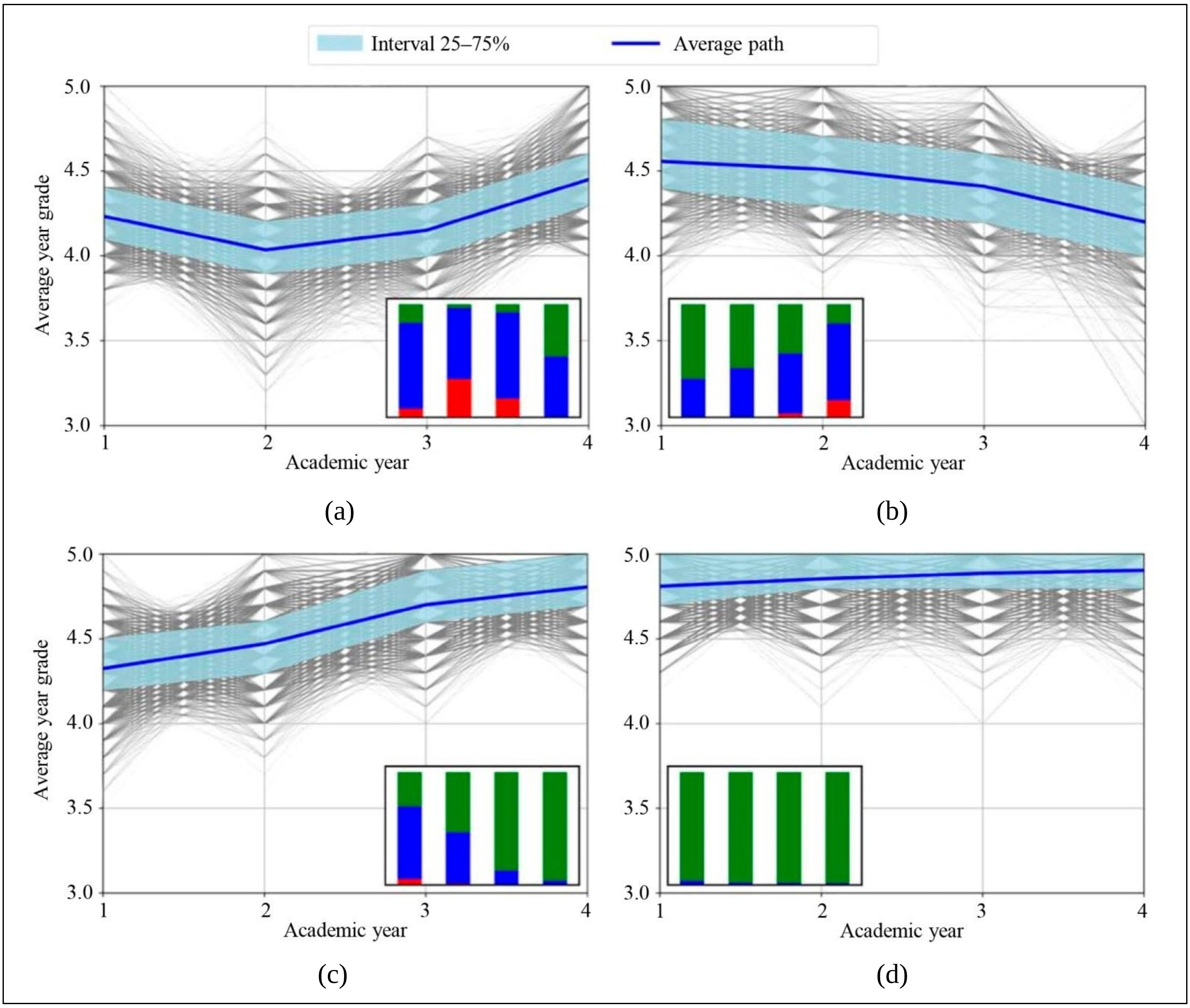

Fig. 3. Path clusters: (a) cluster 4, (b) cluster 5, (c) cluster 6, and (d) cluster 7.

For clarity, each cluster is represented by diagrams showing the proportions of students in different academic performance categories[1]: red, blue, and green bars indicate the shares of mediocre $(\bar{g} < 3.9)$, good $(3.9 \le \bar{g} < 4.5)$, and excellent $(\bar{g} \ge 4.5)$ students, respectively, where $\bar{g}$ is the student's average grade (the arithmetic mean of all his/her end-of-semester grades). These diagrams allow assessing the distribution of academic performance levels of different students across academic years in each path.

For additional visualization of the clustering results, we employed the *principal component analysis* (PCA) method. In Fig. 4, the points corresponding to students are displayed on the plane, with the horizontal axis reflecting the overall academic performance and the vertical one the dynamics of changes. The cross indicates the center of mass of the set.

The repeated runs of the *k*-means algorithm confirmed the stability of the resulting partition, and the visualization results demonstrate the separation of clusters, despite the partial overlap of their boundaries in the 2D projection. This structure provides a reliable foundation for further analysis of the correlation between the digital interests and academic performance of students since the average values of the paths well describe the main differences between the path clusters.

[1] In Russian higher education, the five-grade system is used: 1—"fail," 2—"unsatisfactory," 3—"satisfactory," 4—"good," and 5—"excellent." In the Western grade system, they are equivalent to F, D, C, B, and A, respectively.

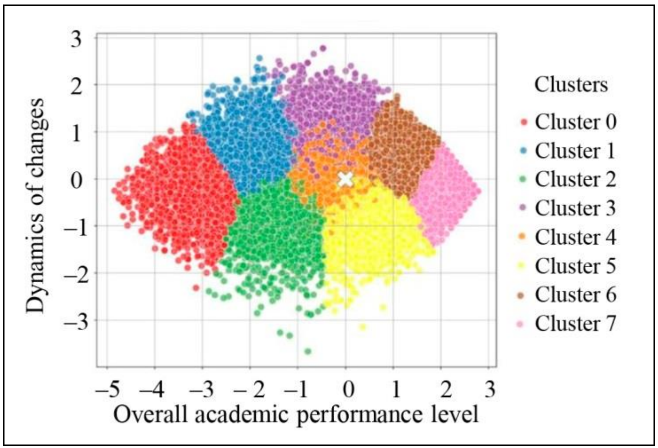

Fig. 4. Student clusters based on academic performance: PCA results.

3. THE GRAPH MODEL OF VK COMMUNITIES AND ITS ROLE IN ACADEMIC PERFORMANCE PATH ANALYSIS

3.1. Initial Data

As the initial data, we took all VK public pages (also called communities or groups) with at least 50 students subscribed; the resulting corpus contains approximately 4500 communities. The approach of using community subscriptions as academic success features has recently proven itself in the identification of the best students by NLP algorithms [14]. Before proceeding to theoretical graph analysis, we checked each community for correlation with academic performance paths using single-factor *analysis of variance (*ANOVA): for 760 communities, the differences between the paths were significant at $p < 0.05$; for 350, at $p < 0.01$.

Let $G = (V, E, W)$ be an undirected graph where V denotes the set of communities and an edge $(c_1, c_2) \in E$ exists if there is at least one common subscriber. The strength of a link between communities was estimated using the *weighted Jaccard coefficient*, first proposed for assessing the similarity of floral compositions [15]:

$$w(c_1, c_2) = \ln\left(1 + \frac{\left|S_{c_1} \cap S_{c_2}\right|}{\left|S_{c_1} \cup S_{c_2}\right|}\right),$$

where S_{c_i} is the set of students subscribed to community c_i. Taking the logarithm allows smoothing out extreme values of the coefficient when the communities overlap greatly or, conversely, are almost disjoint.

To detect communities, we applied the Louvain algorithm [16], which maximizes graph modularity. A single run on the entire graph with *resolution* = 1 produces several abnormally large clusters and many small clusters, so we employed a step-by-step approach. First, the entire graph was clustered at a low resolution value; then the overly large groups were recursively re-partitioned using the same method with *resolution* = 0.7. Additionally, 30% of the weakest links were eliminated, as their total number was very large. The iterative merging of the results provided a final partition into 21 clusters of comparable size.

3.2. The Visualization and Interpretation of Clusters

The final structure of the community graph is shown in Fig. 5: the complete graph with 21 clusters (on the left) and cluster 10 in detail (on the right),

Fig. 5. The visualization of the VK community graph.

Table 2

The characteristics of VK community clusters based on clustering results

Cluster	The description of communities	3/4/5
0	Large communities with universal entertainment content, covering a wide audience	**41**/32/27
1	Communities focused on humorous content, including memes related to e-sports and video games (CS2, PUBG)	**42**/33/25
2	Highly popular communities with memes targeted at a mass audience	**42**/31/27
3	Communities devoted to sports broadcasts and e-sports (soccer, UFC, Dota2)	**46**/32/22
4	Entertainment communities with simple and widespread memes	**42**/33/25
5	Communities with original, unique, or niche memes	**46**/34/20
6	Entertainment communities with repetitive trite memes	**46**/33/21
7	Communities spreading post-ironic memes targeted at a specific audience	**40**/35/25
8	Humorous communities focused on distributing video content	**46**/34/20
9	Communities representing post-irony and absurd humor	**42**/34/23
10	Communities related to educational and scientific topics: programming, internships, research	27/34/**39**
11	Student communities, including faculty public pages and groups like “Eavesdrop”	32/**35**/33
12	Communities focused on student life, including events and activities	27/35/**37**
13	Communities with friendly memes, often focused on pets (especially cats)	32/34/34
14	Communities devoted to aesthetics, fitness, nutrition, and self-care, primarily targeted at a female audience	26/33/**41**
15	Communities about fashion, the beauty industry, online stores, and female memes	25/34/**41**
16	Communities related to self-development, literature, foreign language learning, and aesthetic concepts	30/33/**37**
17	Foreign communities containing female memes and aesthetic content	29/34/**38**
18	Communities devoted to informal fashion, clothing, and footwear trade	**46**/35/19
19	Music communities focused on rap artists and the music industry	**45**/35/20
20	Gaming communities (Dota2, e-sports), as well as groups related to the trade of used cars	**46**/34/20

which includes mainly educational communities. The node colors correspond to the predominant academic performance level of subscribers, namely, red to mediocre, blue to good, and green to excellent students. This coding allows immediately assessing the homogeneity of the clusters in terms of academic performance.

Note that the path clusters ($n = 8$) reflect the dynamics of student academic performance changes over time, while the categories (mediocre, good, and excellent) are calculated based on the individual average grade for the entire study period and are introduced separately for a clear interpretation of the academic performance level in each community. They serve as an additional “projection” of the data.

The description of all clusters and the distribution of students by their academic performance categories are given in Table 2 (based on the author’s subjective generalizations).

4. THE CORRELATION BETWEEN THE DIGITAL INTERESTS AND ACADEMIC PERFORMANCE PATHS OF STUDENTS

4.1. The Formalization of Interests and Calculation of the Log-Odds Ratio

For each student i, we constructed the stochastic interest vector

$$p_i = \left(p_{i,0},\ p_{i,1}, \ldots,\ p_{i,20} \right)$$

with equal weights for all clusters (i.e., $\sum_k p_{i,k} = 1$), where $p_{i,k}$ is the ratio of subscriptions of student i in cluster k to all his/her subscriptions in total. This vector reflects the degree of concentration of the student's interests on particular community clusters.

To identify distinctions in the digital interests of students from different academic performance paths, we used a sample of students with known subscription data: after filtering, 4500 males and 1900 females remained. The *log-odds ratio* [17] was selected for the assessment, measuring the deviation of the share of subscriptions to cluster *k* in academic performance path *t* from the global average:

$$\log\_\text{odds}_{k,t} = \ln\left(\frac{p_{k,t}}{1-p_{k,t}}\right) - \ln\left(\frac{p_{\text{k,g}}}{1-p_{\text{k,g}}}\right).$$

In this formula:

$p_{k,t}$ is the *trimmed mean* share of subscriptions to cluster *k* among students in path *t*. First, for each student *i* from path *t*, the ratio of his/her subscriptions in cluster *k* to all his/her subscriptions is calculated. Then, the trimmed mean is taken from the ratios obtained for all students in path *t*;

$p_{k,\text{g}}$ is the global trimmed mean share of subscriptions to cluster *k* across the entire sample of students. First, for each student, the ratio of his/her subscriptions in cluster *k* to all his/her subscriptions is calculated. Then, the trimmed mean is taken from all these ratios.

To reduce distortion due to atypically low or high individual values, we calculated the average shares for each cluster using trimmed mean: the 10% lowest and highest share values were excluded from the analysis. This approach yielded more reliable and informative estimates robust to outliers.

Positive values of $\log\_\text{odds}$ indicate increased interest in the cluster among students of a particular academic performance path, while negative values indicate a lower share of subscriptions compared to the global average $p_{k,\text{g}}$ (i.e., "avoidance" of this cluster). Here are some illustrative examples of this metric:

- $\log\_\text{odds} = 0.5$ means that the odds of encountering subscriptions to this cluster among students of a given academic performance path are approximately 65% higher than the overall level since $e^{0.5} \approx 1.65$.
- $\log\_\text{odds} = -1$ means that the chances of encountering subscriptions to this cluster among students of a given academic performance path are approximately $e^{-1} \approx 0.37$, i.e., $1/0.37 \approx 2.7$ times lower than the average.

In further analysis, we calculated log_odds separately for males and females. The reasons are as follows:

- **The gender composition of paths is irregular.** In path cluster 0, males account for about 90%; in path cluster 7, only about 57%. During aggregation, this irregularity distorts the global averages $p_{k,\text{g}}$ and, accordingly, the logarithmic coefficients.
- **Subscriptions have gender specifics.** Some communities (clusters 14–17) are targeted at a female audience; their "avoidance" by males is not informative for academic performance conclusions but increases the variance in the overall sample.

Figure 6 shows the heat map of $\log\_\text{odds}$ values for the male sample, revealing marked differences in the interests of such students depending on their academic performance path. The percentage of subscriptions to a given cluster among students of a given academic performance path is indicated below the corresponding log_odds value.

Here are some observations:

- Paths 0–2 (low and declining academic performance). These paths demonstrate a strong bias towards entertainment clusters 0–6 (log_odds ≈ 0.18–0.51) and a moderate bias towards street fashion and rap communities. Also, there is an avoidance of educational cluster 10 and university communities 11–12 (log_odds up to −0.41).
- Path 7 (stably excellent students). The picture is mirrored: negative coefficients for entertainment content and a noticeable jump in interest in clusters 10–12 (education and university communities).

Figure 7 shows a similar heat map for the female sample.

Similar conclusions can be drawn for female students from low academic performance paths: high interest in entertainment clusters and avoidance of educational ones. However, for the best paths (6 and 7), a more uniform distribution of interests is observed. Despite the general obviousness of the conclusions (entertainment prevails among weak students whereas education among strong ones), there are individual preferences within each path. These distinctions in digital interests can be used to build a predictive academic performance model.

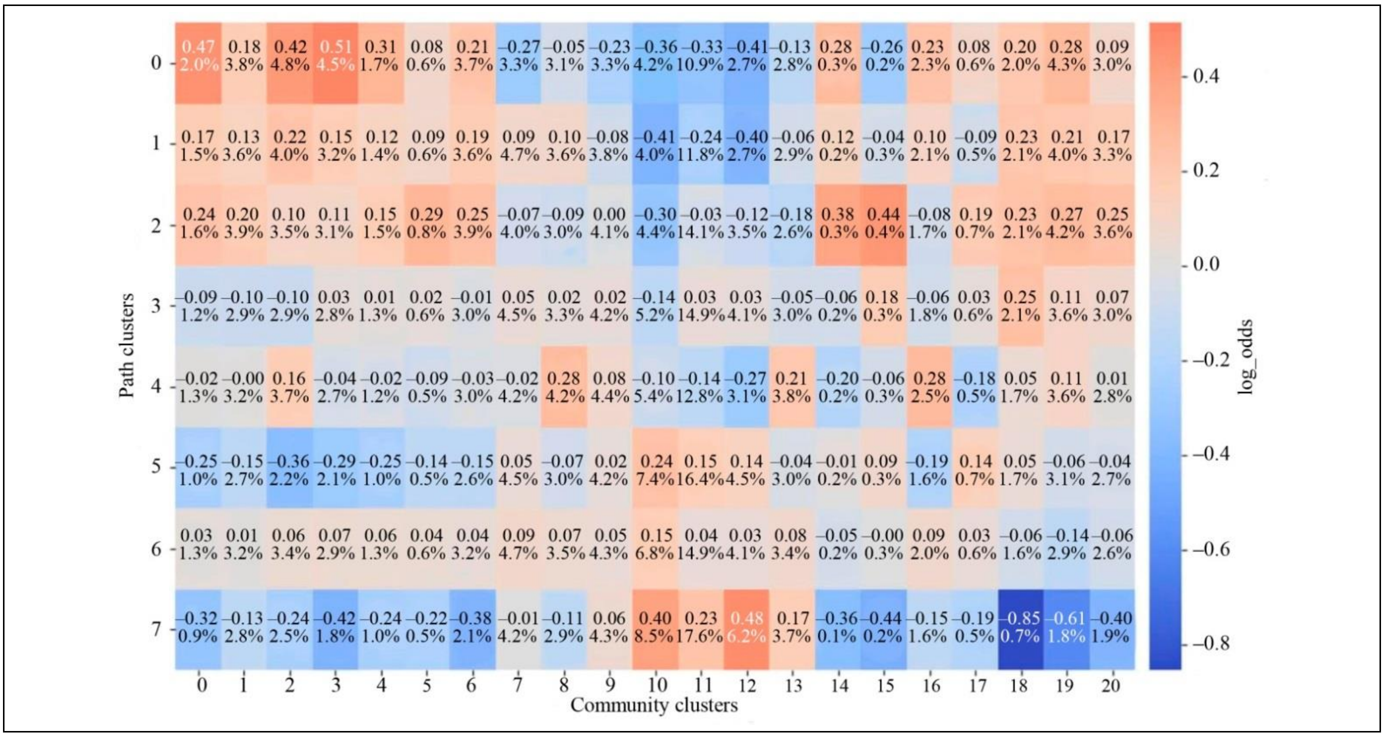

Fig. 6. The heat map of log_odds values for the male sample.

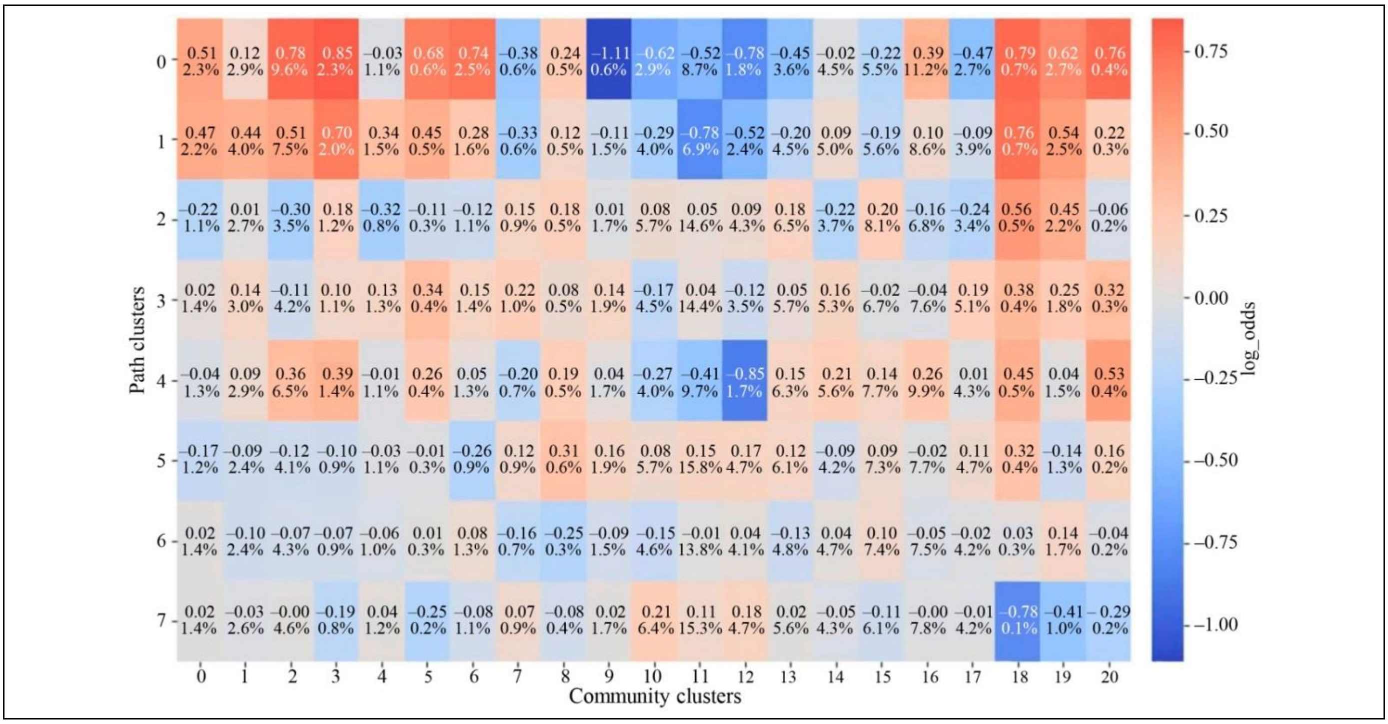

Fig. 7. The heat map of log_odds values for the female sample.

4.2. Testing the Hypothesis on the Independence of Academic Performance Paths from Digital Interests

To formally test the hypothesis on the statistical correlation between the academic performance and digital interests of students, we carried out a contingency table analysis using Pearson's χ^2 test, separately for the male and female samples (Table 3). The global χ^2 test confirmed the existence of a statistically significant correlation between belonging to a particular academic performance path and the subscription profile, among both male and female students.

The distinctions are illustrated in detail in Fig. 8. More specifically, Fig. 8a and 8b present the partial χ^2 test statistic for each of the 21 interest clusters; the χ^2-significant distinctions (α = 0.05) are set off in red whereas the insignificant ones in gray.

Clearly, correlations with academic performance paths became significant in 10 clusters for males and 15 for females. Figures 8c and 8d show estimates of the effect η^2, where the dotted lines indicate the thresholds of 0.01 (small effect), 0.06 (medium), and 0.14 (large) [18]. Most of the significant clusters demonstrate $\eta^2 > 0.3$, and especially among females, about half of the clusters exceed 0.5 (high practical significance of the distinctions revealed). The strongest correlation with academic performance is observed in clusters of educational and student communities (numbers 10–12), as well as entertainment and fashion/music communities (numbers 18–19). This logically fits the subscription patterns observed previously. The data obtained clearly reject the null hypothesis on the independence of academic performance paths from digital interests, confirming that students' digital preferences are representative and substantively related to their academic performance indicators.

Table 3

The global χ^2 test of the independence of academic performance paths from digital interests

Gender	Sample, students	χ^2 test statistic	The number of degrees of freedom	p-value	The critical χ^2 test value (α = 0.05)
Males	~ 4500	557.6	140	1e-51	168.613
Females	~ 1900	755.8		1e-85	

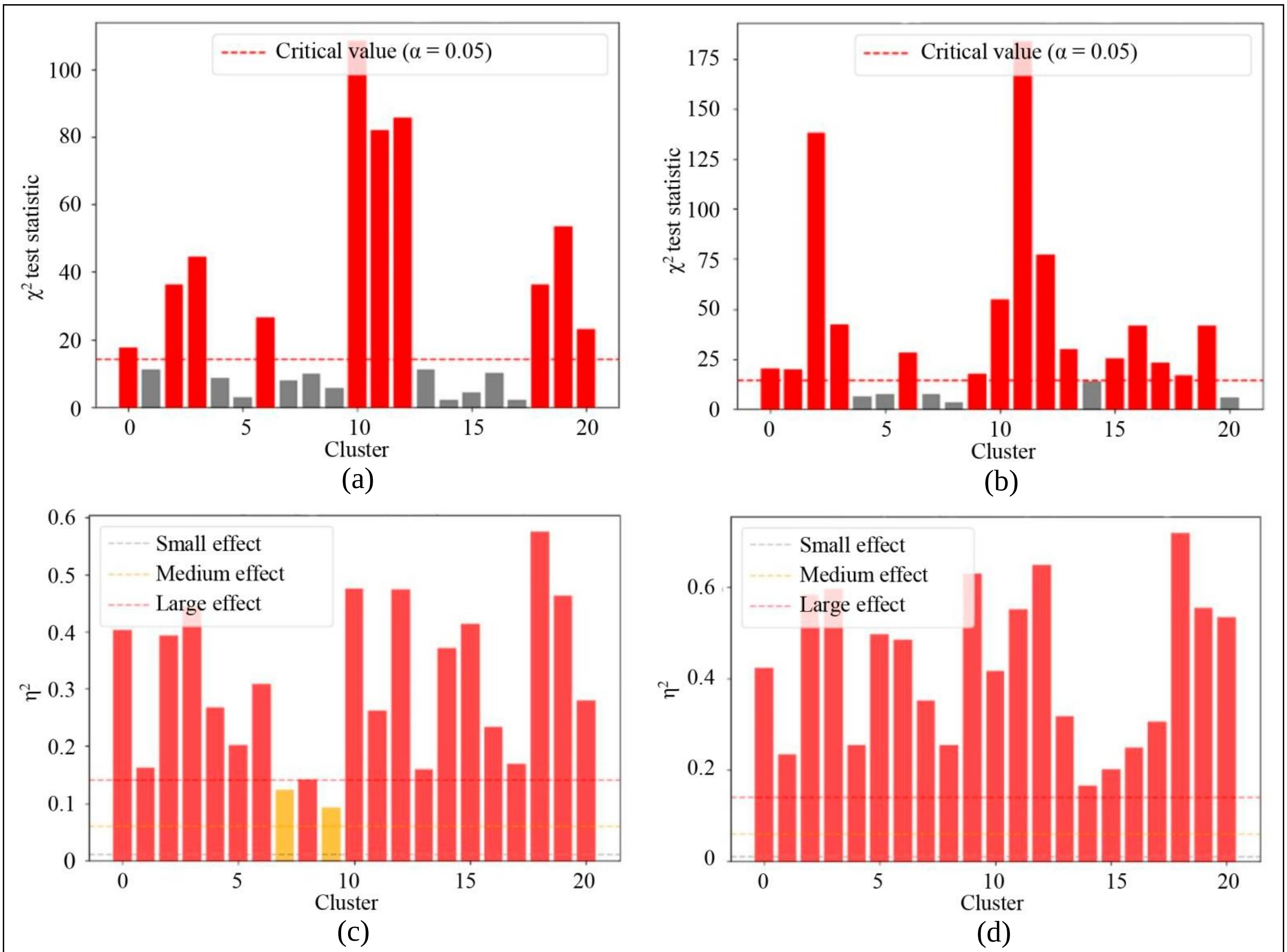

Fig. 8. Partial indicators of the relationship between the clusters of digital interests and academic performance paths: (a), (b) χ^2 statistic (red color – p < 0.05; dotted line – critical χ^2 test value) and (c), (d) estimates of the effect η^2 (dotted lines – thresholds). Bar charts (a) and (c) correspond to males whereas bar charts (b) and (d) to females.

CONCLUSIONS

In this research, we have established a statistical correlation between the topics of subscriptions of students at Bauman Moscow State Technical University on the *VKontakte* social network and their academic performance. According to the results, the digital footprint formed through interests and online activity can be a significant predictor of educational paths and, possibly, other offline indicators. A formal test of the hypothesis using Pearson's χ^2 test has shown high statistical significance $(p < 0.001)$ of the identified correlations for both gender samples, and an analysis of the effect has confirmed the practical significance of the distinctions revealed.

Studies by other researchers also indicate a relationship between social media activity and academic performance, but most of them are limited to small samples and assessments of time spent on social media.

In the future, we plan to expand the analysis by adding the network structure of student friendships in order to build an integrated model for predicting academic performance based on both individual digital interests and their structural position in the student network (centrality measures, community membership, and broker roles). This approach will enable the more accurate diagnosis of academic risks and the formation of personalized recommendations for academic support, from course selection and consulting to inclusion in relevant educational communities.

REFERENCES

1. Smirnov, I.B., Differentiation of Students by Academic Performance in a Social Network, *Cand. Sci. (Ped.) Dissertation*, Moscow: National Research University Higher School of Economics, 2018, p. 210. (In Russian.)
2. Nikitkov, A. and Sainty, B., The Role of Social Media in Influencing Career Success, *International Journal of Accounting & Information Management*, 2014, vol. 22, no. 4, pp. 273–294. DOI 10.1108/IJAIM-02-2014-0009
3. Dokuka, S., Valeeva, D., and Yudkevich, M., How Academic Achievement Spreads: The Role of Distinct Social Networks in Academic Performance Diffusion, *PLOS ONE*, 2020, vol. 15, no. 7, art. no. e0236737. DOI: 10.1371/journal.pone.0236737
4. Tafesse, W., Social Networking Sites Use and College Students' Academic Performance: Testing for an Inverted U-shaped Relationship Using Automated Mobile App Usage Data, *International Journal of Educational Technology in Higher Education*, 2022, vol. 19, art. no. 16. DOI: 10.1186/s41239-022-00322-0
5. Al Mosharrafa, R., Akther, T., and Siddique, F.K., Impact of Social Media Usage on Academic Performance of University Students: Mediating Role of Mental Health under a Cross-sectional Study in Bangladesh, *Health Science Reports*, 2024, vol. 7, no. 1, art. no. e1788. DOI 10.1002/hsr2.1788
6. Chkhartishvili, A., Gubanov, D.A., Melnichuk, V.S., and Sych, V.V., Exploratory Data Analysis and Natural Language Processing Model for Analysis and Identification of the Dynamics of COVID-19 Vaccine Opinions on Small Datasets, *Advances in Systems Science and Applications*, 2023, vol. 23, no. 3, pp. 108–126. DOI: 10.25728/assa.2023.23.3.1381
7. Levenshtein, V.I., Binary Codes Capable of Correcting Deletions, Insertions and Reversals, *Soviet Physics Doklady*, 1966, vol. 10, no. 8, pp. 707–710.
8. Kashpur, V.V., Gubanov, A.V., Feshchenko, A.V., et al., Correlation Between Academic Achievements of High School Students and Their Digital Shadow in Social Network, *Pedagogy and Education*, 2020, no. 4, pp. 37–51. (In Russian.)
9. Smirnov, I.B., Estimating Educational Outcomes from Students' Short Texts on Social Media, *EPJ Data Science*, 2020, vol. 9, art. no. 27. DOI: 10.1140/epjds/s13688-020-00238-1
10. Dudina, V.A., Anisimovskaya, A.I., Gorchenko, A.L., et al., Assessing the Influence of Social Networks and Other Factors on the Academic Performance of Students at the Faculty of Biology, *Molodoi Uchenyi*, 2019, no. 48(286), pp. 438–440. (In Russian.)
11. Giunchiglia, F., Zeni, M., Gobbi, E., et al., Mobile Social Media Usage and Academic Performance, *Computers in Human Behavior*, 2018, vol. 82, pp. 177–185. DOI: 10.1016/j.chb.2017.12.041
12. Rousseeuw, P.J., Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis, *Journal of Computational and Applied Mathematics*, 1987, vol. 20, pp. 53–65. DOI: 10.1016/0377-0427(87)90125-7
13. Tibshirani, R., Walther, G., and Hastie, T., Estimating the Number of Clusters via the Gap Statistic, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2001, vol. 63, no. 2, pp. 411–423. DOI: 10.1111/1467-9868.00293
14. Gorshkov, S., Ignatov, D.I., Chernysheva, A.Yu., et al., Identifying Top-Performing Students via VKontakte Social Media Communities Using Advanced NLP Techniques, *IEEE Access*, 2025, vol. 13, pp. 962–979. DOI: 10.1109/ACCESS.2024.3521857
15. Jaccard, P., Étude Comparative de la distribution florale dans une portion des Alpes et du Jura, *Bulletin de la Société Vaudoise des Sciences Naturelles*, 1901, vol. 37, pp. 547–579. DOI: 10.5169/seals-266450
16. Blondel, V.D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E., Fast Unfolding of Communities in Large Networks, *Journal of Statistical Mechanics: Theory and Experiment*, 2008, no. 10, art. no. P10008. DOI: 10.1088/1742-5468/2008/10/P10008
17. Barnard, G.A., Statistical Inference, *Journal of the Royal Statistical Society: Series B (Methodological)*, 1949, vol. 11, no. 2, pp. 115–139. DOI: 10.1111/j.2517-6161.1949.tb00028.x
18. Cohen, J., *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed., Hillsdale, NJ: Lawrence Erlbaum Associates, 1988.

This paper was recommended for publication by RAS Academician D.A. Novikov, a member of the Editorial Board.

Received May 4, 2025, and revised July 27, 2025. Accepted July 28, 2025.

Author information

Sych, Vladislav Vital'evich. Student, Bauman Moscow State Technical University, Moscow, Russia; engineer, Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia
✉ sycheslavovich@gmail.com
ORCID iD: https://orcid.org/0009-0001-6787-8948

Cite this paper

Sych, V.V., The Correlation Between the Social Network Activity and Academic Performance of Technical University Students: A Case Study of *VKontakte*. *Control Sciences* **4**, 43–54 (2025).

Translated into English by *Alexander Yu. Mazurov*,
Cand. Sci. (Phys.–Math.),
Trapeznikov Institute of Control Sciences,
Russian Academy of Sciences, Moscow, Russia
✉ alexander.mazurov08@gmail.com