Research Article

Possibilities of Predicting a Person's Substance Use Behaviour and Mental Health Through Social Media in a COVID-19 Crisis Context

Vsevolod Konstantinov [1, *], Pavel Ustin [2], Leonid Popov [2]

1. Penza State University, 40 Krasnaya Street, Penza, 440026, Russia; E-Mail: konstantinov_vse@mail.ru
2. Kazan Federal University, 18, Kremlyovskaya street, Kazan, Russia; E-Mails: Pavel.Ustin@kpfu.ru; leonid.popov@inbox.ru

* **Correspondence:** Vsevolod Konstantinov; E-Mail: konstantinov_vse@mail.ru

Academic Editor: Marianna Mazza

Special Issue: Mental Health and Substance Use Disorder During the COVID-19 Crisis

Abstract

The negative psychological consequences of the COVID-19 pandemic and the forced isolation of a large proportion of people worldwide have demonstrated the need to develop ways and technologies to reduce the effects of sudden threats of this type. The basis of any practical work to minimize the negative psychological consequences of the COVID-19 pandemic associated with substance use is the monitoring and diagnosis of the psychological resources of the individual. The article aims to show the possibilities of predicting the behavior of an individual through the content analysis of posts and reposts of their profile on the social network VKontakte on the example of the propensity to use psychoactive substances and to substantiate the possibilities of optimizing and automating such prediction through the use of category markers. Content analysis was carried out by latent semantic analysis of texts extracted from posts and reposts of VKontakte social network users with subsequent content analysis through selecting markers - category words. As a result, a categorical grid was built, which increases the efficiency of content analysis of

posts and reposts of users and is suitable for further automation of such research by machine learning methods.

Keywords

COVID-19; social networks; predictors; qualitative analysis; success; personal profile; substance use

1. Introduction

The era of the COVID-19 pandemic, which led to the social isolation of a large part of the world population in 2020, has actualized new risks and challenges that humanity may face shortly. Analyses of studies show that one of the consequences of the massive spread of coronavirus infection and social isolation has been a significant increase in suicide [1, 2], an increase in loneliness and depression [1, 3], psychological distress and increased inequalities in communication [4, 5], an increase in anxiety, insomnia and maladaptation [6, 7], a decline in mental health and well-being [8, 9], and a sharp increase in substance use [10, 11].

This raises the problem of preventing and mitigating the adverse effects on the population caused by sudden events such as the recent COVID-19 pandemic. One of the resources for reducing such negative phenomena associated with substance use is the development of tools and technologies that would allow for the automated prediction of individual behavior and mental health on multi-million samples. A variant of the development of such technologies is social networks - virtual social and psychological platforms, which are one of the main ways for people to interact in the modern world. The reason for the popularity and mass use of social networks by a significant part of the population is related to the intensive development of information technologies and a substantial transformation of the traditional model of social reality. Consequently, the virtualization of social processes, when there is an intensive merger of social and virtual realities and the emergence of the phenomenon of "I - virtual", reflecting the "I - real" of the individual, expands its social and psychological boundaries.

For example, a study by L.E. Buffardi and W.K. Campbell shows the manifestations of users' narcissism on Facebook through the specific traits of their social interaction, self-promotion, and the attractiveness of the main photo [12]. V. Barker showed the peculiarities of motivation for using social networks and the gender-specific social services visiting among adolescents [13]. F.G. Deters and M.R. Mehl showed that the experimental increase in the frequency of status updates of Facebook users reduced the degree of their loneliness as users began to feel more connected with their friends daily [14]. A.L. Forest and J.V. Wood showed that people with low self-esteem regard Facebook as an attractive platform for their self-expansion, but this social network cannot enrich their interpersonal life since their messages are characterized by low positivity and high negativity, which cause other users' unwelcome reaction [15]. According to F.T. McAndrew's, H.S. Jeong's research results, the most active Facebook users are women, young people, and those not currently in a severe relationship [16]. C.J. Blomfield Neira and B.L. Barber show that social media act as one of the significant factors in the positive adaptation of young people to real life and have no negative impact on them [17].

Interaction in social networks occurs both through direct communication of people with each other and through various products of their virtual activity. The effects of a person's virtual activity are understood as multiple elements of their profile in social networks - metrics or psychometrics. These include posts, reposts, photo-audio-video content, likes, communities, etc. These metrics, being products of a person's virtual activity, form their virtual image, which in turn reflects their authentic self. Consequently, the metrics of personal profiles in social networks are the predictors that potentially open vast opportunities for predicting a person's behavior in real life.

This statement is proved by numerous studies revealing the possibility of predicting a person's vigilance characteristics, behavior, and personal characteristics in real life via quantitative and qualitative indicators of their profile in social networks (number and content of posts, likes, avatars, photos, etc.). Among them is the possibility of predicting extraversion, devotion, self-control, emotional resilience, and expressiveness [18], narcissism and extraversion [19], bashfulness and loneliness [20], self-esteem, self-respect [21], sexual orientation [22], mental well-being [23], self-esteem [24], anti-social behavioral trends [25], academic progress [26], etc.

Predictive markers can be different variables, which refer to personal profile elements or the special traits of online activity in a virtual environment. For example, the study [27] proves the possibility of predicting the unique characteristics of a social network user by reference to the features of his musical preferences in a virtual environment.

The study [28] shows that through the number of a user's likes of various content on social networks, one can predict ethnicity and party affiliation, gender, sexual orientation, and religious confession with a high degree of probability.

Facial images [29], Main Profile frame, The Wall, and Photos [30], characteristic features of status updates [14, 19, 30], and features of participating in various online activities can also serve as predictive markers of activity, behavior and personality traits on social networks [24].

Qualitative analysis of texts in social networks and predictive possibilities based on psycho-linguistic analysis are presented in the papers [31-33]. The algorithm of formalization of text units in qualitative research through indexing is substantiated in them.

The research [31] reveals algorithms for predicting the professional interests of an individual through linguistic and statistical analysis of user texts from the VKontakte social network. Via normalization of input texts, the authors developed three thesauri for automatic indexing (Humanities, Mathematics, and Natural Science professional interests), which made it possible to show the effectiveness of a text classifier based on linear logistic regression.

The researcher [32] deals with the task of highlighting text styles based on statistical analysis. Several scientific, official and fictional literature-style frames were created and indexed for numerical traits. Based on these features, a tree-based decision classifier was constructed. This classifier showed the best results for official-style texts and the worst for media texts.

The research shows possible methods of distinguishing a text mode in a natural language from the point of view of both the text author's attitude to the message and the reference of the reality message. The main result is that all methods involve initial indexing followed by text markup, which computer-assisted instruction can do.

Thus, products of users' virtual activity in social networks enlarge the possibilities of predicting behavior, training, and personal traits via quantity and content analysis of their social network profile details (posts, reposts, likes, etc.). Analyzing user texts in social networks through indexing and creating classifiers is one of the informative methods in this case.

The proposed study reflects the solution of one of the tasks within the framework of the ongoing project on developing a neural network psychometric model of cognitive-behavioral predictors of personal life activity based on social networks.

The paper aims to show the possibilities of predicting the behavior of a person related to substance use through the content analysis of posts and reposts of their profile in the social network VKontakte on the example of propensity and to substantiate the possibilities of optimization and automation of such prediction through the use of marker-categories.

2. Methods

The study was based on empirical data based on content analysis of posts and reposts of users of the VKontakte social network posted in their profiles. The VKontakte social network is the Russian analog of Facebook. This virtual platform connects millions of users (according to the company's reporting data, more than 100 million active users per month) and allows them to communicate and interact with each other. This social network was chosen due to its popularity among Russian users and the difficulties in working with international virtual platforms (such as Instagram, Facebook, etc.). In addition, this network is quite informative and includes many personal profile metrics that can be analyzed.

The study is based on empirical data obtained by monitoring the personal pages of users of the social network "VKontakte" and analyzing the content of their posts and reposts. A specially developed information-analytical system conducted Monitoring and data collection from open personal profiles. Collection of texts of posts and reposts within the framework of this information and the analytical method was carried out with the help of a script in Python language using the library requests through the open API interface of the social network "VKontakte" (VK API) from the available access to personal profiles of users of the social network "VKontakte".

First, a research sample was formed that included more than 117,000 people with accounts on the VKontakte social network. The sampling frame was made up of social network users who agreed to participate in an online survey, where they were asked voluntarily to evaluate how often they had consumed alcohol in the past year. Respondents were given 5 options: 1. «I do not consume alcohol in principle»; 2. «I do it very rarely, several times a year»; 3. «I do not often consume, several times a month»; 4. «I do it quite often, several times a week»; 5. «I consume very often, almost every day». Respondents who were offered to answer the questions were chosen randomly with the help of an automated call-out system. The finite number of users who agreed to participate in the survey and answered the questions was more than 80,000.

Based on the data analysis and structuring results, the list of subjects was refined: issues with empty pages, no posts and reposts, and no activity in their personal profiles were excluded. Subjects who closed their social networking profiles were also excluded.

The final data processing was performed on a sample of more than 60,000 subjects. The total number of extracted texts was more than 20,000,000 units.

Regarding the given answers, respondents were divided into 5 groups, where 2 major research samples were identified: Group 5 - non-substance users, and Group 1 – substance users. The remaining groups (2,3,4) performed as interim samples, allowing further assessment of both the main differences in the content of posts and reposts of the persons who consumed and did not

consume psychoactive substances and checking the assumption about the monotonous consistency of increasing use of category words from group to group.

Content analysis of the texts of users' posts and reposts was done using natural language processing methods. The study solved the problem of transforming arbitrary text into a mathematically formalized and labeled array of data - numbers. To normalize random texts, the method of text index construction, which belongs to latent semantic analysis (LSA) methods, was used. The method is based on the principles of factor analysis, particularly the identification of latent relationships of the phenomena or objects under study. The choice of this method for qualitative research in the study was because it provides the opportunity to obtain results suitable for both manual analysis using rules and machine learning.

This method divides the text into individual words and calculates some statistical characteristics of terms, for example, the number of uses of a word, the number of sources (posts and reposts) in which the word is used, etc. Thus, the formalization of texts by indexing them allowed us to identify respondents' most frequently used terms associated with substance use. This indexing is done by combining text indices from individual profile pages into generalized indices suitable for analyzing the total sample of profiles.

After the formalization stage, the content analysis method was used for the substantive study of the results with the construction of a categorical grid - a working template for the substantive analysis of posts and reposts with identified and coded categories and subcategories. The content analysis determined the novelty of the approach since latent semantic analysis itself does not involve intellectual analysis of indexing results by the researcher, directly using the resulting index as formalised input data for building classifiers using machine learning. Using a categorical grid allowed us to reduce the dimensionality of the data set for machine learning by including only those significant words both statistically and within the theoretical model proposed in the study. This, in turn, allowed optimization of the index-based classifier model in terms of performance (accelerated machine learning) and protected against overtraining, i.e., the detection of latent statistical relationships that are irrelevant to the feature under study.

In the study, we used the content analysis method to differentiate the content of posts and reposts of users of social networks into specific classification units - empirical indicators. Based on the selected practical hands, we identified the indices of the texts of posts and reposts with the identified categories - we assigned individual indexes to one or another empirical indicator.

Verification of the categorical grid, i.e., testing of the hypothesis about its effectiveness (in particular, words-categories from the network of predictors of substance use should reflect a person's belonging to a group related to substance use, and vice versa) was carried out by studying the frequency characteristics of words-categories on generalized indices of all groups of people. Thus, the solution to the task of content analysis of posts and reposts included the following stages of work:

Collection of texts of posts and reposts from the pages of social network users among the sample of people considered in the framework of substance use research;

Construction of indices of texts of posts and reposts for each person from the sample, considering his/her profile in the social network as a document in terms of the indexing method;

Construction of generalized indices for groups of people in the sample characterized as associated and not associated with substance use, as well as for intermediate groups;

Aggregation of the generalized indices to find words that are unique to the group of those associated and not associated with substance use;

Construction of a categorical grid based on the obtained aggregated indices and formulation of the selected flat grid's effectiveness hypothesis.

Verification of the hypothesis of the effectiveness of the selected categorical grid by analyzing the frequency characteristics of words-categories from it within the framework of all obtained aggregated indices.

3. Results

The indexing process (Construction of indexes of texts of posts and reposts) started with the tokenization of texts extracted from the bars and reposts - their division into separate token words. During tokenization, the text was cleared of punctuation and odd symbols, and some different semantic elements of interest were distinguished - hashtags and URL links. For these elements, separate indexes were created due to the unique character of their semantics compared to common words. Collocations from the regex library were used for tokenization.

Before changing common words into indices, we transformed them into a definite root form to consider different variants of the same word in the text. We used the pymorphy2 library, which uses the Russian language corpus OpenCorpora to bring words to their regular form.

After processing, each word was given an index reflecting the number of uses of the word in posts and reposts from the user's social media profile. Thus, indexing of the form "Word" - "Number of uses" occurred for each shape from the total sample separately for common words, for hashtags, for URL links from posts and from shares, resulting in 6 indices per profile: 1. Index of post texts; 2. Index of hashtags of posts; 3. Index of post links; 4. Index of texts shares; 5. Repost hashtag index; 6. Index of shares references. Table 1 shows a fragment of the post-text index for one personal profile.

Table 1 Example of post index of a single profile in the VKontakte social network.

Word	Times used
mine	40
to	39
all	38
publicize	38
photo	38

Next, generalised indices were constructed. The general research sample (more than 60,000 research-subjects) was divided into 5 groups according to their substance use level. Group 1 - very frequent substance users (N = 4666); Group 2 - frequent substance users (N = 7203); Group 3 - infrequent substance users (N = 25926); Group 4 - very infrequent substance users (N = 14701); Group 5 - non substance users (N = 8998). The main analysis of differences was performed on the two polar groups, Group 1 and Group 5. The other groups were used to analyze the nature of category word differences when moving from group to group

In the process of generalized indices aggregation, subsets of indices were identified as indicators of the most commonly used words, unique or more characteristic of the five groups (from non-substance consuming to maximum psychoactive substance consuming).

The generic approach was used to form these subsets. i.e. the gradation of words according to their qualitative characteristics (the frequency of use), selection of top N words in the sorted index and calculating the difference between the generalized indices for all groups in terms of set theory. It was precisely this difference that contained words unique to the group. (Table 2). The approach was implemented for N = 1000 words for each of the five groups. All terms beyond our defined limit in N = 1000 were excluded from further analysis.

Table 2 A fragment of a set of unique words used by people associated with substance and non-substance use.

Posting indices for those unrelated to substance use		Indexes of posts related to substance use	
Word	Frequency of use rating	Word	Frequency of use rating
apartment	1	Cross over	1
respect	2	reminiscence	2
center	3	hangout	3
news	4	like	4
programme	5	buzz	5
sign	6	Birthday party	6

Further, a categorical grid of cognitive-behavioral predictors of substance use, including categories and subcategories, was constructed based on the selected word indices. This grid was based on the theoretical model proposed based on S.L. Rubinstein's concept of personality [34] and L.M. Popov's concept of the psychological organization of a person as a subject of development and self-development [35]. This model included 6 main categories. The first category - I am (Demographic characteristics), had the following subcategories: age, gender, income, profession, nationality. The second category - I want (Motivational orientation) includes the following subcategories: needs, motives, goals, meanings, self-actualization, self-development, and will. The third category - I can (Personality characteristics) includes the following subcategories: abilities, personality traits, openness to experience, and orientation. The fourth category - I must (Volitional Characteristics), consists of the following subcategories: will and duty. The fifth category - I feel (Emotional characteristics), includes the following subcategories: experiences, emotions, and feelings. The sixth category - I act (Behavioural Characteristics and Pastimes) contains the following subcategories: activities, planning of actions and operations, leisure time, and hobbies. The defined subsets of indices were further divided into the categories of the categorical grid.

The evaluation of the effectiveness of the proposed categorical grid consisted of checking the validity of the selected words. That is, each selected word is assigned to categories. Subcategories had to be truly reflective of their corresponding group, namely, used to a greater extent by people in that group (e.g., not associated with substance use) and to a lesser extent by people in other groups (e.g., not associated with substance use and associated with substance use). In addition, the assumption of monotonicity in the pattern of increasing use of category words from group to

group was tested; that is, if a category word is drawn from the categorical grid for the most unrelated substance users, it is most used among unrelated substance users (group 5), least used in group 4, least used in group 3, least used in group 2, and least used among related substance users (group 1). Thus, the test examined the number of people using the word category among each of the groups for each word category from each grid constructed.

The verification results showed that most of the word categories were correctly selected, with the frequency of occurrence of words not exceeding 10-15% for posts and 20-30% for reposts, consistent with the frequency characteristics of natural language vocabulary. These frequency ranges can be used to benchmark high-frequency word characteristics in content analysis.

The test results show that most of the word categories are characterized by a monotonous increase in the degree of use of word categories from group to group, which suggests that the intermediate position of groups 2, 3, and 4 between unrelated and substance use-related subjects is correct. The categorical grid for issues associated with substance use in this test no longer shows a bias towards the characterization of non-associated substance use, indicating that its construction is correct. Thus, the categorical grid is constructed correctly. It allows predicting the propensity to use psychoactive substances based on the content analysis of the texts of posts and reposts, i.e., it sets markers of tendency to use psychoactive substances in the form of certain words-categories. Using content analysis based on the constructed categorical grid, the results presented in Table 3 and Table 4 were obtained.

Table 3 The most frequently used main categories in the text of the posts of subjects not related to substance use and associated with substance use.

Posting indices for those unrelated to substance use		Indexes of posts related to substance use	
Main Category	Frequency of use rating	Main Category	Frequency of use rating
I act	40%	I act	26%
I want	30%	I want	24%
I am	12%	I am	2%
I can	10%	I can	5%
I should	7%	I should	4%
I feel	2%	I feel	38%

Table 4 The most frequently used main categories in the text of the reposts of subjects not related to substance use and associated with substance use.

Posting indices for those unrelated to substance use		Indexes of posts related to substance use	
Main Category	Frequency of use rating	Main Category	Frequency of use rating
I act	30%	I act	25%
I want	23%	I want	17%
I am	24%	I am	19%
I can	14%	I can	-
I should	9%	I should	-
I feel	-	I feel	38%

Consequently, behavioral characteristics, in particular, characterize the non-substance use group. They need to engage in activities and plan their actions through sequential tasks. They are motivationally orientated, know what they want, seek to fulfill their needs, engage in self-development, and identify themselves in a particular professional role. Often, their orientation is not characterized by personality characteristics or volitional qualities but rather by the motivational side. They ignore emotional factors and focus on results.

The substance use group has predominantly emotional characteristics, i.e., they mainly think about their experiences, emotions, and feelings. They also have behavioral traits such as action planning and preoccupation with activities but are primarily concerned with their emotional state. They know what they want, give meaning to their actions, fulfill their needs, self-actualize, and develop personally. Their abilities and personality traits are not so important to them. They are free to choose their actions and desires.

4. Conclusion

The study results showed the effectiveness of using markers in the form of categories and subcategories in the content analysis of posts and reposts of users of the VKontakte social network, which increases the effectiveness of predicting their behavior and personality characteristics related to substance use. The proposed algorithm formalizes arbitrary text into a set of word categories that can be numerically represented, for example, as a vector, which makes it suitable as a data source for the machine learning-based neural network model under development. A set of arbitrary texts from a social network user profile is indexed and processed by statistical methods using a categorical grid to identify characteristic words-markers associated with substance use, after which the resulting set of numerical statistical characteristics of these words serves as input data for the neural network model. It, in turn, outputs a prediction of the propensity to use psychoactive substances, being trained on a training set formed based on the collected data. The obtained algorithm is currently being tested on the neural network model for predicting the subject's substance use. .Further prospects for the study are seen in expanding the list of psychoactive substances used by respondents (particularly smokers and non-smokers) and in expanding the sample to include users from countries other than the post-Soviet space.

5. Limitations of the Study

The results were obtained on the basis of empirical evidence downloaded from the VKontakte social network, where the users are primarily residents of the post-Soviet states with their inherent regional and cultural characteristics. Consequently, transferring the interpretation of the results to the residents of other regions with different regional and cultural factors requires special precautions and additional research.

Author Contributions

Conceptualization and methodology, V.K., L.P. and P.U.; software, P.U.; investigation, data mining, V.K., P.U.; writing—review and editing, V.K. and P.U.; supervision, V.K. All Authors contributed to write paper and approved the final manuscript.

Funding

The study (all theoretical and empirical tasks of the research presented in this paper) was supported by a grant from the Russian Science Foundation (Project No. 19-18-00253, "Neural network psychometric model of cognitive-behavioral predictors of life activity of a person on the basis of social networks"), https://rscf.ru/project/19-18-00253/.

Competing Interests

The authors have declared that no competing interests exist.

References

1. Ahmed F, Sifat RI. Strategic assessment of mental health and suicide amid COVID-19 pandemic in Bangladesh. Int J Health Plann Manage. 2021; 36: 980-985.
2. Sher L. The impact of the COVID-19 pandemic on suicide rates. QJM. 2020; 113: 707-712.
3. Krendl AC, Perry BL. The impact of sheltering in place during the COVID-19 pandemic on older adults' social and mental well-being. J Gerontol B. 2021; 76: e53-e58.
4. Sifat RI, Ahmed F, Miah MR, Khisa M. Effects of COVID-19 on livelihood, health, and psychology of hijra population: Insights from Dhaka, Bangladesh. J Homosex. 2023; 70: 2096-2112.
5. Chu IY, Alam P, Larson HJ, Lin L. Social consequences of mass quarantine during epidemics: A systematic review with implications for the COVID-19 response. J Travel Med. 2020; 27: taaa192.
6. Rossi R, Socci V, Talevi D, Mensi S, Niolu C, Pacitti F, et al. COVID-19 pandemic and lockdown measures impact on mental health among the general population in Italy. Front Psychiatry. 2020; 11: 790.
7. Tull MT, Edmonds KA, Scamaldo KM, Richmond JR, Rose JP, Gratz KL. Psychological outcomes associated with stay-at-home orders and the perceived impact of COVID-19 on daily life. Psychiatry Res. 2020; 289: 113098.
8. Ahmed F, Sifat RI. The impact of the COVID-19 pandemic on the mental health of the rickshaw-puller in Bangladesh. J Loss Trauma. 2021; 26: 782-789.
9. Sani G, Janiri D, Di Nicola M, Janiri L, Ferretti S, Chieffo D. Mental health during and after the COVID-19 emergency in Italy. Psychiatry Clin Neurosci. 2020; 74: 372.
10. Isralowitz R, Khamenka N, Konstantinov V, Gritsenko V, Reznik A. Fear, depression, substance misuse and related conditions among multi-national medical students at the peak of the COVID-19 epidemic. J Loss Trauma. 2021; 26: 401-404.
11. Gritsenko V, Skugarevsky O, Konstantinov V, Khamenka N, Marinova T, Reznik A, et al. COVID 19 fear, stress, anxiety, and substance use among Russian and Belarusian university students. Int J Ment Health Addict. 2021; 19: 2362-2368.
12. Buffardi LE, Campbell WK. Narcissism and social networking web sites. Pers Soc Psychol Bull. 2008; 34: 1303-1314.
13. Barker V. Older adolescents' motivations for social network site use: The influence of gender, group identity, and collective self-esteem. Cyberpsychol Behav. 2009; 12: 209-213.
14. Deters FG, Mehl MR. Does posting Facebook status updates increase or decrease loneliness? An online social networking experiment. Soc Psychol Personal Sci. 2013; 4: 579-586.

15. Forest AL, Wood JV. When social networking is not working: Individuals with low self-esteem recognize but do not reap the benefits of self-disclosure on Facebook. Psychol Sci. 2012; 23: 295-302.

16. McAndrew FT, Jeong HS. Who does what on Facebook? Age, sex, and relationship status as predictors of Facebook use. Comput Hum Behav. 2012; 28: 2359-2365.

17. Blomfield Neira CJ, Barber BL. Social networking site use: Linked to adolescents' social self-concept, self-esteem, and depressed mood. Aust J Psychol. 2014; 66: 56-64.

18. Kluemper DH, Rosen PA, Mossholder KW. Social networking websites, personality ratings, and the organizational context: More than meets the eye?[1] J Appl Soc Psychol. 2012; 42: 1143-1172.

19. Ong EY, Ang RP, Ho JC, Lim JC, Goh DH, Lee CS, et al. Narcissism, extraversion and adolescents' self-presentation on Facebook. Pers Individ Differ. 2011; 50: 180-185.

20. Ryan T, Xenos S. Who uses Facebook? An investigation into the relationship between the Big Five, shyness, narcissism, loneliness, and Facebook usage. Comput Hum Behav. 2011; 27: 1658-1664.

21. Sorokowski P, Sorokowska A, Oleszkiewicz A, Frackowiak T, Huk A, Pisanski K. Selfie posting behaviors are associated with narcissism among men. Pers Individ Differ. 2015; 85: 123-127.

22. Wang Y, Kosinski M. Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. J Pers Soc Psychol. 2018; 114: 246-257.

23. Manago AM, Taylor T, Greenfield PM. Me and my 400 friends: The anatomy of college students' Facebook networks, their communication patterns, and well-being. Dev Psychol. 2012; 48: 369-380.

24. Tazghini S, Siedlecki KL. A mixed method approach to examining Facebook use and its relationship to self-esteem. Comput Hum Behav. 2013; 29: 827-832.

25. Kosinski M. Facial width-to-height ratio does not predict self-reported behavioral tendencies. Psychol Sci. 2017; 28: 1675-1682.

26. Polivanova K, Smirnov I. What's in my profile: VKontakte data as a tool for studying the interests of modern teenagers. Educ Stud. 2017: 134-152. doi: 10.17323/1814-9545-2017-2-134-152.

27. Nave G, Minxha J, Greenberg DM, Kosinski M, Stillwell D, Rentfrow J. Musical preferences predict personality: Evidence from active listening and Facebook likes. Psychol Sci. 2018; 29: 1145-1158.

28. Kosinski M, Stillwell D, Graepel T. Private traits and attributes are predictable from digital records of human behavior. Proc Natl Acad Sci U S A. 2013; 110: 5802-5805.

29. Segalin C, Celli F, Polonio L, Kosinski M, Stillwell D, Sebe N, et al. What your Facebook profile picture reveals about your personality. Proceedings of the 25th ACM international conference on Multimedia; 2017 Oct 23-27; Mountain View, CA, USA. New York: ACM Journals.

30. Chen L, Gong T, Kosinski M, Stillwell D, Davidson RL. Building a profile of subjective well-being for social media users. PloS One. 2017; 12: e0187278.

31. Stepanenko AA, Shilyaev KS, Rezanova ZI. Attribution of professional interests of social network users based on subject-oriented groups and personal pages [In Russian]. Vestnik Tomskogo gosudarstvennogo universiteta Filologiya. 2018: 130-144. doi: 10.17223/19986645/52/8.

32. Dubovik AR. Automatic determination of stylistic belonging of texts by their statistical parameters (In Russian). Comput Linguist Comput Ontol. 2018: 29-45. doi: 10.17586/2541-9781-2017-1-29-45.

33. Egikyan S. Up-to-date methods of the modality analysis in natural language texts (In Russian). Program Syst Theory Appl. 2017; 8: 133-167. doi: 10.25209/2079-3316-2017-8-3-133-167.

34. Rubinstein SL. Being and consciousness. Man and the world. St. Petersburg: Peter; 2003. 512p.

35. Popov L, Ustin P. Cognitive-behavioral concept and possibilities of its implementation in students' life activity (In Russian). Psikhologicheskii zhurnal. 2020; 42: 26-35. doi: 10.31857/s020595920013324-2.