# Classification of Russian Internet users: preliminary results of cluster analysis

Vladimir Dmitrievich Nechaev[1], Elena Victorovna Brodovskaya[1], Yurij Vladimirovich Kaira[2], Anna Yurievna Dombrovskaya[1]

[1]Moscow State University for the humanities M.A. Sholokhov
Verkhnyaya radishchevskaya street, 16-18, Moscow, 109240, Russia
[2]Russian Presidential Academy of National Economy and Public Administration (Oryol branch)
Bulvar Pobedy, 5a, Oryol, 302028, Russia

**Abstract.** The paper presents the preliminary results of the cluster analysis devoted to the database research of «World Internet Project in Russia – 2012». The purpose of cluster analysis is to classify the respondents in accordance with the characteristics of the Internet content they use and the dominant roles they play in Internet communication. Social profiles of different user groups in Russia can help to identify different strategies of their online behavior.

[Nechaev V.D., Brodovskaya E.V., Kaira Y.V., Dombrovskaya A.Y. **Classification of Russian Internet users: preliminary results of cluster analysis** *Life Sci J* 2014;11(12):330-335] (ISSN:1097-8135). http://www.lifesciencesite.com. 63.

**Keywords:** Internet communication, cluster analysis, social profiles of users

## Introduction

Modern Russia experiences dynamic changes in the sphere of information, thus Runet development turns out to be one of the most significant tendencies which transforms communicative, political and cultural environment of the Russian society. The Internet has become a centre where "traditional" values are being transformed whereas "new" values and behavior models of Russian citizens are actively crystallizing.

Interdependence of communicative and social components of virtual space is an important research subject for a number of Russian politologists and scientists all over the world (B.Barber [1], M.Yu Pavlyutenkova [2], A.N. Chumikov [3], E.V. Morozova [4], P. Norris [5], L.M. Weber [6] and others).

They analyses political effects of Internet communication. Special attention is paid to the networking approach and studies of nature, structure and functions of online networks. There are a number of sociological studies devoted to Runet users and their characteristics, [7, 8] one of them is «World Internet Project in Russia - 2012» [9].

Among the leading roles chosen by Russian users one can find "shoppers" (14%), “bloggers” (15%), “music lovers” (17%), "gamers” (14%) [10]. The results of «World Internet Project in Russia - 2012» [9] helped to specify these categories. The respondents include 60% of gamers (24% of them are active gamers (play once a day or more often), 58% show interest in services and goods (14% of them are active shoppers (do it online once a month or more often); 80% listen to music online (9% are active music listeners (do it several times a day). Alongside with the traditional roles scientists managed to define the groups of researchers and pick-ups.

Taking all this into consideration, we defined six typical Runet roles existing at present and classified them in the following way:

- «bloggers» – users who have online diaries or blogs, post new content there (from several times a month upto several times a day); users who read online diaries or blogs (from several times a month upto several times a day);
- «gamers» –users who play online games (from several times a month upto several times a day);
- «music lovers» –users who listen to music online (from several times a month upto several times a day);
- «shoppers» –users who shop, pay their bills and book online; they also use Internet banking;
- «researchers» – users, who educate themselves through search of definitions, concepts or facts, some of them are distant learning students;
- «pick-ups» - users who have accounts on dating sites, make friends online.

The so-called web-surfers form a special group of users as they usually combine several roles. 45% of Runet users surf the Web, which can be considered as sufficient evidence for the existence of a certain subculture of Internet behavior.

## Method

To collect data used mass pool of Russians (N=1600, sample is representative for the type area of residence, district, gender and age). To verify the received data we clustered the selection of World Internet Project in Russia – 2012 with the help of

SPSS.17. It was necessary because the classification of Russian users based on the content they use and their involvement into Internet communication, is of theoretical nature and it only provides comparative analysis of sociological research data.

**Main results**

The following variables were defined for clusterisation (Table 1).

**Table 1.The list of variables**

| # | Variables |
|---|---|
| 1 | Intensity of e-mailing |
| 2 | Intensity of instant messaging |
| 3 | Intensity of chatting |
| 4 | Intensity of sending e-mails with enclosed files |
| 5 | Intensity of making Internet calls |
| 6 | Intensity of posting information in a personal blog |
| 7 | Intensity of posting pictures or photos online |
| 8 | Intensity of uploading audio/video files online |
| 9 | Intensity of posting messages or comments on discussion forums/chat-rooms |
| 10 | Intensity of status updating in a social network |
| 11 | Intensity of commenting |
| 12 | Intensity of reading/searching information on the Internet |
| 13 | Intensity of searching for travel information |
| 14 | Intensity of searching for job opportunities |
| 15 | Intensity of reading of Internet blogs |
| 17 | Intensity of reading/searching for jokes or humorous content |
| 18 | Intensity of search for health information |
| 19 | Intensity of playing online games |
| 20 | Intensity of listening/downloading music from the Internet |
| 21 | Intensity of listening/downloading video from the Internet |
| 22 | Intensity of listening to radio online |
| 23 | Intensity of using religious sites |
| 24 | Intensity of gambling and betting online |
| 25 | Intensity of web-surfing |
| 26 | Intensity of searching for information/reviews on products, services |
| 27 | Intensity of online shopping |
| 28 | Intensity of online booking |
| 29 | Intensity of paying bills online |
| 30 | Intensity of online banking |
| 31 | Intensity of investing into shares/securities/funds online |
| 32 | Intensity of search for a definition or concept |
| 33 | Intensity of search for a particular fact |
| 34 | Intensity of search for information on education |
| 35 | Intensity of distant learning |
| 36 | Intensity of visting sites with sexual content |
| 37 | Intensity of visiting social networks or sites with video content |

The cluster analysis has the following parameters:

– Ward's method;

– chi-square for ordinal variables;

– The standardization of variables was unnecessary because the scales of meanings have the same proportionality.

The number of clusters is determined by the comparison results received after the application of the following three methods:

The first method shows the table of merges (the difference between the number of observations for clusterisation (888) and the level of merging at which the distance coefficient between the objects leaps, showing the transition from a strongly bound to weakly bound state (884)) – 4 clusters (Table 2).

**Table 2. Final merging in cluster analysis**

| level | Cluster merges | | Coefficients | The first level of clusterisation | | Next level |
|---|---|---|---|---|---|---|
| | Cluster 1 | Cluster 2 | | Cluster 1 | Cluster 2 | |
| 873 | 21 | 29 | 1619,93 | 848 | 845 | 874 |
| 874 | 18 | 21 | 1629,164 | 853 | 873 | 880 |
| 875 | 30 | 52 | 1639,05 | 871 | 850 | 879 |
| 876 | 19 | 22 | 1649,275 | 860 | 857 | 881 |
| 877 | 32 | 37 | 1661,062 | 872 | 864 | 884 |
| 878 | 23 | 45 | 1672,921 | 868 | 867 | 883 |
| 879 | 30 | 181 | 1685,077 | 875 | 844 | 881 |
| 880 | 18 | 28 | 1698,133 | 874 | 866 | 882 |
| 881 | 19 | 30 | 1714,161 | 876 | 879 | 885 |
| 882 | 18 | 24 | 1730,477 | 880 | 870 | 885 |
| 883 | 23 | 27 | 1748,59 | 878 | 865 | 886 |
| **884** | **32** | **33** | **1774,289** | **877** | **815** | **886** |
| 885 | 18 | 19 | 1803,186 | 882 | 881 | 887 |
| 886 | 23 | 32 | 1835,167 | 883 | 884 | 887 |
| 887 | 18 | 23 | 1886,245 | 885 | 886 | 0 |

The second method is based on the icicle diagram (the number of groups of observations is graphically distributed into clusters by observations which do not belong to any of the clusters, starting from the first levels of merging) – 4-6 clusters.

The third method is based on the dendrogram (the number of clusters with minimal and compatible, close in-cluster scale distances) – 4 clusters. The scale distances are too large if the number of clusters is less than 4; when the number is bigger than 4, in-cluster distances become incompatible compared to other clusters.

Thus, the final number of clusters is 4 as such distribution has the highest discriminating potential.

**Results of clusterisation**

*Cluster 1 – Internet as information space*

The representatives of this cluster are the most active users of Internet services, all of which are practically equally used except Internet banking and

some other financial online transactions. This situation can be explained by the high level of anxiety about the security of financial operations online. These respondents also actively use audio and video content on the Internet, though they do not upload their own content. They also have a greater experience and a longer history of Internet communication, this group often uses mobile Internet applications. These respondents treat the Internet as a primary source of information and entertainment, though, they do not only communicate with reference groups virtually. As a result of their Internet communication, the respondents began to communicate more actively with people who have similar interests, friends, colleagues, family. They constantly combine being online with other activities. This group includes people who willingly express their political views on the Internet, they oppose the authorities control and are more inclined to trust the Internet information. (Table 3)

The social profile of the first cluster: slight prevalence of women, the average age is 28, level of education – intermediate, half of the representatives are employed, one third of them are students, the income level is medium, family status - most respondents are single; the family size is average in the sampling.

**Table 3. Semantic reasoning for cluster nomination based on the analysis of the respondents' answers**

| Variables | Cluster # 1 | Cluster # 2 | Cluster # 3 | Cluster # 4 |
|---|---|---|---|---|
| Primary sources of information | Hierarchy of information sources:<br>- Internet;<br>- face-to-face communication;<br>- TV | face-to-face communication as a primary source of information | face-to-face communication as a primary source of information | Hierarchy of information sources:<br>- Internet;<br>- face-to-face communication;<br>- TV |
| Perception of political functions of the Internet | Positive assessment of Internet political functions | Neutral assessment of Internet political functions | Negative assessment of Internet political functions | Neutral assessment of Internet political functions |
| Assessment of security and control on the Internet | Negative attitude to authorities control on the Internet | Respondents support absolute freedom of speech on the Internet | Positive attitude to authorities control on the Internet | Neutral attitude to authorities control on the Internet |
| Assessment of reliability of information on the Internet | Respondents think that information is reliable | Respondents think that information is unreliable | Respondents think that information is reliable | Respondents think that information is reliable |
| Level of anxiety about the security of financial transactions on the Internet | High level of anxiety connected with the security of financial transactions on the Internet | Low level of anxiety connected with the security of financial transactions on the Internet | Do not have any financial transactions on the Internet | Middle level of anxiety connected with the security of financial transactions on the Internet |
| Intesity of Internet use | Respondents use the Internet most often | The intensity of the Internet use is intermediate | The intensity of the Internet use is intermediate | The intensity of the Internet use is low |
| Ways of Internet access | Mainly home access | Respondents actively use various sources of Internet access | Access at home or at work | Mainly home access |
| Mobile Internet | These respondents use mobile internet most often | These respondents do not use mobile Internet very often | These respondents use mobile Internet most often | These respondents do not use mobile Internet |
| Duration of Internet use | A long period of Internet use | The longest period of Internet use | The shortest period of Internet use | The shortest period of Internet use |
| Assessment of the nature of the Internet impact on communication with various groups | Due to the Internet usage their communication with friends, colleagues and people with similar interests significantly increased, their communication with family members also slightly rose | Due to the Internet usage their communication with friends, colleagues and people with similar interests slightly increased | Due to the Internet usage their communication with friends and people with similar interests slightly increased | The Internet usage had no impact on the intensity of their interpersonal communication |
| Existence of the experience of combining Internet activities with some other ones | They always combine using the Internet with other activities | They sometimes combine using the Internet with other activities | They rarely combine using the Internet with other activities | They don't combine Internet activities with any other ones |
| Intensity of the usage of Internet services | Active users of all Internet information and communication services | Moderate users of all Internet information and communication services | Primarily use social networks | Use the Internet to transmit information rather than to communicate or entertain themselves |
| Type of preferable Internet content | Entertaining and audiovisual content prevails | Audiovisual content prevails | Audiovisual content prevails, active Internet radio listeners | Types of roles:<br>- web surfers;<br>- gamers;<br>- shoppers. |
| Gender characteristics | There are slightly more women | Equal ratio of men and women | There are slightly more men | There are significantly more women |
| Age group | Young people | Lower margin of the middle age | Young people | Higher margin of the middle age |
| Education level | Intermediate | Upper intermediate | The least educated | The most educated |
| Social status | The employed and almost one third is made up by students | The employed and only some people are students (cluster of the employed) | The employed and one third is made up by students (cluster of students) | The employed and retired |
| Subjective assessment of welfare | Intermediate level | High level | Intermediate level | Low level |
| Marital status | Single | Primarily married | Primarily single | Primarily married |
| Family size | Middle-sized families | Middle-sized families | The largest families | The smallest families |

*The second cluster – the Internet as a routine background*

Representatives of the second cluster perceive the Internet as a background of their daily routine which accompanies them wherever they go. These people have an insignificant experience of implementing mobile Internet devices, though they have the longest track record of the Internet connection. Due to the Internet usage they started to communicate slightly more with friends, colleagues and people who share similar interests. Among a wide range of the Internet services they prefer

communication ones as well as watching video- and listening to audio-content. They have a neutral attitude to any political activities on the Internet, which is probably determined by some fear to express their political views via the Internet. The Internet is a space of absolute speech freedom for them even though such democracy can entail promotion of extremist ideas. However, they demonstrate lack of trust to the credibility of the Internet content. They are the least concerned about providing any financial Internet security.

*Social profile of the second cluster*:

Equal ratio of men and women, average age is 34 years, education is upper intermediate, primarily employed, are characterized by the highest level of welfare, mostly married with a middle family size in the sample group.

*The third cluster - the Internet as a civilian space of social networks.*

Representatives of the third cluster have a more thorough attitude to the choice of both the Internet content and the format of the Internet communication and since they prefer social networks to all other communication opportunities, they become the most active listeners of the Internet radio and consume audio- and video materials with a moderate intensity. Primarily, they use the Internet at home, though apart from that they capitalize on all opportunities of the Internet connection at their working places. They are characterized by an insignificant experience and the shortest track record of the Internet usage, however, they use mobile Internet devices quite intensively. To all means of mass communication, including the Internet, they prefer face-to-face communication in real life, primarily with friends. Due to the Internet their communication with friends and people with similar interests slightly increased. Sometimes they combine being online with some other activities. Judging by their opinions, the Internet does not extend any political rights and freedoms and thus does not enable users either to have any noticeable impact on the authorities or to understand politicians' agenda better. They support the necessity to increase control over the Internet and at the same time they rely on the online information. No financial transactions are carried out by them on the Internet.

*Social profile of the third cluster*: men slightly prevail over women; an average age is 29 years; education is on the upper intermediate level; mainly employed people, however, there is the largest number of students; characterized by an average level of welfare; percentage of single respondents is higher than average and those who are married have the largest family size in the sample group.

*The fourth cluster - the Internet as an instrument of transmitting information and as an entertainment space for leisure time.*

Representatives of this cluster use the Internet more seldom than others, mainly for one-way transmission of information. However, representatives who make up this cluster turn out to be the most active web-surfers, online gamers, online shoppers and users of online financial services. They do not resort to mobile Internet devices as a rule and, on the whole, have the shortest track record of the Internet connection. According to subjective assessment of the respondents, online communication had no impact on the intensity of interpersonal communication. Among a wide range of Internet services they tend to choose communication ones, but they rarely use them because they prefer real-life communication with friends. They do not ordinarily combine online activities with any others. Their attitude to political activities on the web is quite neutral, they do not care for control over the Internet. At the same time they assess the credibility of online content critically. Moreover, they are not concerned about security of financial operations on the Internet.

*Social profile of the fourth cluster*:

Evident prevalence of women; an average age – 42 years (this is the oldest cluster); the education level is the highest in the sample group; they are primarily employed, but there is also a high percentage of the retired; are characterized by the lowest level of welfare; percentage of the divorced and widowed is higher than average; the smallest family size.

**Findings**

As a result of clustering on the basis of all the data yielded in «World Internet Project in Russia - 2012» we can draw the following conclusions.

*Firstly*, the initial working hypothesis on the differentiation of all Russian users according to the dominant roles which they choose in online communication was not proved. It does not mean though that these roles cannot be found in the virtual space. The identified peculiarity rather indicates lack of consolidated groups of users united by common roles. In our opinion, it is determined, on the one hand, by the asymmetry of the Internet penetration and spreading in Russia (a considerable differentiation of experience and technologies of the Internet communication which are available for various groups of users), and, on the other hand, by a "digital gap" between generations which is constantly

widening from the point of view of communicative competence online, meeting communicative demands and, on the whole, choosing a certain lifestyle.

*Secondly,* comparative study of the basic characteristics of all the four clusters demonstrates differentiation not on the dominant roles, but on such criteria as:

- intensity of the involvement in the Internet communication;
- extent of orientation on informational functions of the Internet (the first type);
- extent of orientation on political functions of the Internet (the second type);
- extent of orientation on economic functions of the Internet (the third type);
- extent of orientation on entertaining functions of the Internet (the fourth type);
- extent of orientation on specific instrumental functions of the Internet (such as transmission of messages) (the fifth type).

On the basis of the above mentioned differentiation several types of users' attitude to the global network can be identified. In the situation when intensity of involvement in the Internet communication coincides with the high extent of orientation of the 1st, 2nd and 4th types, virtual space is regarded as the basic source of information which is constantly consumed. Internet users choose and implement various roles online several times per day, consequently, no permanent dependence on any of them can be observed. Provisionally, such a position can be called *"informational man"*. These characteristics are very similar to the description of *the first cluster*.

In the case when intensity of usage differs in the average value and coincides with the extent of orientation of the 1st, 3rd and 4th types, the Internet is regarded as a background of everyday life. Roles which are realized with the help of the global network are as diverse as in the first cluster. However, first, their implementation has a *routine character*, secondly, almost no roles with political content are observed. This description is typical for the representatives of *the second cluster*.

In the situation when a middle level of intensity of using Internet services goes with the extent of orientation of the fourth type in the respondents' assessment, there appears a type of online behavior which is called *"the Internet for similar people"*. The space of the Internet communication is limited by communities in social networks (it is characteristic of representatives of *the third cluster*).

Finally, there comes the fourth combination of criteria: low intensity of the Internet communication complemented by the extent of orientation of the 3rd, 4th and 5th types forms a highly *instrumental perception* of the Internet which, in practice, can be narrowed down to a single function, to one or several roles. At the same time fixation of roles becomes possible thanks to the rational choice of a certain strategy of online behavior (*the fourth cluster*).

Consequently, differentiation of the users' attitude to the Internet and its basic functions is primarily determined by such a factor as a scale of communication space: the smaller it is, the more civilian are the users' demands and the more precise is their choice of traditional online roles.

*Thirdly,* the scale of the user's communication space depends on a number of socio-demographic and socio-economic factors. There are the following leading factors:

- age group;
- level of education;
- level of welfare;
- employment etc.

While conducting cluster analysis we managed to identify several significant correlations:

- the higher the respondents' level of education is, the more thorough and instrumental becomes their attitude to the Internet communication (*the fourth cluster*);
- the lower the level of education is, the smaller is the scale of communication chosen by the user (the third cluster);
- the higher the level of subjective assessment of the well-being is, the higher is the corresponding level of routinization in using the Internet (*the second cluster*);
- the lower the level of subjective assessment of the well-being is, the higher is the extent of orientation on specific instrumental functions of the Internet (*the fourth cluster*) etc.

It is necessary to point out that if taken separately from communication factors, socio-demographic ones "*do not work*". For instance, the expected correlation between the age group and the type of the consumed content was not identified.

At the same time, identification of clusters, i.e. consolidated groups of consumers, has both theoretical significance (excelling methods of studying the users' structure of roles) as well as practical importance (social profile-making of various Internet groups has serious prospects from the viewpoint of the development of the so called recommendation systems).

**Thanks**

With support from the Institute of Advamced Research and Technology Humanities opf Sholokhov Moscow State University for the

Humanities. The project "Youth on the Internet the differentiation of values and strategies of political behavior" (2013-2014).

**Corresponding Author:**
Dr.Nechaev Vladimir Dmitrievich
Moscow State University for the humanities M.A. Sholokhov,
Verkhnyaya radishchevskaya street, 16-18, Moscow, 109240, Russia

**References**

1. Barber, B., 2001. The uncertainty of digital politics: Democracy's uneasy relationship with information Technology. Harvard International Review, 23: 42-48.
2. Pavlyutenkova, M.Yu., 2012. Communication Mechanisms and Technologies of the GR-activity. GR-relations with the State: Theory, Practice and Mechanisms of Interaction between Business and Civil Society and the State. Study guide. Moscow: ROSSPEN, pp. 35.
3. Chumikov, A.N., 2009. Political "Communicavistics": Current Objectives and Technologies of Practical Implementation. Political studies, 5: 55-67.
4. Morozova, E.V. and I.V. Miroshnichenko, 2011. Networking Communities in the Conditions of Emergency Situations: New Opportunities for Citizens and for the Authorities. Political studies, 1: 140-152.
5. Norris, P., 2001. Digital Divide. Civic Engagement, Information Poverty, and the Internet Worldwide. Cambridge University Press, pp: 234.
6. Weber, L. M., A. Loumakis and J. Bergman, 2003.Who participates and why? An analysis of citizens on the Internet and the mass public. Social Science Computer Review, 21: 26-42.
7. Online Studies and Methodology of Social Sciences: New Horizons, New (and not so New) Difficulties, 2010. Regional Information Centre "North-East".
8. Monitoring of the online audience. ROMIR. Date Views 22.05.2013 www.romir.ru/.
9. World Internet Project. Date Views 01.06.2013 www.worldinternetproject.net/#news/.
10. Online people. Who are they? From a Demographic Portrait to a Behavioral One. Date Views 14.02.2014 www.onlinemonitor.ru/.

7/29/2014