Experimental study reliability and functional stability of the social graph

E K Algazinov¹, V V Garshina¹, V A Stepantsov¹ and E N Desyatirikova²

¹ Computer Science Department, Voronezh State University, Voronezh, Russia
² Department Automatic Control Systems Department, Voronezh State Technical University, Voronezh, Russia
E-mail: garshina@cs.vsu.ru

Abstract. The article is devoted to the results of an experimental study the integrity, reliability and efficiency of the social network graph as a complex network in terms of transferring messages between users. The computed network parameters that affect the evaluation of the efficiency of its functioning are determined; the integral coefficient of network stability loss is developed; formulated heuristic rules for assessing the functional stability of a social network based on the parameters of the social graph are formulated.

1. Introduction
In the field of information technology, a direction has been developed — the Social Network Analysis (SNA), which is devoted to research on the nature of interactions of people in various social groups and communities. In this area of research, subtasks have been formed that focus on both structural analytics and the semantic interpretation of information reflected in social networks (fields of application: sociology, marketing, political science, psychology). These studies demonstrate great practical importance both now and a great potential for development in the future.

Journals "Social Networks" and "Connections" are focused on the analyses of social networks as well as the international academic community INSNA — "The International Network of Social Network Analysts". Particular attention is paid to the analysis of social networks in Russia Mail.Ru Group, Yandex, the Russian summer school for information retrieval (RuSSIR). There are analytical web services for monitoring and analysis of social networks: IQBuzz, Brand Analytics, SEUSLAB, etc. The problems of social-network analysis are dealt with by Russian such scientists as I. A. Evin [1], A. A. Koblyakov, L. E. Zhukov, G. V. Gradosel'skaya, N. V. Korytnikova, T. E. Savitskaya, O. P. Kuznetsov, D. V. Lande [2], V. M. Sazanov, as well as foreign researchers: M. S. Granovetter [3], M. Newman, L. Freeman, S. Milgram, A. E. Barabashi [4], S. Storgats, D. Watts and others.

In the theory of complex networks, Internet communities are treated as a special case of a non-scaled network — the social graph (web graph) [4, 5]. In the social graph nodes will be groups (communities) and user pages (profiles). Ribs (links) will be links to posts, communities or other users, subscriptions to pages (profiles) and profiles of other people, "friendship" (users
are in the list of friends from each other), tags. User interaction is most effectively described by complex graphs, which are characterized by the "small world" effect [6], as well as a high degree of clustering and a small average interstitial path length. These networks have a probability distribution and follow a power law function: there are a small number of hubs (node-concentrators) and a huge number of nodes with a small number of connections (random Erdos-Renyi networks) [7].

One of the functions of a social network is the ability to disseminate information from group to group or from person to group [1, 3, 8, 9]. Depending on the type of information, it is possible to advance the information or block it [10]. The purpose of the pilot study is to determine approaches to assessing the reliability of a social network, that is, the ability of a friend graph to perform its function (transfer information from contact to contact). We consider the resistance of the graph to destruction as a measure of reliability, for how long, and in what ranges of parameters the graph retains its functionality in external attacks that lead to destruction. The indicator of the functional stability of the social graph to destruction can be used to assess the degree of information communication between users, identify the weakest points, the removal of which leads to a violation of the functional stability of the social graph.

2. Materials and Methods

The reliability of the network can be changed by deliberately removing the most significant nodes, a giant connected component or bridges — weak links, in the event of the withdrawal of which the network breaks up into unrelated parts. Weak links structural fill hole, connecting and making integral clusters chain [3]. The most effective removal of bridges, because when even 80% of the nodes of the giant connectivity component are removed, the remaining nodes continue to form a connected cluster [1]. In this regard, in order to assess the reliability of complex networks (social graph), for research experiments, it was decided to carry out a partition into disjoint parts by removing the bridges.

Destroying a social network, we change its topology and state, and therefore the values of network parameters in general and in particular knots. The network under consideration can be represented as a system — set of interrelated elements, each of which is connected directly or indirectly with every other element, any two subsets of this set cannot be independent. The state of the system at any point in time is a set of essential properties that the system possesses at this moment. Based on this, we can define the network as a system that can have many states:

\[
S = \{S_0, S_1, \ldots, S_n\},
\]

where \(S\) — is the set of states of the system, \(S_0, S_1, \ldots, S_n\) — is the state of the system at time \(t\), where \(t = 0, 1, \ldots, n\).

It is logical to assume that the network will have an initial state of \(S_0\), when no action has yet been taken to destroy the system. Define the action \(D_j\) — an action to remove a specific node, where \(j\) — is the node selected for deletion. The action can be effective or not effective. The efficiency \(E(D_j)\) in our case will be the result of destruction, obtained at certain costs (node removal). The effectiveness of the action for each node will be different and will depend on the structure of the network and the elements it contains.

A set of properties of the system reflecting its state will be the vector of social graph parameters:

\[
X = \{x_0, x_1, \ldots, x_k\},
\]

Each state of the system \(S_t\) will correspond to a vector of parameters \(X\). For each subsequent removal of the element (node), the network state and its parameters will change. Therefore, it is expedient to calculate the selected characteristics of the graph after each node removal. Each
subsequent state of the system $S_t$ must be compared with the previous state $S_{t-1}$ to observe the dynamics of the damage and to evaluate the effectiveness of the perfect action.

Initially, seven characteristics of the social graph were chosen: average node degree, average geodesic distance, diameter, density, fragmentation, centrality of mediation and width. Further on we decided to define the parameters for finding the node that should be deleted, and the parameters that reflect the state of the network when it is partitioned. When selecting a vertex for deletion, characteristics such as node load (centrality by mediation) and the average node degree were evaluated. Thus, the Girvana-Newman algorithm [9] was used, in which the edges or nodes with the highest centrality index by intermediation are successively removed. However, taking into account only one node load, there is a risk of destruction of the giant connectivity component. Earlier it was shown [1] that the giant component of connectivity is not inexpedient and ineffective, since it requires high costs with a low result (low efficiency). The removal of weak connections — bridges and drivers — will be a great efficiency, for which it was decided to take into account the average degree of the node. As a result, the node to be deleted must have a high centrality in mediation and should not belong to the connectivity component, that is, it must have a small degree of connections.

In this regard, the following parameters were chosen, in our opinion, reflecting the stability of the social graph:
- mediation centrality (node load) — the number of shortest paths from each vertex to the remaining vertices passing through a particular node;
- the degree of the node — the number of links that the node has;
- density is the ratio of the number of available edges of a graph to the maximum possible number of edges of a given graph;
- width — the average distance between all pairs of nodes of the network for which there is a path of transition from one to another, taking into account the missing values;
- fragmentation — displays pairs of nodes that cannot be linked to each other.

The first two of the above characteristics are necessary when selecting a node for deletion. The object to be deleted must have a large load and a small node. A node that has such parameters will be a bridge connecting the network into a single whole. After each successive removal of the bridge, the density, width and fragmentation of the network are calculated to determine the threshold for the stability indicator, below which the network is still functioning. If the threshold is exceeded, the network will be completely destroyed and information from one end of the network to the other cannot be transmitted. To determine the functional stability of the social network of contacts, a single indicator of the loss of stability of the friendly relations graph was formed: $\text{Stab (stability):}$

$$\text{Stab} = d + f + w,$$

where $d$ (density) is the density, $f$ (fragmentation) is the fragmentation, $w$ (width) is the network width obtained after the destruction action was performed. This additive indicator was evaluated during the experiments at each of the stages of network destruction (the range of the indicator values varies in the interval $[0; 2]$, the larger the value, the less the network stability). For convenience in the representation and reduction of values to the interval $[0; 1]$, the loss of stability of the graph of friendly relations was reduced to the integral coefficient of loss of stability $K$ in the state of the network $i$:

$$K^i = \frac{\text{Stab}_i}{\text{Stab}_{\text{max}} - \text{Stab}_{\text{min}}}, \quad \text{where } \text{Stab}_{\text{max}} = 2, \text{ Stab}_{\text{min}} = 0. \quad (4)$$

Data for the experiments were obtained from the social network "VKontakte" using a developed application in Python 3.4. To obtain data, authorization and access to user
information were realized using the VKontakte API — (built-in API for creating applications). Calculation of the characteristics of graphs and their visualization was carried out using the software tools NetDraw and UCINET.

The experiments were conducted on six social networks: four graphs of friendships between users of the social network VKontakte and two complex networks from the UCINET database. The tested networks contained 50, 63, 82, 100, 108 and 150 nodes.

For each network state, the selected integrity parameters of the graph were calculated: density, fragmentation, width; an index of loss of stability, an integral coefficient of loss of stability, as well as additional parameters: diameter, average node degree and average geodetic distance. After each node removal, in connection with the change in the configuration of the graph, the parameters and the reliability factor were recalculated. The calculations were carried out to verify the degree of reflection by the parameters of the dynamics of the destruction of the graph.

Consider the process of destroying the social network as the example of the graph of friendships on the VKontakte network. The investigated network is an ego-network, contains 100 nodes (users) and 891 edges (links) and is shown in figure 1.

The visual display of the social network of contacts makes it possible to note the presence of a strongly connected component and several clusters without using methods for analyzing network parameters. In its initial state $S_0$, the system has the indicators shown in table 1.

![Figure 1. Investigated Social Network of Contacts before Destruction.](image)

As can be seen from table 1, the network has a fragmentation equal to zero before performing any destructive action, in the initial state. The diameter and average geodetic distance have small values, which indicates a high network connectivity and rapid dissemination of information from one node to another. With the data obtained, all subsequent results will be compared to determine the dynamics of network disruption.
Table 1. Values of the Network Parameters for the State $S_0$.

| Parameters                        | Values |
|-----------------------------------|--------|
| density ($d$)                     | 0.09   |
| fragmentation ($f$)               | 0      |
| width ($w$)                       | 0.46   |
| diameter                          | 3      |
| average geodetic distance         | 1.92   |
| average node degree               | 9      |
| index of loss of stability ($Stab$)| 0.55   |
| integral coefficient of loss of stability ($K$) | 0.275 |

The first node to be deleted is node N 1 — the user whose friend list was downloaded from the social network. It is he who has the greatest load and through it all the shortest paths pass. After the action $D_1$ (removal of the node N 1), the state of the network $S_1$ has changed, and the parameters of the graph have also changed.

It is important that even after the first destroying action, isolated elements were formed, as well as a clique whose nodes were connected to other nodes only by remote node N 1. A comparison of the obtained characteristics of the state of the network $S_1$ with the initial values of $S_0$ gives a graphic representation of the change in the structure and connectivity of the network (see table 2). The decrease in density is due to the decrease in the number of connections in the graph. A sharp increase in diameter can be explained by the removal of a node that has a high centrality in mediation, which has led to an increase in the path of information (message, signal).

The expediency and efficiency of the removal of node N 1 was visible in the graphical display of the network, but in the state $S_1$ it becomes not so easy to determine the next node to be removed. Having estimated the load of all nodes (Network $>$ Centrality $>$ Betweenness $>$ Nodes) and excluding nodes with a large number of links, we get node N 11.

In figure 3 shows the load values of the nodes, where the values of interest will be presented in the first column of Betweenness. After the action of $D_{11}$ (removal of the 11th node), the network enters the state $S_2$ (table 2).

Since the network has several connections between subgroups, according to figure 2, the first actions to remove the nodes will not be highly effective separately. It is necessary to make a number of iterations to remove all weak links, information bridges between blocks. With each removal of the joint, the other node will receive a high load and become a priority for removal.

Using these rules to select the necessary vertices for deletion, we will successively delete the nodes numbered 7, 8, 3, 67, having received the network $S_6$.

As can be seen in figure 4, after the action of $D_{67}$, the network ceased to be connected in comparison with its initial state. To understand the dynamics of changes in the state of the graph, we will analyze the value of the obtained parameters after the perfect destruction actions in accordance with table 2.

Chaotic increase and decrease in the values of the parameters "diameter" and "average geodetic distance" can be explained by the loss of the graph of its subgroups (cliques) or vertices that turned out to be isolated during the removal of the node. The above characteristics are calculated for the connected component, and not for the network as a whole, taking into account the lost blocks. In figure 4 depicts a network, the connected component of which, performs its function to transfer a message from one boundary node of the graph to another. However, the
lack of compactness leads to a large distance, which information must overcome in order to reach the addressee, which reflects the value of the diameter.

The slow drop in density is explained by the fact that when removing selected nodes, the network loses a small number of connections. Initially, it was supposed to take into account the percentage of lost links as the effectiveness of the network destruction action, however, to lose a large number of links, it is necessary to remove the node belonging to the giant connected component.
Figure 4. Network Structure after Removing Nodes 7, 8, 3, 67.

Table 2. Values of the Network Parameters after the Action $D_{67}$.

| Parameters                        | $D_0$ | $D_1$ | $D_{11}$ | $D_7$ | $D_8$ | $D_3$ | $D_{67}$ |
|----------------------------------|-------|-------|----------|-------|-------|-------|----------|
| density ($d$)                    | 0.09  | 0.07  | 0.07     | 0.07  | 0.07  | 0.07  | 0.06     |
| fragmentation ($f$)              | 0     | 0.46  | 0.46     | 0.48  | 0.59  | 0.65  | 0.66     |
| width ($w$)                      | 0.46  | 0.78  | 0.78     | 0.79  | 0.82  | 0.83  | 0.84     |
| diameter                         | 3     | 9     | 9        | 10    | 8     | 8     | 10       |
| average geodetic distance        | 1.92  | 3.18  | 3.33     | 3.43  | 3.07  | 2.85  | 3.09     |
| average node degree              | 9     | 7     | 7        | 7     | 7     | 7     | 6        |
| index of loss of stability ($Stab$) | 0.55  | 1.30  | 1.32     | 1.34  | 1.47  | 1.55  | 1.56     |
| integral coefficient of loss of stability ($K$) | 0.275 | 0.65  | 0.66     | 0.67  | 0.732 | 0.775 | 0.78     |

In the course of the experiments tests were carried out to destroy the giant connected component — the nodes with the largest number of connections (hubs $D_h$) were removed. The test results presented in table 3, showed that with a significant decrease in the density and increase in width, the fragmentation does not reach its maximum value given by us, and the coefficient of stability loss $Stab = 1.38$. Since the connection between subgroups of the network is not violated, there is no removal of bridges — nodes that provide a single connection between other users or clusters (groups of users), the network continues to perform its functions of information transfer, has connectivity and small fragmentation. To destroy a network by
removing nodes in the connectivity component, it is necessary to perform many actions, which determines the process of destruction itself as long and inefficient.

Table 3. The Values of the Network Parameters after Removing the Hubs $D_h$.

| Parameters                          | Action   |
|-------------------------------------|----------|
|                                     | $D_0$   | $D_h$   |
| density ($d$)                       | 0.09     | 0.05    |
| fragmentation ($f$)                 | 0        | 0.51    |
| width ($w$)                         | 0.46     | 0.82    |
| diameter                            | 3        | 9       |
| average geodetic distance           | 1.92     | 3.58    |
| average node degree                 | 9        | 5       |
| index of loss of stability ($Stab$) | 0.55     | 1.38    |
| integral coefficient of loss of stability ($K$) | 0.275 | 0.69 |

Continuing the consecutive removal of bridges, namely, by blocking the nodes 19, 83 and 39, we get an unconnected destroyed network (figure 5). After removing the last vertex number 39, the network becomes divided into separate, unconnected components, and completely loses its functionality. This did not require the destruction of the giant connectivity component, and the loss of nodes was only 8 % of the original value.

Figure 5. Completely Destroyed Network of Contacts.
It was possible to completely destroy the network by removing the 86 or 49 node and break the last link between the two clicks, but taking into account the disconnection and isolation of the giant connected component from all other nodes, the resulting network partition was quite enough. Based on the obtained data, it can be concluded that fragmentation and width with each subsequent removal of the node increase, and the density and the average degree of the node decrease. However, only with values (see table 4 and figure 6) of the above parameters, the coefficient of loss of stability increases. To destroy the network, it is necessary not only to reduce the density of the graph, but also fragmentation, fragmentation of the entire network as a whole — the presence of unconnected elements.

| Parameters                        | Action          |
|-----------------------------------|-----------------|
| density (d)                       |                |
| fragmentation (f)                 |                |
| width (w)                         |                |
| diameter                          |                |
| average geodetic distance         |                |
| average node degree               |                |
| index of loss of stability (Stab) |                |
| integral coefficient of loss of stability (K) |        |
| $D_0$                             | 0.09 0.07 0.07 0.07 0.07 0.07 0.06 0.06 0.06 0.06 |
| $D_1$                             | 0.07 0.07 0.07 0.07 0.07 0.07 0.06 0.06 0.06 0.06 |
| $D_{11}$                          | 0.07 0.07 0.07 0.07 0.07 0.07 0.06 0.06 0.06 0.06 |
| $D_7$                             | 0.07 0.07 0.07 0.07 0.07 0.07 0.06 0.06 0.06 0.06 |
| $D_8$                             | 0.07 0.07 0.07 0.07 0.07 0.07 0.06 0.06 0.06 0.06 |
| $D_3$                             | 0.07 0.07 0.07 0.07 0.07 0.07 0.06 0.06 0.06 0.06 |
| $D_{67}$                          | 0.07 0.07 0.07 0.07 0.07 0.07 0.06 0.06 0.06 0.06 |
| $D_{19}$                          | 0.07 0.07 0.07 0.07 0.07 0.07 0.06 0.06 0.06 0.06 |
| $D_{83}$                          | 0.07 0.07 0.07 0.07 0.07 0.07 0.06 0.06 0.06 0.06 |
| $D_{39}$                          | 0.07 0.07 0.07 0.07 0.07 0.07 0.06 0.06 0.06 0.06 |

**Figure 6.** Change of Density, Width and Fragmentation Values for each State of the System; where on the abscissa axis: 0 — is the initial state of the system ($S_0$), 9 — ($S_9$) the final state of the network at the stages of destruction.

When other networks were destroyed, similar results were obtained (see figure 7). Testing was performed on networks containing 50, 63, 82, 108 and 150 nodes. Initially, it was planned
to include the number of iterations (actions to destroy the graph) in the additive indicator of
stability loss as a possible characteristic of the efficiency of the destruction process. But then,
during the testing it was revealed that the number of stages of destruction does not depend on
the size of the network, but depends on the structure and complexity of the network. A network
with many connections between subgroups will be more difficult to destroy than networks with
fewer connections between components. In the test, the number of iterations varied from 1 to
20. There are networks for which several actions are sufficient to achieve a complete loss of
functional stability by the network.

Using the example of changing fragmentation values for all social networks studied, we
concluded that, regardless of the number of stages of destruction, fragmentation will change
in the same way. The only exception is Network N 6 with 50 nodes, since the network initially
had a rather strongly fragmented structure.

When the network N 3 was destroyed with the number of nodes equal to 108, the residual
strong coupling component was initially obtained, which was not supposed to be further
destroyed, but at this stage of failure, the network loss indicator took the value $Stab = 1.61$
(Integral loss of network stability $K = 0.805$), which indicates the incomplete destruction of
the graph. Indeed, with a detailed examination of figure 8 it becomes clear that the destruction can
continue inside the giant connected component.

Having done the necessary actions to break the integrity of the graph, we obtained the value
of the stability loss indicator of the network $Stab = 1.75$ (integral loss of network stability
$K = 0.875$). Figure 9 clearly shows the final partition of the graph into disconnected elements
and a decrease in the giant connectivity component.

3. Results and Discussion
The integral coefficient of loss of functional reliability of the social network proposed in this
paper helps to assess how much the network is resistant to destruction, how long and to what
extent the changing parameters of the network retains its functionality under external influences.

In the course of the tests, the hypothesis was confirmed that the fragmentation and width
grow, as well as the density drops with each subsequent stage of network destruction. Hypotheses
about the interconnectedness of the average degree of the node and the diameter and structure of
the web graph were not confirmed, since the computation of these characteristics is performed not
for the entire network, but for the components remaining after the subdivision into subgraphs.
The diameter of the network and the average geodetic increased until all the components of the
network were interconnected and formed a single whole. In case of a cluster or click loss, these parameters were calculated by the program for a giant connected component or for a component of the largest size. It is worth emphasizing that the average geodetic distance does not exceed 4 in any network state, which indicates a strong connectivity and stability of social networks in general.

The values of the integral coefficient of stability of the analyzed social networks were determined experimentally. For the initial state of the system, the estimates ranged around $K = 0.2$ ($Stab = 0.55$). For the state of the networks at which they ceased to function the value $K = 0.85 \pm 0.01$ ($Stab = 1.7 \pm 0.02$).

The following dependencies or rules were also formed:

- If the density value of the graph is $d \geq 0.9$, the network is not destroyed and is in the initial state $S_0$;
- if the density value $d \leq 0.05$, then the probability of complete destruction of the network is high;
- If the fragmentation value is $f = 0$, then the network is in the state before the destruction of $S_0$;
- If the fragmentation value $f \geq 0.8$, the network is considered completely destroyed;
- If the network width is $w = 0$, the network is considered to be whole;
- For a width $w \geq 0.8$, we can speak of the termination of the graph as a system.

4. Conclusion
In the article the approach to definition the stability of functioning of social networks and its resistance to structural changes is offered. The paradigm of reliability is as following.
Figure 9. The Final Destruction of the Network N 3.

1. Social networks are quite resistant to external attacks.
2. To effectively destroy the graph, you need a targeted removal of nodes.
3. Removing weak links, articulations in a graph is much more effective than destroying a giant connected component.
4. The selection of the node necessary for removal is justified by the values of the average degree of the vertex and centrality by mediation.
5. The number of stages in violation of network reliability does not depend on the size of the graph, but depends on the structure.
6. Parameters such as fragmentation, density and width clearly demonstrate the state of the entire network at each of the stages of the partition.

References
[1] Evin I A 2010 Komp’juternye issledovaniya i modelirovanie: Uchrezhdenie Rossijskoj akademii nauk Institut mashinovedenija im. A. A. Blagonravova RAN (in russ.) 2 121–141
[2] Snarskij A A and Landje D V 2015 Modelirovanie slozhnyh setej (in russ.) (Kiev: NTUU "KPI"
[3] Granovetter M S 1973 The American Journal of Sociologys 78 1360–1380
[4] Barabási A L and Albert R 1999 Science 286 509–512
[5] Albert R and et al 2000 Nature 406 378–382
[6] Watts D J and Strogatz S H 1998 Nature 393 440–442
[7] Baronchelli A, i Cancho R F, Pastor-Satorras R, Chater N and Christiansen M H 2013 Trends in Cognitive Science 17 348–360
[8] Fortunato S 2010 Physics Reports 486 75–174
[9] Girvan M and Newman M E 2002 Proceedings of the National Academy of Sciences 99 7821–7826
[10] Gubanov D A, Novikov D A and Chchartishvili A G 2010 Social’nye seti modeli informacionnogo vlijanija, upravlenie i protivoborstvo (in russ.) (Moscow: Izdatel’stvo fizikomatematicheskoj literatury)