

Dependence of the reliable graph failure probability on the topological indicators of its elements

I V Pronin

Omsk State Transport University, Omsk, Russia

Abstract. An analysis is made of the dependence of the probability of network failure, presented in the form of reliability graphs, on the topological indicators of its elements, in particular, the variety of centrality proposed in the article in mediation. In connection with the study of the reliability of large networks (networks containing millions of elements), it became necessary to find critical groups of elements that have the greatest impact on the reliability of the system. The article solves the problem of quickly finding such elements without laborious analytical calculation. Monte Carlo analytical and statistical methods are used as methods in the article. As a result of the study, a method is obtained that allows one to find such groups of elements with a certain accuracy. As a confirmation of the great influence on the reliability of the system of the found group of elements, a repeated calculation of reliability was made with an increased probability of failure of the most influential nodes. The results obtained can effectively distribute resources to the most significant system nodes, and, accordingly, increase the reliability of these systems. The absence of a laborious calculation in the method makes it effective when working with large networks.

1. Introduction

Highly reliable systems, that is, systems expressed as a graph consisting of elements, each of which also has a low probability of failure, are the object of study in a scientific article. Functional edges between these elements usually form an irreducible graph, the overall reliability of which is rather difficult to find using analytical calculation, especially if the investigated graph is irreducible or multipolar [1].

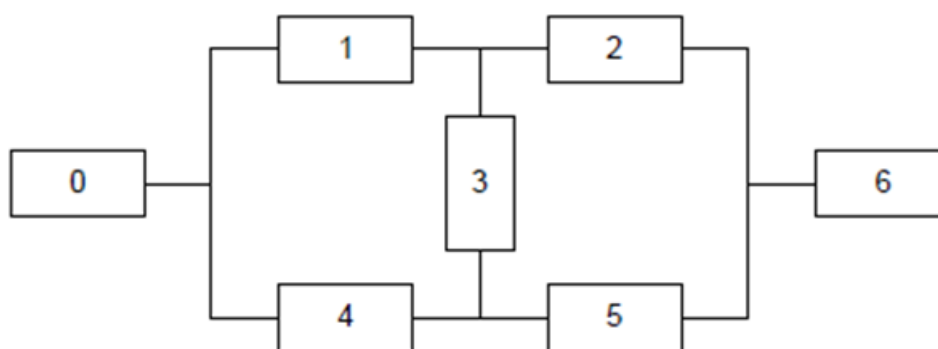


Figure 1. Example of the irreducible graph with poles 0 and 6

One example of such systems is telecommunication networks (including the Internet), consisting of many nodes connected to each other. To increase reliability, it is possible to use a certain resource S , the distribution of which between different groups of elements can reduce the probability of system failure:



$$Q = Q(s_1, s_2, \dots, s_n) \rightarrow \min. \quad (1)$$

The resource S can be money, human resources, etc. For the most optimal distribution of the resource S , it is necessary to find critical groups of elements of the system, i.e. elements that have the greatest impact on system reliability. Analytical calculation to find critical groups of elements can also be quite time-consuming.

The article solves the problem of finding a method for detecting elements that have the greatest impact on the reliability of the system without the aid of complex calculation, but on the basis of their topological indicators, such as betweenness centrality.

2. Topological indicators of graph elements

Let us determine what indicators of the reliability graph element can influence the probability of system failure.

Consider the following topological indicators of the elements [2]:

- parameter «degree»,
- parameters «closeness centrality» and «harmonic centrality»,
- parameter «betweenness centrality».

In addition to these values, in article [2] other types of centralities are also given, however, according to the author, their influence on the reliability of the system is insignificant.

2.1 Parameter the degree of the node.

The reliability graphs considered in this article are non-oriented, therefore, all vertex edged are analyzed together, for this, the number of edges that start or end at this node is calculated. Accordingly, the degree is calculated as the sum of the line:

$$\deg(i) = \sum_j a_{ij}. \quad (2)$$

2.2 The parameters closeness centrality and harmonic centrality.

The first parameter C_c (closeness centrality) characterizes the closeness of the node to the rest of the network nodes. As applied to graphs that are not strongly related:

$$C_c(i) = \frac{1}{\sum_{d(i,j) < \infty} d(i,j)}, \quad (3)$$

Where $d(i, j)$ is the length of the path between the vertices i and j . In the case, if vertex j is not reachable from vertex i , then the path length is $d(i, j) = \infty$, therefore, such vertices are excluded from consideration.

For such graphs, it is possible to calculate the C_{HC} parameter harmonic centrality, when instead of the inverse of the sum of the distances, the sum of the inverse distances is taken (assuming that $\infty^{(-1)} = 0$):

$$C_{HC}(i) = \sum_{j \neq i} \frac{1}{d(i,j)} = \sum_{d(i,j) < \infty, j \neq i} \frac{1}{d(i,j)}. \quad (4)$$

2.3 The parameter betweenness centrality.

The betweenness centrality (C_B) parameter is the fraction of the shortest paths between all the network nodes that pass through this node. The index $C_B(v)$ for the vertex v is defined as follows:

$$C_B(v) = \sum_{i \neq v \neq j \in V} \frac{\sigma_{ij}(v)}{\sigma_{ij}}, \quad (5)$$

Where σ_{ij} is the number of shortest paths from vertex i to vertex j of the graph, and $\sigma_{ij}(v)$ is the number of shortest paths from i to j passing through v .

2.4 The relative betweenness centrality

We introduce our own topological indicator of the nodes of the graph — *the relative betweenness centrality*. This value shows the ratio of the number of paths passing between the poles through a certain vertex to the total number of paths between the poles.

$$C'_B(v) = \frac{\sum_{i \neq v \neq j \in V} A_{ij}(v)}{\sum_{i \neq v \neq j \in V} A_{ij}}, \quad (6)$$

where A_{ij} is the number of all paths from vertex i to vertex j of the graph, and $A_{ij}(v)$ is the number of all paths from i to j passing through v .

3. System reliability determination

As mentioned earlier, the reliability graphs considered in the article are irreducible, therefore, an analytical calculation to determine the probability of a system failure will be too time-consuming. Reliability assessment by one of the Monte Carlo methods will be the optimal solution, since, in addition to the accuracy and speed of these methods, it is easy to analyze the influence of certain elements on the reliability of the system (this analysis will be described in detail in paragraph 4 of this article). Among the Monte Carlo methods, we will focus on the combined method of weighing/stratification described in [1], since it combines the best qualities of weighing and stratification methods, namely: accuracy, speed, and the ability to work with systems whose elements have different elements failure probability.

Here is a brief description of this method. Unlike the usual Monte Carlo method, when the state of the system is played N times, and after all the experiments have been completed, the probability of system failure is defined as the ratio of the number of unsuccessful experiments (i.e., experiments during which the system failed) to the total number of experiments N , the essence of the combined method is as follows: first, the weighing method is used, during which the probability of failure of the system elements can be brought to a convenient form (for example, increase it to increase the probability of failure of the entire system or the same for all its components), and then break options outcome experience in layers, then produce the required number of experiments only required layers. For example, when calculating the probability of failure of reliability graphs during the writing of this article, the separation by layers occurred according to the number of elements that failed during the experiment, which made it possible to exclude “useless” layers with the number of failed elements equal to 0 and 1. It is also worth noting that the order of application of the methods of weighing and stratification can be interchanged (i.e., first divide the experiment into layers, and then in these layers bring the probabilities to the desired form).

4. Experiment

To confirm the theory of the influence of a certain group of elements (in this case, the elements through which the greatest number of paths pass) on the reliability of the system, we will analyze Barabashi-Albert graphs [3]. In these graphs, we select special vertices, poles, the probability of failure of which is 0 (that is, when playing an experiment, these vertices cannot fail), among which one pole will be the “main”. We will consider the system workable if, after playing the state of the

system, the “main” pole will have a path to each of the other poles, otherwise the system is considered to be failed.

We have a counter on each vertex. When playing each experiment during the work of the combined method of weighing / stratification, we will save the numbers of the vertices that have failed in this experiment. If the system turns out to be unworkable, the counter for each of these vertices will increase by 1. Thus, it is possible to determine which of the vertices lead to the failure of the system most often.

In addition, we calculate the relative betweenness centrality by each vertex for each pair of poles under consideration, i.e. pairs of “main” and ordinary pole.

The result of the calculation will be a table whose rows will contain the value of *relative betweenness centrality* by the vertex for each of the pairs of poles. By adding a column of counters to this table, we can check whether the *relative betweenness centrality* affects the number of system failures.

To study the dependence of the reliability of the system on its topological indicators, the following programs were used:

1) Software implementation of the combined method for calculating the probability of failure, taking into account which elements led to the failure of the system as a whole.

2) A program that counts the number of paths between two vertices passing through each vertex. Java was used as a programming language. Input data for programs is indicated as the number of vertices and edges, as well as pairs of vertices connected by edges.

To establish the dependence of the reliability of the system on its topological indicators, several graphs with the number of vertices 50 and 100 were studied. The research results showed, on the whole, the same result for several graphs.

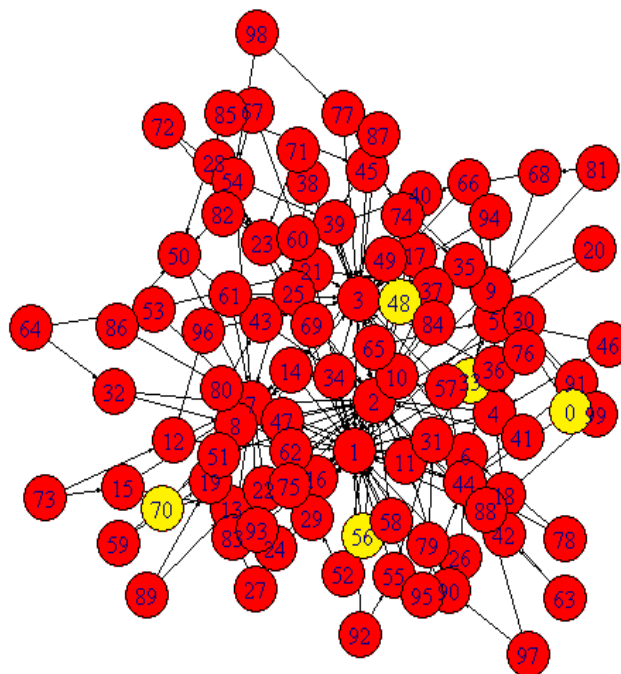


Figure 2. Investigated graph

Figure 1 shows one of the studied graphs. The number of its elements is 100, and the vertices 0 (main pole), 33, 48, 56, 70 are declared as poles. The graphical image of the graph was obtained using the igraph package of software environment R. We put the experiment as follows – we calculate the

number of paths between certain vertices passing through each of the vertices, using the number of paths we find the relative betweenness centrality (it can be easily calculated as the ratio of the number of paths of a particular vertex to the number of paths between the considered pair of poles). After that, we calculate the probability of system failure and find the vertices that most often lead to failure and correlate these lists. The results (for the first 10 most traveled nodes) are summarized in table 1. It is worth noting that not all the vertices, the failure of which led to the exit of the system, turned out to be in the top ten significant vertices, so for further analysis, add them to the end of the table. We will also calculate the total probability of system failure with a given failure probability of each element and summarize this data in table 2.

For understanding, we give a description of the columns of table 1. "Vertex" – the number of the vertices in the studied graph. "Path" – the value of relative betweenness centrality for a particular pair of main and ordinary poles. "Sum" – the sum of the relative betweenness centralities for all three paths, sorting was performed on this column. "Rank" – shows what position the vertex takes when sorting the elements of the graph by the column "Sum". "Failures" – the number of failures of each vertex that led to the failure of the system as a whole, counted when playing states of the system.

Table 1. Ranking of nodes according to the total relative betweenness centrality

Rank	Vertex	Path				Sum	Failures
		0 - 33	0 - 48	0 - 56	0 - 70		
1	1	0,996007	0,994721	0,996966	0,995385	3,983078916	1657
2	2	0,99187	0,992191	0,996473	0,99392	3,974453755	576
3	3	0,992393	0,992448	0,99248	0,992221	3,969541031	43
4	7	0,980011	0,980623	0,979032	0,989027	3,92869284	614
5	9	0,988477	0,958585	0,958875	0,958236	3,864172707	1144
6	44	0,956037	0,888387	0,858668	0,866744	3,569836593	599
7	5	0,845957	0,886909	0,849892	0,85731	3,440067967	44
8	54	0,765603	0,765925	0,75508	0,748938	3,035545345	52
9	6	0,641146	0,580486	0,530296	0,571018	2,322945917	46
10	21	0,508009	0,549968	0,53505	0,519948	2,112974819	49
12	40	0,363421	0,931589	0,360747	0,338102	1,993858133	563
58	22	0,011841	0,013169	0,006561	0,531689	0,563260607	626

Table 2. Calculation of system failure probability

Parameter	Value
Node failure probability p	0,02
System failure probability evaluation Q	$1,963692983119093 \cdot 10^{-3}$
System failure probability evaluation Dispersion D	$8.9494165852299 \cdot 10^{-10}$
Standard deviation $3 \cdot \sigma$	$8.974672655148437 \cdot 10^{-5}$

As can be seen from table 1, most often lead to the failure of the vertex system with maximum total relative betweenness centrality, however, this correlation is not entirely accurate. It is also worth considering separately the vertices 40 and 22, the failure of which also often led to the failure of the system, however, when sorted by the total relative betweenness centrality, they did not fall into the top ten vertices.

Node 40 has a high relative betweenness centrality only in the case of counting the paths between poles 0 and 48, between the other poles the relative betweenness centrality is low.

Node 22 does not have a high relative betweenness centrality when considering any pair of poles. The multiple failures that led to the failure of the entire system can be explained by the proximity of this vertex to one of the poles and to the most connected vertices.

To confirm the great influence of these vertices on the probability of system failure, we re-evaluate the probability of system failure, but for “influential” nodes we reduce the probability of failure by 100 times. Such a calculation using different failure probability for different vertices is possible using the combined method of weighing / stratification. These “influential” nodes were already found in the previous assessment of the probability of system failure, these are nodes 1, 2, 7, 9, 22, 40, 44.

Table 3. Calculation of the failure probability of a system with a reduced probability of failure of “influential” nodes and comparing it with the past result

Parameter	Calculation without a reduced probability of failure of “influential” nodes	Calculation with a reduced probability of failure of “influential” nodes
Node failure propability p	0,02	0,02
“Influential” node failure propability $p_{\text{вн}}$	-	0,0002
System failure propability evaluation Q	$1,963692983119093 \cdot 10^{-3}$	$1.9851481292487396 \cdot 10^{-7}$
System failure propability evaluation Dispersion D	$8.9494165852299 \cdot 10^{-10}$	$5.947484590695101 \cdot 10^{-14}$
Standard deviation $3 \cdot \sigma$	$8.974672655148437 \cdot 10^{-5}$	$7.316239561158171 \cdot 10^{-7}$

As can be seen from table 3, a 100-fold decrease in the probability of failure for seven elements in a graph containing 100 elements resulted in a 10,000-fold decrease in the probability of failure of the entire system. It is also worth noting that despite the reduced probability of failure of “influential” nodes, their failure still led to the failure of the entire system more often than other nodes. New “influential” vertices also did not appear after a change in the probability of failure of the “influential” vertices. This is shown in table 4, in which the graph elements are sorted by their number of failures, as a result of which the system became unworkable (the table is sorted by the number of failures and only the first 10 vertices are shown).

Table 4. Sorting graph elements by the number of failures that caused the system to fail

Vertex	Failure number
1	1312
9	877
7	475
22	475
40	463
44	462
2	454
82	38
28	37
87	35

5. Conclusion

In the course of the investigation, results were obtained that confirm the theory that the *relative betweenness centrality* proposed in this article affects the reliability of the system. Nevertheless, during the experiment, there were cases in which elements that did not have a high value of relative

betweenness centrality had a significant impact on the reliability of the system. It is most likely that, in addition to the proposed relative betweenness centrality, the probability of failure of the reliability graph is less affected by other topological indicators, for example, closeness centrality. It is planned to continue research in this area.

6. References

- [1] Zadorozhniy V N, Udina M N 2016 Accelerated calculation and parametric optimization methods of the highly reliable structurally complex systems *Omsk Scientific Bulletin* vol **4** pp 132-138
- [2] Bredikhin S V, Lyapunov N G, Scherbakova N G, Urgenson A N 2016 Nodes "centrality" parameters of the scientific articles citation network *Problems of computer science* vol **1** pp 39-57
- [3] Zadorozhniy V N, Udin E B 2009 Exact theory of Barabashi-Albert graphs *Omsk Scientific Bulletin* vol **3** pp 13-19
- [4] Luke D A 2015 *A User's Guide to Network Analysis in R* (New York: Springer-Verlag) p 238
- [5] Sedgewick R Wayne K 2011 *Algorithms, 4th edition* (Boston: Pearson Education, Inc.) p 956
- [6] Zadorozhnyi V N, Yudin E B 2016 Dynamic Equations of Node Degrees in Growing Networks with Connection Losses *Dynamics of Systems, Mechanisms and Machines: conference proceeding (Russia Omsk State Technical University)* 7819111–5
- [7] Zadorozhnyi V N 2017 Preferential attachment random graphs with vertices losses *International Siberian Conference on Control and Communications (SIBCON): conference proceeding Astana, Kazakhstan* 7998456–6
- [8] Zadorozhnyi V N, Yudin E B, Yudina M N 2018 Distributions of Degrees in Growing Graphs with Loss of Arcs *Moscow Workshop on Electronic and Networking Technologies: proceeding Moscow* 8337251–7
- [9] E. Manzi, M. Labbe, G. Latouche, F. Maffioli 2001 Fishman's sampling plan for computing network reliability *IEEE Transactions on Reliability* vol **50** pp 41-46
- [10] Pierre L'Ecuyer, Gerardo Rubino, Samira Saggadi, Bruno Tuffin 2011 Approximate Zero-Variance Importance Sampling for Static Network Reliability Estimation *IEEE Transactions on Reliability* vol **60** pp 610-604

Acknowledgments

The author thanks his scientific supervisor, V N Zadorozhniy, for his help in writing this article.