Method of Calculation of Information Protection from Clusterization Ratio in Social Networks

Vitalii Savchenko [1], Volodymyr Akhramovych [2], Oleksander Matsko[3] and Ivan Havryliuk [4]

[1,2]*State University of Telecommunications, Solomianska str.7, Kyiv, 03110, Ukraine*
[3,4]*The National Defense University of Ukraine named after Ivan Cherniakhovskyi, Povitroflotsky av. 28, Kyiv, 03049, Ukraine*

Abstract
The article investigates the dynamic models of the information protection system in social networks taking into account the clustering coefficient, and also analyzes the stability of the protection system. In graph theory, the clustering factor is a measure of the degree to which nodes in a graph tend to group together. The available data suggest that in most real networks, and in particular in social networks, nodes tend to form closely related groups with a relatively high density of connections; this probability is greater than the average probability of a random connection between two nodes.
There are two variants of this term: global and local. The global version was created for a general idea of network clustering, while the local one describes the nesting of individual nodes. There is a practical interest in studying the behavior of the system of protection of social networks from the value of the clustering factor. Dynamic systems of information protection in social networks in the mathematical sense of this term are considered. A dynamic system is understood as any object or process for which the concept of state as a set of some quantities at a given moment of time is unambiguously defined and a given law is described that describes the change (evolution) of the initial state over time. This law allows the initial state to predict the future state of a dynamic system. It is called the law of evolution.
The study is based on the nonlinearity of the social network protection system. To solve the system of nonlinear equations used: the method of exceptions, the joint solution of the corresponding homogeneous characteristic equation. Since the differential of the protection function has a positive value in some data domains (the requirement of Lyapunov's theorem for this domain is not fulfilled), an additional study of the stability of the protection system within the operating parameters is required. Phase portraits of the data protection system in MatLab / Multisim are determined, which indicate the stability of the protection system in the operating range of parameters even at the maximum value of influences.

Keywords
dynamic models, information protection system, social networks, clustering coefficient, nonlinearity, exception method, homogeneous characteristic equation, function differential, system stability, phase portrait

1. Introduction

Descriptions of dynamical systems for various problems depending on the law of evolution are also various: with the help of differential equations, discrete mappings, graph theory, Markov chain theory, and so on. The choice of one of the methods of description determines the

III International Scientific And Practical Conference "Information Security And Information Technologies", September 13–19, 2021, Odesa, Ukraine
EMAIL: savitan@ukr.net; 12z@ukr.net; macko2006@ukr.net; ivan.havryliuk@gmail.com
ORCID: 0000-0002-3014-131X (A.1); 0000-0000-0002-6174-5300 (A. 2); 0000-0003-3415-3358 (A. 3); 0000-0002-3514-0738 (A. 4)

specific form of the mathematical model of the corresponding dynamic system [3].

The mathematical model of a dynamic system is considered to be given if the parameters (coordinates) of the system are introduced, which unambiguously determine its state, and the law of evolution is specified. Depending on the degree of approximation to the same system, different mathematical models can be matched.

Theoretical study of the dynamic behavior of a real object requires the creation of its mathematical model. In many cases, the procedure for developing a model is to compile mathematical equations based on physical laws. Usually these laws are formulated in the language of differential equations. As a result, the coordinates of the state of the system and its parameters are interconnected, which allows us to begin to solve differential equations under different initial conditions and parameters.

2. Related works

In the article [1] the definition of the clustering coefficient in the case of (binary and weighted) directional networks is extended and the expected value for random graphs is calculated. In [2], it is noted that the properties of the small world of neighboring connections are higher than in comparative random networks. If a node has one or no neighbors, in such cases the local clustering is traditionally set to zero, and this value affects the global clustering factor. It is proposed to include the coefficient θ for isolated nodes in order to estimate the clustering coefficient, except in cases from the determination of Watts and Strogats. In [3] a method of determining trust and protection of personal data in social networks was developed. In article [4-6] the clustering coefficients for social networks, including power ones, are considered. In [7], a comparison of different generalizations of the clustering coefficient and local efficiency for weighted undirected graphs is made. In the article [8] the analysis of the clustering coefficient on the social network twitter is carried out. In [9], an analysis of the clustering coefficient through triads of connections was performed. In the article [10] the dependence between the clustering coefficient and the average path length in a social network is investigated. In [11,13] the use of clustering methods of social networks for personalization of educational content is investigated. The article [12,15] discusses the behavior of the clustering coefficient for complex networks. In [14], it was concluded that based on the results of the experiment, it can be concluded that among the clustering algorithms there is no universal algorithm that would be significantly ahead of others on all data sets. The leaders of benchmarking are the algorithms Spinglass and Walktrap. From the considered analysis of the works, it can be concluded that currently the protection of users in social networks is considered primarily as a technical problem that does not take into account the structural parameters of the network and its topological features. This emphasizes the relevance of the topic of work regarding the construction of a protection system based on structural parameters, taking into account network clustering.

3. Formulation of the research task

It is necessary to investigate the dynamic system of information protection in the social network (SN) from the clustering factor. Carry out modeling of a nonlinear protection system taking into account the clustering factor in SN. Investigate the stability of the protection system in the SN.

4. Main part

4.1. Nonlinear solution of the protection system in the SN, taking into account the action of a specific parameter - the clustering factor

Analysis of graphical dependences of a linear system [3] indicates the nonlinearity of the system. Therefore, in the system of equations (1) we introduce nonlinear components (2):

$$\begin{cases} \frac{dI}{dt} = Z_p Z + (C_v + C_K) I \\ \frac{dZ}{dt} = -\left(\frac{\sum_{v \in V} C_{v1}}{N^2}\right) - I(C_{d2} + C_{d1}) \end{cases} \quad (1)$$

where: $\sum_{v \in V} C_{v1}$ - the total number of connections in the network, N - the number of vertices in the network.

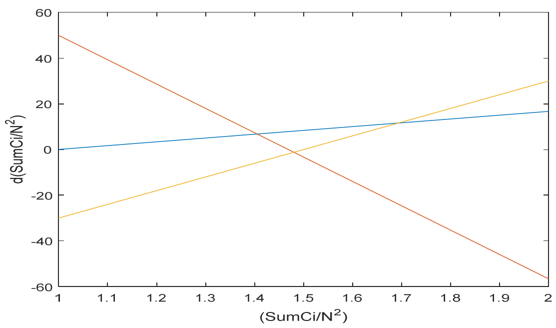


Figure 1: Differential of the clustering coefficient function

$$\begin{cases} \frac{dI}{dt} = Z_p Z + (C_v + C_K) I + L_2(I^2) + L_3(I^3) + \ldots \\ \frac{dZ}{dt} = -\left(\frac{\sum_{v \in V} C_{v1}}{N^2}\right) - I(C_{d2} + C_{d1}) + K_2(Z^2) + K_3(Z^3) + \ldots \end{cases} \quad (2)$$

where: L_2, L_3, etc. K_2, K_3, etc. some linear operators. We consider the nonlinearity of the system to be weak, which allows us to find a solution for each equation of the system (2) by the method of successive approximation, putting:

$$I = I_1 + I_2 + I_3 \ldots$$

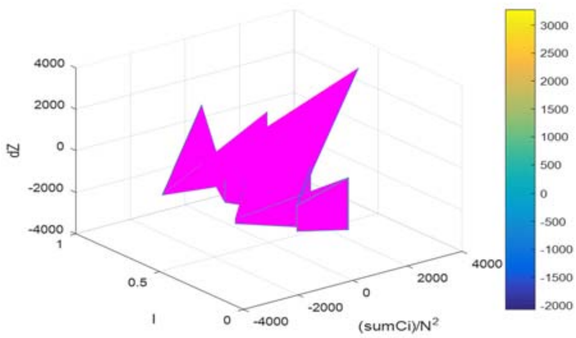


Figure 2: Differential of protection function

Since the differential of the protection function has a positive value in some data domains (the requirement of Lyapunov's theorem for this domain is not fulfilled), an additional study of the stability of the protection system within the operating parameters is required

$$Z = Z_1 + Z_2 + Z_3 + \ldots$$

Let at

$$dI = 0, \ \frac{dI}{dt} = 0, \text{ and } dZ = 0, \ \frac{dZ}{dt} = 0$$

$$I = I_0 \operatorname{Sin} \omega t, Z = Z_0 \operatorname{Sin} \omega t$$

We obtain a system of equations:

$$\begin{cases} \frac{dI}{dt} = Z_p Z + (C_v + C_K) I - L_2(I_0^2 \operatorname{Sin}^2 \omega t) - \\ - L_3(I_0^3 \operatorname{Sin}^3 \omega t) - \ldots \\ \frac{dZ}{dt} = -\left(\frac{\sum_{v \in V} C_{v1}}{N^2}\right) - I(C_{d2} + C_{d1}) - K_2(Z_0^2 \operatorname{Sin}^2 \omega t) - \\ - K_3(Z_0^3 \operatorname{Sin}^3 \omega t) - \ldots \end{cases} \quad (3)$$

Let's rewrite the system and present it as follows:

$$\begin{cases} \frac{dI}{dt} = \alpha Z + \beta_1 I - \sum_{k=2}^{\infty} L_k I_0^k \sin^k \omega t, \\ \frac{dZ}{dt} = \beta_2 I + \gamma - \sum_{k=2}^{\infty} K_k Z_0^k \sin^k \omega t, \end{cases} \quad (4)$$

where:

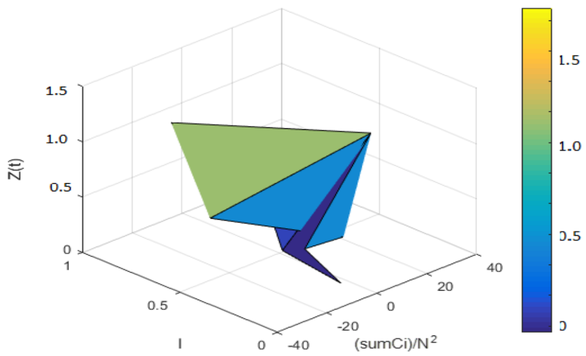
$$\alpha = Z_p, \beta_1 = C_v + C_K, \beta_2 = -\left(C_{d2} + C_{d1}\right), \gamma = -\left(\frac{\sum_{v \in V} C_{v1}}{N^2}\right)$$

Next, use the exception method:

$$\frac{dZ}{dt} = \beta_2 I + \gamma - \sum_{k=2}^{\infty} K_k Z_0^k \sin^k \omega t \Rightarrow$$

$$I = \frac{1}{\beta_2}\left(\frac{dZ}{dt} - \gamma + \sum_{k=2}^{\infty} K_k Z_0^k \sin^k \omega t\right) \Rightarrow$$

$$\frac{dI}{dt} = \frac{1}{\beta_2}\left(\frac{d^2 Z}{dt^2} + \frac{1}{\omega}\sum_{k=2}^{\infty}\left(k K_k Z_0^k \sin^{k-1} \omega t \cos \omega t\right)\right) \quad (5)$$



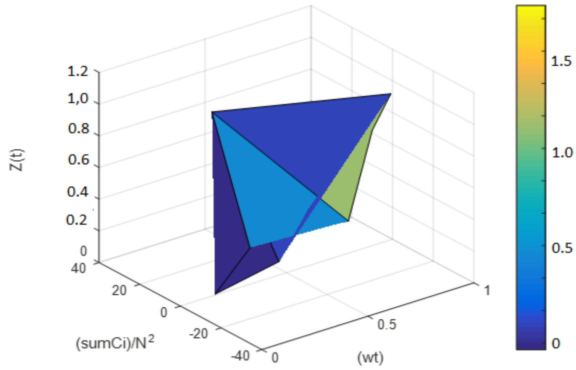


Figure3: Graphs by dependence (4)

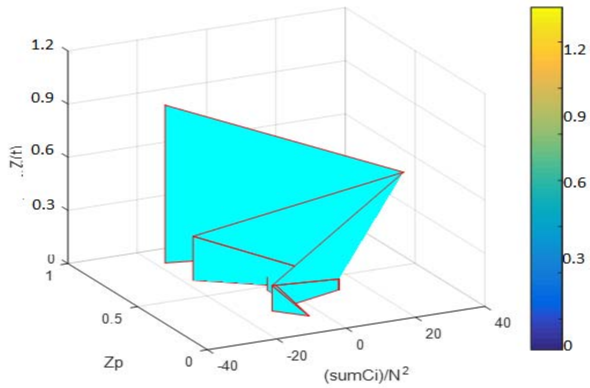


Figure 4: Graphs by dependence (5)

Substitute all the found expressions (5) in the first equation of system (4):

$$= \frac{1}{\beta_2}\left(\frac{d^2Z}{dt^2} + \frac{1}{\omega}\sum_{k=2}^{\infty}\left(kK_k Z_0^k \sin^{k-1}\omega t\cos\omega t\right)\right) = \alpha Z + \frac{\beta_1}{\beta_2}\left(\frac{dZ}{dt} - \gamma + \sum_{k=2}^{\infty} K_k Z_0^k \sin^k \omega t\right) - \sum_{k=2}^{\infty} L_k I_0^k \sin^k \omega t \quad (6)$$

or:

$$\frac{d^2Z}{dt^2} - \beta_1\frac{dZ}{dt} - \alpha\beta_2 Z = -\frac{1}{\omega}\sum_{k=2}^{\infty}\left(kK_k Z_0^k \sin^{k-1}\omega t\cos\omega t\right) - \beta_1\gamma + \beta_1\sum_{k=2}^{\infty} K_k Z_0^k \sin^k \omega t - \beta_2\sum_{k=2}^{\infty} L_k I_0^k \sin^k \omega t \quad (7)$$

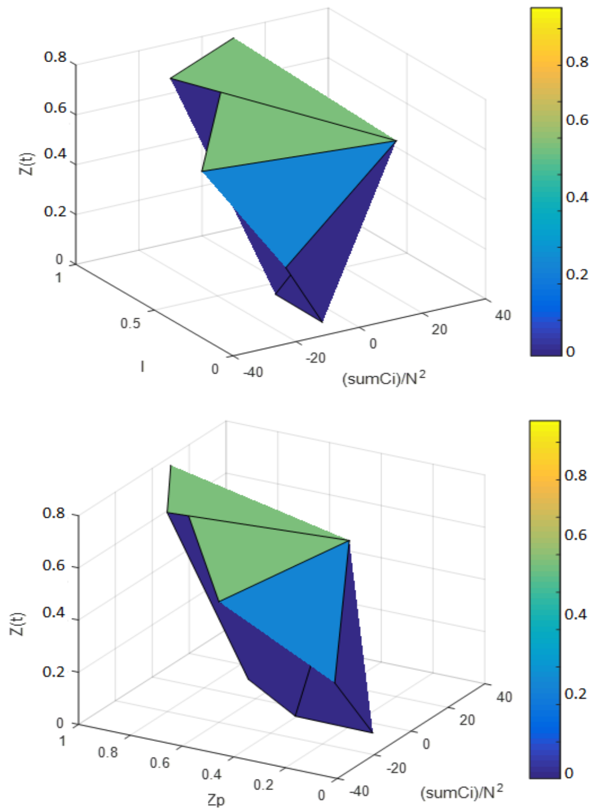
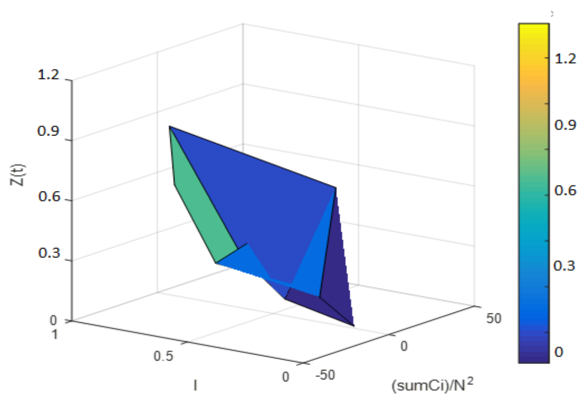


Figure 5: Graphs by dependence (7)



Now we find a common solution of the corresponding homogeneous equation:

$$Z'' - \beta_1 Z' - \alpha\beta_2 Z = 0 \quad (8)$$

The characteristic equation has the form: $\lambda^2 - \beta_1\lambda - \alpha\beta_2 = 0$. Consider the case of the positive discriminant of this equation:

$$D = \beta_1^2 + 4\alpha\beta_2 > 0 \Rightarrow \lambda_{1,2} = \frac{\beta_1 \pm \sqrt{\beta_1^2 + 4\alpha\beta_2}}{2} \quad (9)$$

From:

$$Z_{одн}(t)=c_1e^{\frac{\beta_1+\sqrt{\beta_1^2+4\alpha\beta_2}}{2}t}+c_2e^{\frac{\beta_1-\sqrt{\beta_1^2+4\alpha\beta_2}}{2}t}$$

joint solution of a homogeneous equation.

To find the general solution of the inhomogeneous equation we use the method of variation of arbitrary constants:

$$Z_{одн}(t)=c_1(t)e^{\frac{\beta_1+\sqrt{\beta_1^2+4\alpha\beta_2}}{2}t}+c_2(t)e^{\frac{\beta_1-\sqrt{\beta_1^2+4\alpha\beta_2}}{2}t}$$

where: $c_1'(t),c_2'(t)$ are from the system:

$$\begin{cases} c_1'(t)e^{\frac{\beta_1+\sqrt{\beta_1^2+4\alpha\beta_2}}{2}t}+c_2'(t)e^{\frac{\beta_1-\sqrt{\beta_1^2+4\alpha\beta_2}}{2}t}=0, \\ c_1'(t)\frac{\beta_1+\sqrt{\beta_1^2+4\alpha\beta_2}}{2}e^{\frac{\beta_1+\sqrt{\beta_1^2+4\alpha\beta_2}}{2}t}+ \\ +c_2'(t)\frac{\beta_1-\sqrt{\beta_1^2+4\alpha\beta_2}}{2}e^{\frac{\beta_1-\sqrt{\beta_1^2+4\alpha\beta_2}}{2}t}=N(t), \end{cases}$$

where:

$$N(t)=-\frac{1}{\omega}\sum_{k=2}^{\infty}\left(kK_kZ_0^k\sin^{k-1}\omega t\cos\omega t\right)- \\ -\beta_1\gamma+\beta_1\sum_{k=2}^{\infty}K_kZ_0^k\sin^k\omega t- \qquad (11)$$

From equations (10, 11) we obtain:

$$c_1'(t)e^{\frac{\beta_1+\sqrt{\beta_1^2+4\alpha\beta_2}}{2}t}=-c_2'(t)e^{\frac{\beta_1-\sqrt{\beta_1^2+4\alpha\beta_2}}{2}t}\Rightarrow$$

$$\Rightarrow c_2'(t)e^{\frac{\beta_1-\sqrt{\beta_1^2+4\alpha\beta_2}}{2}t}\left(\frac{\beta_1+\sqrt{\beta_1^2+4\alpha\beta_2}}{2}+ +\frac{\beta_1-\sqrt{\beta_1^2+4\alpha\beta_2}}{2}\right)=N(t) \qquad (12)$$

or:

$$c_2'(t)e^{\frac{\beta_1-\sqrt{\beta_1^2+4\alpha\beta_2}}{2}t}\sqrt{\beta_1^2+4\alpha\beta_2}=-N(t) \qquad (13)$$

where will we get:

$$c_2(t)=\frac{1}{\sqrt{\beta_1^2+4\alpha\beta_2}}\int N(t)e^{\frac{-\beta_1+\sqrt{\beta_1^2+4\alpha\beta_2}}{2}t}dt \qquad (14)$$

$$c_1(t)=\frac{1}{\sqrt{\beta_1^2+4\alpha\beta_2}}\int N(t)e^{\frac{-\beta_1-\sqrt{\beta_1^2+4\alpha\beta_2}}{2}t}dt \qquad (15)$$

Given (13,14,15) we have:

$$Z(t)=\int\Big(N(t)-e^{\frac{-\beta_1-\sqrt{\beta_1^2+4\alpha\beta_2}}{2}t}\frac{e^{\frac{\beta_1+\sqrt{\beta_1^2+4\alpha\beta_2}}{2}t}}{\sqrt{\beta_1^2+4\alpha\beta_2}}\Big)dt- \\ -\int\Big(N(t)-e^{\frac{-\beta_1-\sqrt{\beta_1^2+4\alpha\beta_2}}{2}t}\frac{e^{\frac{\beta_1-\sqrt{\beta_1^2+4\alpha\beta_2}}{2}t}}{\sqrt{t\beta_1^2-4\alpha\beta_2}}\Big)dt, \qquad (16)$$

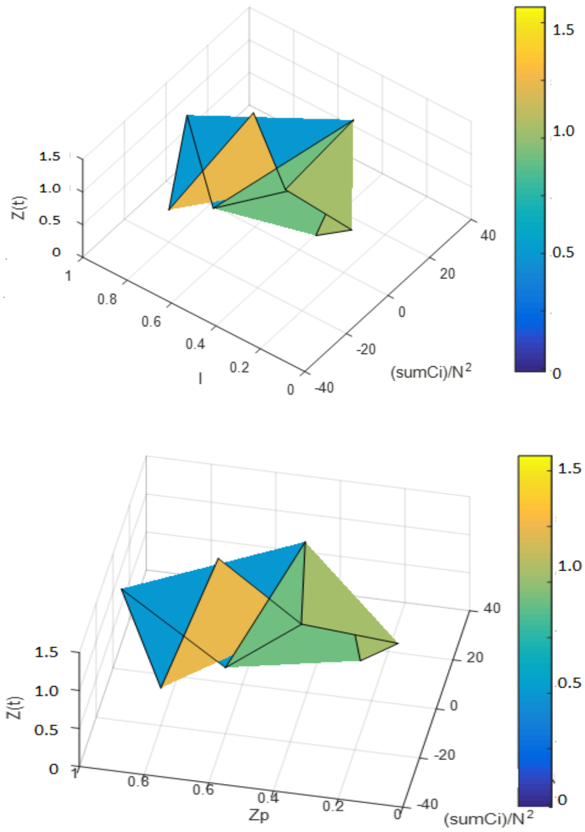


Figure 6: Graphs by dependence (16)

3.2. Define the phase portrait of the data protection system

Initial equation:

$$\frac{d^2Z}{dt^2}-\beta_1\frac{dZ}{dt}-\alpha\beta_2Z=-\frac{1}{\omega}\sum_{k=2}^{\infty}\left(kK_kZ_0^k\sin^{k-1}\omega t\cos\omega t\right)- \\ -\beta_1\gamma+\beta_1\sum_{k=2}^{\infty}K_kZ_0^k\sin^k\omega t-\beta_2\sum_{k=2}^{\infty}L_kI_0^k\sin^k\omega t \qquad (17)$$

The solution will be implemented in the program MatLab / Multisim. Let's make the scheme (Fig. 7).

The phase portrait is presented in the form of an ellipse, which indicates the stability of the personal data protection system.

The results of the program are presented in Fig. 8, 9.

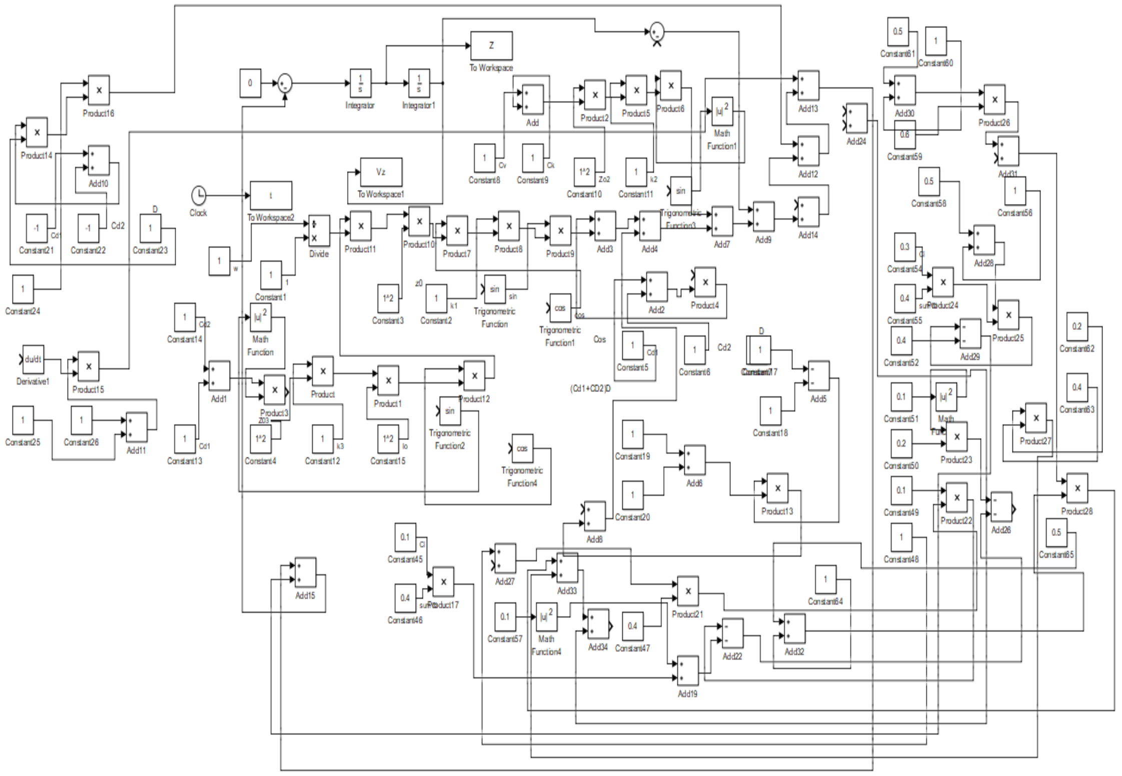


Figure 7: Block diagram of the phase portrait program in the Multisim program, taking into account the attack block

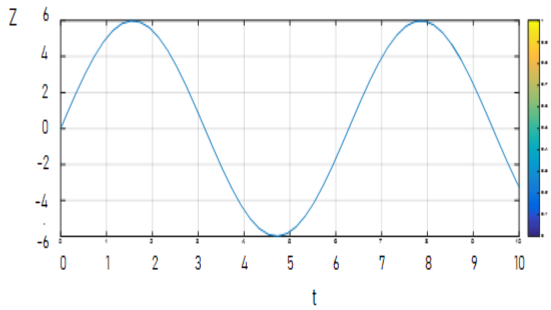


Figure 8: Harmonic oscillations of the protection system on time Z=f(t)

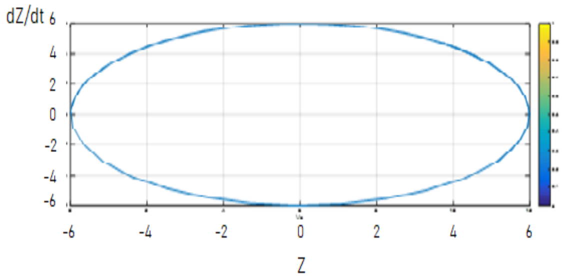


Figure 9: Phase portrait of the protection system on clustering factor

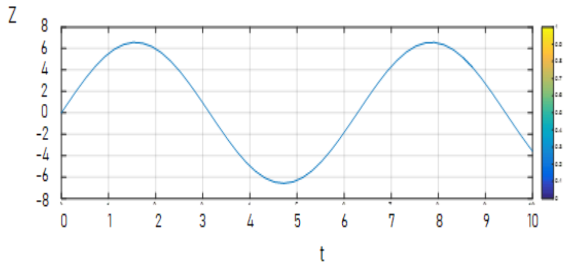


Figure 10: Harmonic oscillations of the protection system on time Z=f(t) taking into account the attacks

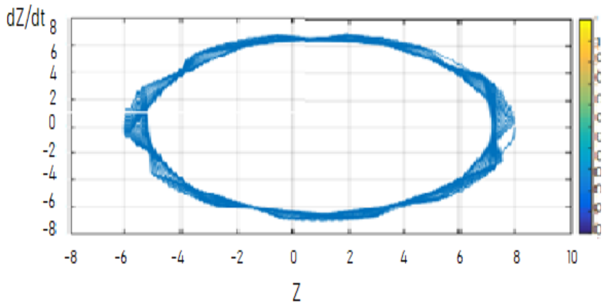


Figure 11: Phase portrait of the protection system on clustering factor taking into account the attacks

5. Analysis of the obtained results

In contrast to previous research by scientists, it has been proven that the SN protection system is stable even from external maximum influences and a specific parameter of the clustering coefficient in the operating range of parameters.

6. Conclusions

For the first time in the article the dynamic model of the information protection system in social networks is investigated taking into account the clustering coefficient, and also the analysis of the stability of the protection system is carried out. A nonlinear equation of information protection is obtained. It is shown that the protection index changes depending on the clustering coefficient. Phase portraits of the protection system are obtained, which indicate the resistance of the system to external influences and the clustering coefficient in SN.

7. References

[1] Giorgio Fagiolo. Clustering in complex directed networks. Phys Rev E Stat Nonlin Soft Matter Phys. . 2007 Aug;76(2 Pt 2):026107. doi: 10.1103/PhysRevE.76.026107. Epub 2007 Aug 16.

[2] Marcus Kaiser (2008). Mean clustering coefficients: the role of isolated nodes and leafs on clustering measures for small-world networks. New Journal of Physics 10 (8): 083042. Bibcode:2008NJPh...10h3042K. arXiv:0802.2512. doi:10.1088/1367-2630/10/8/083042.

[3] O. Laptiev, V. Savchenko, A. Kotenko, V. Akhramovych, V. Samosyuk, G. Shuklin, A. Biehun. Method of Determining Trust and Protection of Personal Data in Social Networks. International Journal of Communication Networks and Information Security (IJCNIS) 2021. №. 1, April 2021. Pp. 15-21.

[4] V. Savchenko, V. Akhramovych, A. Tushych, I. Sribna, I. Vlasov. Analysis of Social Network Parameters and the Likelihood of its Construction. International Journal of Emerging Trends in Engineering Research (IJETER). ISSN: 2347 – 3983, Vol. 8. No. 2, February 2020. pp. 271 – 276.

[5] P. Shchypanskyi, V. Savchenko, V. Akhramovych, T. Muzshanova, S. Lehominova, V. Chegrenets. The Model of Secure Social Networks Activity Based on Graph Theory. International Journal of Innovative Technology and Exploring Engineering (IJITEE). ISSN: 2278–3075, Vol. 9. Issue 4, February 2020, pp. 1803 – 1810.

[6] Serhii Yevseiev, Oleksandr Laptiev, Sergii Lazarenko, Anna Korchenko, Iryna Manzhul. Modeling the protection of personal data from trust and the amount of information on social networks. Number 1 (2021), «EUREKA: Physics and Engineering» pp.24–31. DOI:10.21303/2461-4262.2021.001615

[7] Yui Van, Eshvar Humare, Rik Vandenberhe, Patrik Diupon (2017). Comparison of various generalizations of clustering coefficient and local efficiency for weighted undirected graphs. Neural computing. 29 (2): 313–331. doi : 10.1162 / NECO_a_00914 . Retrieved 8 August 2020.

[8] M. Gracheva, Y. Iakobi, V. Stepanenko, Y. Luneva. Analysis of the characteristics of social graphs built according to the data of the social network Twitter. gramar_98.98@mail.ru.

[9] I. Evin. Introduction to the theory of complex networks. Computer research and modeling. 2010 T. 2 № 2 P. 121–141.

[10] S. Yevseiev, V. Ponomarenko, O. Laptiev, O. Milov and others. Synergy of building cybersecurity systems: monograph / Edited by– Kharkiv: PC TECHNOLOGY CENTER, 2021. – 188 p.

[11] H. Mamedova, F. Agaev, L. Zeinalova. Using social media to personalize e-learning.

İnformasiya texnologiyaları problemləri, 2019, №1, 27–34.

[12] J. Pavlov, I. Chepliukova. On the asymptotics of the power structure of configuration graphs with constraints on the number of edges. Discrete Math, 2018. T. 30. Issue. 1. P. 77–94.

[13] V. Starodubtsev. Personalization of the virtual educational environment. Pedagogical Education in Russia, 2015, №7, p.24–29.

[14] M. Firsov. Benchmarking graph clustering algorithms for decision-making problems. Basic research. – 2017. – № 12 (part 1) – C. 138-142.

[15] J. Pavlov. On the asymptotics of the cluster coefficient of a configuration graph with an unknown distribution of vertex degrees. Information and its application, 2019, T 13, issue 3, 9–13.