Models of the threat of virus idea dissemination in information-telecommunication networks

Dmitriy Moiseev [1] and Vera Miryanova [1]

[1] *Sevastopol state university, 33 Universitetskaya str., Sevastopol, 299053, Russia*

Abstract
The work is devoted to solving the problems associated with the spread of viruses and ideas in the field of research on the processes of reintegration of post-conflict societies. As you know, information and telecommunication networks currently include all kinds of means of switching subscribers, the most common and popular are social networks, which in fact provide an almost complete set of opportunities for exchanging multimedia information between users. The virus must be aware that some «information networks" and "are instantly perceived by active participants in social networks». Effective protection of subscribers from the threat of spreading the virus idea is a serious problem, especially for the development and reintegration of post-conflict societies, since the modern Internet provides not only mobilization and technological opportunities, but also has an informational and psychological impact on individual and mass consciousness. The virus idea, in turn, should be understood as a certain information message, which is often "thrown in" by the media and instantly picked up by the active part of social network subscribers.

Keywords
resources unmanned vehicles, detection of vulnerabilities, information criterion, statistical distance

1. Introduction

As you know, information and telecommunications networks, at the moment, include all kinds of means of switching subscribers, the most common and popular are social networks, which actually provide almost a complete set of opportunities for exchanging multimedia information between users (subscribers). In the second decade of the XXI century, there are a huge number of different social networks in the world, from which you can distinguish some of their prevalence depending on the country or region, so in the Russian Federation today the most popular are «vk.com», «ok.ru», «my.mail.ru», in USA there are «Facebook», «MySpace», «Twitter» and «LinkedIn»; «Nexopia» — in Canada, «Bebo» — in Great Britain, «Facebook», «dol2day» — in Germany. The current problem of such networks, as you know, is their low level of information security.

The virus idea, in turn, should be understood as a certain information message, which is often "thrown in" by the media and instantly picked up by the active part of social network subscribers.

When analyzing the information flows of the socio-media environment, it is necessary to base on the network theories set forth in the works of M. Granovetter, M. Castels, P. Lazarsfeld, J. Moreno, as well as the ideas of E. Toffler about information and communication technologies as a factor of socio-political changes. It becomes possible to use modern information technologies of agent - based modeling and big data processing to study the processes of deliberate organization of collective actions of a socio-political nature that form value-semantic guidelines and attitudes of Internet users.

When considering issues related to modeling processes occurring in information and telecommunications networks, the main approach is to use models of influence, information management and confrontation (see Figure 1:).
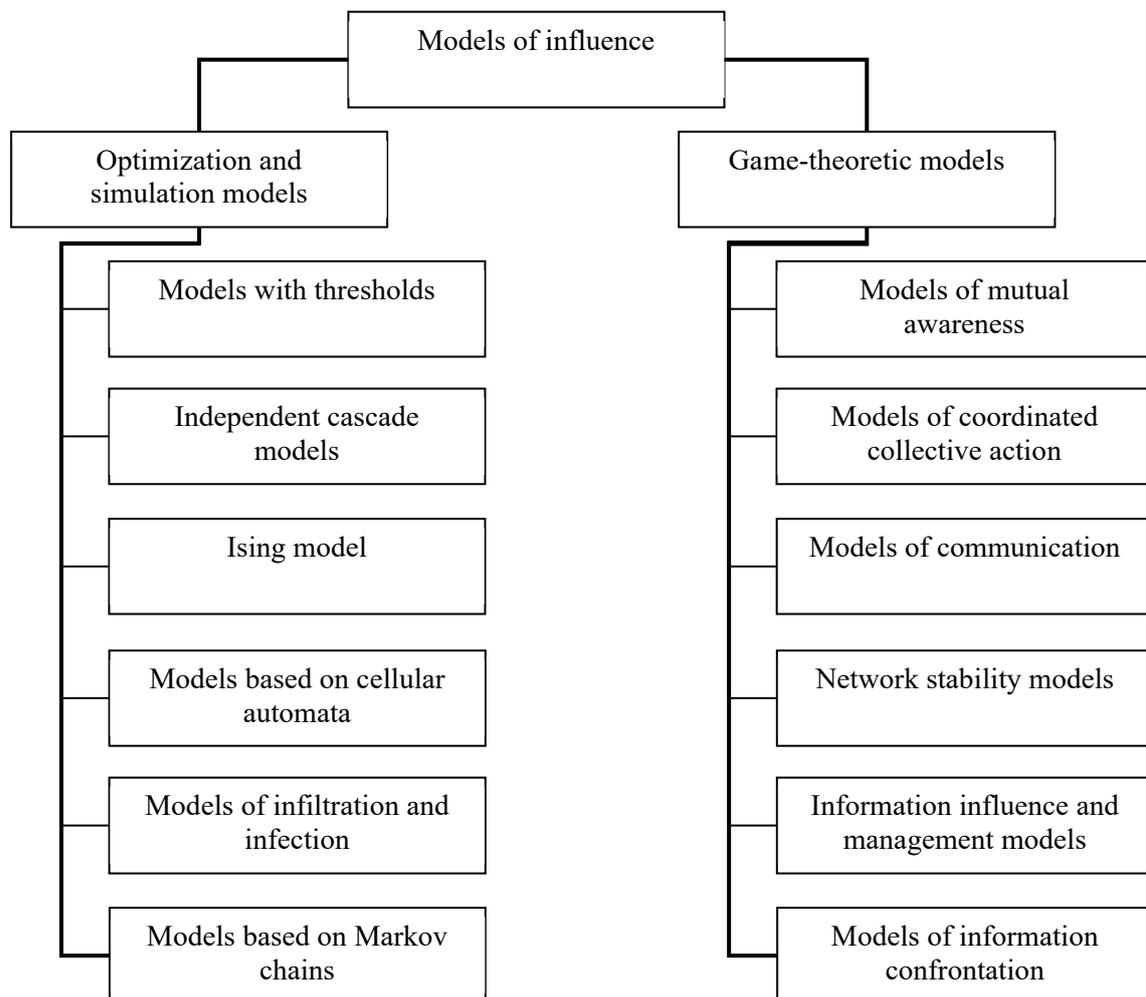
Figure 1: Classification of influence models

As you know, intrusion detection systems (IDS) are one of the mandatory components of the social network security infrastructure [2]. A special role in the protection of social networks is played by the creation of preventive protection systems.

Currently, signature intrusion detection methods are being actively developed (intrusion detection methods based on signatures - templates of typical attacks created based on the headers or contents of network packets [3-6]. And behavioral methods are based not on models of information attacks, but on models of "normal" functioning [7-11].

Working with insights, identifying them, classifying them, and determining the best ways to use them should be done using an interdisciplinary approach that synthesizes statistical, mathematical, and programming methods [12]. The "preprocessing" stage involves identifying the main features and factors, compressing the dimensions of the data multistructure, and other preparatory calculations, including checking the data for quality and the presence of rough outliers. At this stage it is assumed to use the principal component and multidimensional scaling methods. At the stage of "data analysis", hidden insights are supposed to be identified by the mutual correlation (quantitative and qualitative) of the user's characteristics of the media environment [13-14]. To do this, we plan to use the canonical correlation method. Factor analysis allows you to determine the dominant factors that have the main contribution to the variance of a multidimensional data array. An important part of the study will be clustering the data set in a multidimensional feature space. For this purpose, non-hierarchical ones will be used (k-medium, k-medoids, etc..) and hierarchical methods of cluster analysis.

For various information occasions, the analysis of the level of radicality of messages for different social groups, depending on the number of messages, was carried out (Figure 2:).
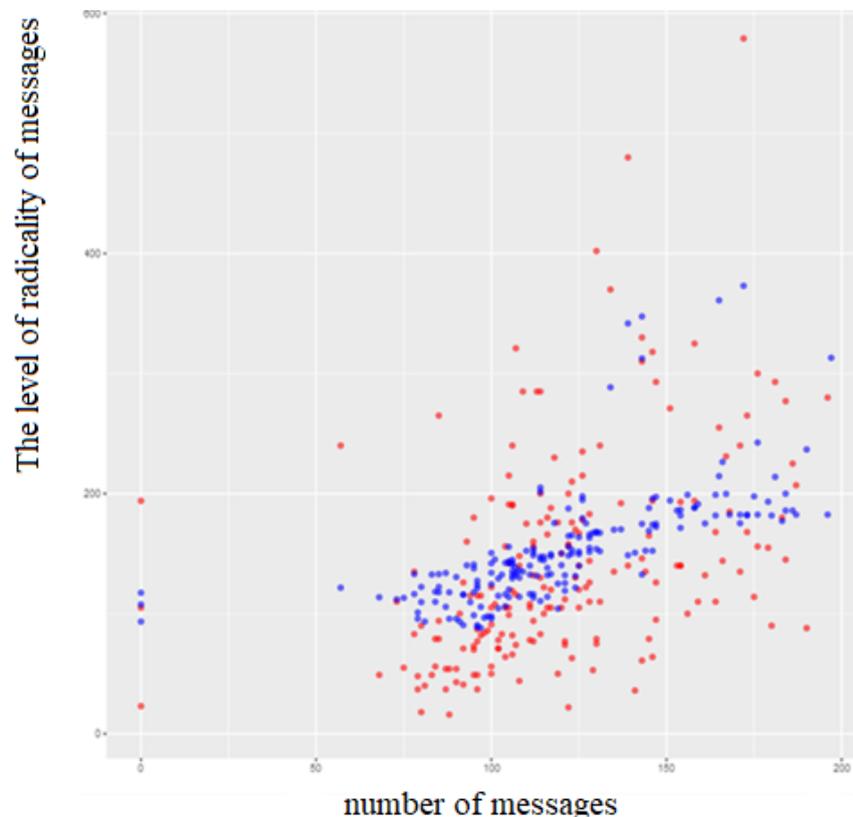
Figure 2: Analysis of the radicality of social groups ' messages for an informational occasion 1 (blue color – positive, red color-negative)

In general, the trend of the level of radicality shows a positive trend in the intensity of messages, regardless of belonging to a particular social group and the nature of the messages. At the same time, it should be noted that the level of radical positive messages has less volatility (Figure 3:).

At the "postprocessing"stage, it is supposed to filter the main "patterns" of features, visualize and interpret the obtained groups or classes of features that determine the main hidden patterns in the data (insights). At the "Analytics" stage, it is supposed to work with insights and determine the most optimal ways to use them. The purpose of this stage is to determine the type of patterns and cause-and-effect relationships between sets of signs of social media activity. To do this, we will use the method of multivariate linear regression, including the use of generalized linear models. Also, at this stage it is planned to develop discriminating rules for the selection of signs of insights of a particular type identified at the previous stage. The established rules can also be represented as a decision tree or regression tree, which is widely used in machine learning and data mining and is a means of predicting user targets based on a set of social media activity characteristics.

For a group of "chronological" data, it is planned to use special methods for the diagnosis and forecast of time series. In particular, such analysis should be based on an objective division of series into trend, quasi-periodic and stochastic ("noise") components. Often the deterministic components of a series are masked by a "noise" component and manifest themselves through the autocorrelated structure of the series. The selection and analysis of such components should be carried out using different methods. The project proposes to use a new method of empirical decomposition of time series into a finite number of orthogonal irregular components - the EMD method. The EMD method does not depend on a strictly defined basis of decomposition functions (Fourier analysis) or on the need for preliminary determination of the parent wavelet (wavelet analysis), which, in fact, is more applicable for a time series with a regular structure. The method has a high degree of localization in the time domain of the decomposition and allows you to effectively isolate the trend component, quasi-periodic

and noise components from non-stationary irregular time series. Using the EMD method will allow you to build a significantly better model of the analyzed time series. The model of trend and quasiperiodic components can serve as polynomials of different degrees and periodic functions, while the model of the stochastic component should be based on autoregression of time series residuals after removing the deterministic components.

Figure 3: Analysis of predictors of radical social groups

The critical point for scripting is the method by which this work can be done. The most obvious, effective and optimal method from the point of view of the subsequent computer algorithmization of the scenario is the decision tree method and other combinational methods based on the basic "trees"method.

Figure 4: An example of a graphical representation of a decision tree for an arbitrary data set

When building trees, the "divide and conquer" strategy is used, which is as follows. In the nodes, starting from the root, the attribute is selected, the value of which is used to divide all the data into 2 classes. The process continues until the stop criterion is met, which is possible in the following situations:

- All (or almost all) data of a given node belongs to the same class.
- There are no signs left on which to build a new partition.
- The tree has exceeded the preset "growth limit".

The decision tree grows while there is still the possibility of constructing new partitions. There may be a situation where the groups are too small and there are too many branch points. In this case, the model is referred to as overridden. It can be inconvenient to use such decision trees in practice. In this case, some reasonable compromise branching depth should be chosen.

Figure 5: shows a decision tree based on an array of data from a 75% sample of 590 social media user records.
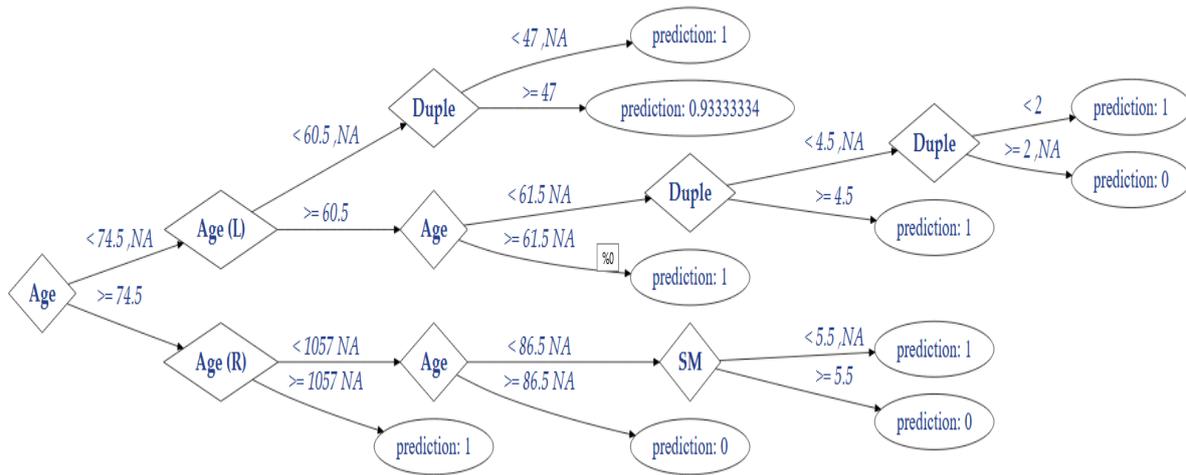

Figure 5: A decision tree for identifying abnormal records (users on the network)

The most important parameters of the model are the intensity of incoming messages and the number of reposts for an informational occasion. The adequacy of the developed model directly depends on the "plausibility" of these values.

2. Conclusions

Data collection technologies for analyzing the socio-media environment of post-conflict societies will be further developed, as well as big data analysis methods for identifying cause-and-effect relationships in the reintegration processes of post-conflict societies.

Methods of agent-based modeling of information and propaganda influence in post-conflict societies using the Internet will be further developed, which will allow analyzing the structural dynamics of reintegration processes in such societies.

The scientific novelty of the research is as follows:
- development of agent-based modeling methods for social media analysis in post-conflict societies;
- improving the use of big data methods to search for insights in social media analysis in post-conflict societies;
- development of scenarios for the development of political processes in post-conflict societies using simulation.

After analyzing the results of the conducted research, it can be concluded that the effectiveness of the chosen strategy for organizing information management significantly depends on the current situation and the selected values of the control parameters. On the one hand, the presented results demonstrate the consistency of the results obtained, and on the other hand, they have the necessary stochastic component. The non-triviality of the presented dependencies is visible, which, indirectly, confirms the relevance of the problem being solved in the work.

3. Acknowledgements

4. References

[1] Y. Y. Tarasevich, V. A. Zelepukhin, Academic network as an excitable environment Comp. research and modeling 7 (2015) 177–183.

[2] J. Goldenberg, B. Libai, E. Muller, Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-Mouth 2 (2001) 11–34.

[3] Md. F. Dewan, M. Z. Rahman, Ch. M. Rahman, Mining Complex Network Data for Adaptive Intrusion, Moscow: BINOM, 2006.

[4] D. Dasgupt, Artificial Immune Systems and Their Applications, Moscow – Fizmatlit, 2006.

[5] A. V. Skatkov, Information Technologies for Critical Infrastructures, Sevastopol: SNTU, 2012.

[6] D. Y. Yeung, Y. Ding, Host-Based Intrusion Detection Using Dynamic and Static Behavioral Models, Journal of Pattern Recognition 36 (2003) 229–243.

[7] C. C. Michael, Ghosh Simple, State-Based Approaches to Program-Based Anomaly Detection ACM, Transactions on Information and System Security 3 (2002) 341–349.

[8] Y. Chen, A. Abraham, B. Yang, Hybrid Flexible Neural-Tree-Based Intrusion Detection Systems, International Journal of Intelligent Systems 22 (2007) 337–352.

[9] T. Shon, J. Moon, A Hybrid Machine Learning Approach to Network Anomaly Detection, Journal of Information Sciences 177 (2007) 3799-3821.

[10] D. V. Moiseev, A. A. Bryukhovetskiy, A. V. Skatkov, Intelligent decision - making support on the level of encryption of information transmitted in the UMV information exchange channels, IOP Conf. Ser.: Mater. Sci. Eng. 734 012086 (2020). https://doi.org/10.1088/1757-899X/734/1/012086.

[11] A. V. Skatkov, A. A. Bryukhovetskiy, D. V. Moiseev, Adaptive vulnerability detection model for unmanned vehicles drugs based on artificial immune systems, IOP Conference Series: Materials Science and Engineering 734 012028 (2020). DOI: iopscience.iop.org/article/10.1088/1757-899X/734/1/012028.

[12] O. Yarmak, Online Surveys in Sociology: Opportunities, Drawbacks and Limitations, 11th International Conference on Computer Science and Information Technologies CSIT 4 (2017) 476–477.

[13] A. Skatkov, A. Bryukhovetskiy, V. Shevchenko, Monitoring of qualitative changes of network traffic states based on the heteroscedasticity effect, Application of Information and Communication Technologies, AICT 2016 - Conference Proceedings, Baku, 7991765 (2016).

[14] A. V. Skatkov, A. A. Bryukhovetskiy, D. V. Moiseev, Intelligent monitoring system for solving large-scale scientific problems in cloud computing environments, Information and control systems 2(87) (2017) 19–25.