UDC 519.876.5

Mathematical models of information concealment cases spreading dynamics in social media

Y. V. Nakonechna[1]

[1] *National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Institute of Physics and Technology*

Abstract

The realities and needs of the globalizing information society require special attention to the problems of information security, both in a technical sense and in a humanitarian one.Cybernetic space and social media as a partial case of it are a constant struggle for the attention of users and the information impact that can be used for their own purposes in the context of information operations and information wars. The directed influence on various actors of the media, representatives of social groups, producers and consumers of content in conditions of information confrontation can pose a direct threat to human security, the security of cyberspace and the security of the state in general. Thus, detection of tools of information influence during the conduct of information wars as information concealment cases in time, is an important condition for ensuring the state of security of society and the state. The use of analytical approach and mathematical modeling enables to prevent, detect and counteract the concealment of information in cyberspace in the subject area of cybernetic security.

Keywords: Mathematical model, SIR model, Daley-Kendall model, cellular automata, information spreading, information operation, rumor spreading, social media

Formulation of the problem

Consealment of information in cyberspace is one of the types of information operations - the interrelated sequences of information influences to achieve the goal. They allow you to hide the fact of their conduction, but at the same time get the target effect, have a low cost of conducting, and the real effect of their realization (change in the company's stock, changing mood active part of the population, loss of confidence in a political or public figure). Such concealment of information, in particular the distribution of fake news, rumors and mysteries, leads to sharp differences of opinion, various conflicts, including political ones. Since most of the exchange of information is currently taking place in cyberspace, timely detection and counteraction to concealment of information is of high value.

Analysis of the literary sources. The spread of misinformation is the new normal. Misinformation spreads in uncontrollable ways in social networks. Misinformation includes rumors, hoaxes and fake news. Therefore, throughout the experience of social media, users face the problem of determining data authenticity and quality. As explained previously, it is hard to rate reliability of a source in such a user-generated-content platform, where sources of information might, by mistake or intentionally, propagate false information. In turn, this causes the spread of polluted information. Thus, it would be hard to determine the factual quality of propagated data and how much consideration should be given to it [1].

False rumors have affected stock prices and the motivation for large-scale investments, for example, wiping out 130 billion dollars in stock value after a false tweet claimed that Barack Obama was injured in an explosion [2]. Indeed, our responses to everything from natural disasters [3] to terrorist attacks [4] have been disrupted by the spread of false news online.

Many methodologies were proposed with the hypothesis that social network users are likely to react to suspicious misinformation in a specific way L. Zhao, J. Yin, and Y. Song [5] suggest, that influences that affect social network users' behavior and willingness to combat rumors during social crises were investigated. A model was developed to better understand how social network users react to rumors during crises. Their approach uses structural equation modeling to evaluate the influence factor of a user's behavior. The authors concluded that people are prone to disputing false new and rumors. This provides further evidence that the detection of misinformation is based on the social network's user, although collaborative wisdom can also be effective.

An automatic detection of rumors spreading on Weibo, a blogging network, mainly popular in China, was proposed in [6]; a graph kernel-based hybrid SVM classifier was used to detect the misinformation. This classifier captures propagation patterns in addition to semantic features such as topics and opinions to evaluate credibility of information.

But almost no studies are devoted to differences between spreading of truthful and false information or exploring why false news may spread with another characteristics than the truth. Regardless of importance of these results they can contain relative error because of smaller sample size of the study and fact that more shares per message can not equate to deeper, broader, or more rapid diffusion). Many scientists engaged in the study of the mathematical aspect of the distribution of fake and mysticism in cyberspace. Thus, some mathematical principles and models were proposed that describe the motion of information flows and characterize the movement of some informational message on the network.

Kermack - McKendrick model [7] was proposed to explain the rapid rise and fall in the number of infected patients observed in epidemics such as the plague (London 1665-1666, Bombay 1906) and cholera (London 1865). It assumes that the population size is fixed (i.e., no births, deaths due to disease, or deaths by natural causes), incubation period of the infectious agent is instantaneous, and duration of infectivity is same as length of the disease. It also assumes a completely homogeneous population with no age, spatial, or social structure. It was the first model possible to apply on different social processes.

Of course, modern models use a combined scientific approach, using the theory of graphs [8], the theory of fuzzy sets [9], equations for describing unstable dynamical systems [10], cellular automata [11], analytical approaches [12], applied mathematical and social sciences - sociological, psychological, and philosophical approaches.

Presentation of the main material

The subject of the study are mathematical models of the dynamics of news spreading in the network of cases of concealment of information based on different approaches of information spreading process modeling, depending on the peculiarities of the information spreading environment and the participants in the dissemination process in the cyberspace.

According to different conditions of news getting, there are few ways of perceiving the news by a network member :

- Spreading of information that does not require a decision (rumors, news). In this case, a person regardless of desire from the state of an uninformed member of the information interaction becomes a carrier of information (the simplest example is advertising, when a person does not make a decision or learn about a new product).
- Distribution of information provided by the agent decision . This process is known as "diffusion of innovations". Knowledge base and interest in certain knowledge directly affects the speed and quality of information coming to the future of its media.
- Combined way in which perceiveing of information is happening depending on special points e.g. probability of information perception, number of social connections, the credibility of sourse and so on.

Based on this, there are different approaches to information spreading modeling.

Epidemiological approach

There is a close resemblance between the dynamics of the spread of the epidemic and the dissemination of information. The classic epidemiological SIR (susceptible - infected- removed) model is used in epidemiology to understand the news dissemination process in the online media network.

Extended SIR model of information spreading

The deterministic epidemiological model SIR (susceptible - infected- removed) describes epidemic transmittion from one individual (agent) to another. This model is based on a homogeneous population mix where every person is equal to other. However, the degree of uniformity of interaction among users will result in the loss of a homogeneous mixing concept. For this reason, the SIR model does not work for a map of a complete scenario for disseminating information among users.

The classic SIR model is characterized by the presence of three types of objects: S are not infected, but susceptable, I are infected, R are healed objects that have immunity. In context of the subject consider S are network agents (users) who have not received an informational message, I have received the message and consider the information given in it to be relevant, R have forgotten the news/lost interest in it.

The general structure of the system described by the SIR model can be represented as:

$$S(t) + I(t) + R(t) = N$$

where $S(t)$, $I(t)$, $R(t)$ show number of objects in each class.

$$\begin{aligned}
\frac{dS(t)}{dt} &= -\frac{\beta I(t)}{N} S(t), \\
\frac{dI(t)}{dt} &= \frac{\beta I(t)}{N} S(t) - \gamma I(t), \\
\frac{dR(t)}{dt} &= \gamma I, \\
\frac{dS(t)}{dt} + \frac{dI(t)}{dt} + \frac{dR(t)}{dt} &= 0.
\end{aligned}$$

where β is the frequency of infication (receiving news by the user),

γ is the frequency of treatment or the rate of immunization (the rate of forgetting a message or its transition to an irregular state).

Note that the distribution of news is possible only under the condition $\beta < \gamma$.

The social network is characterized by time variability. This means that agents can join the network or leave it. Denote by parameter μ the average network connection frequency per unit time. The parameter δ will assume the average frequency of the agent output from the network per unit time. The probability of transition from the invulnerable state to the vulnerable is denoted by the parameter α. We add this condition to the system of equations of the model. The system of equations will be [7, 8]:

$$\frac{dS(t)}{dt} = -\frac{\beta I(t)}{N}\mu(N - S(t)) + \alpha R(t),$$
$$\frac{dI(t)}{dt} = \frac{\beta I(t)}{N}S(t) - \gamma I(t) - \delta I(t),$$
$$\frac{dR(t)}{dt} = \gamma I - \delta R(t) - \alpha R(t),$$
$$\frac{dS(t)}{dt} + \frac{d(t)I}{dt} + \frac{dR(r)}{dt} = 0$$

RnSIR model of rumors and news spreading

The $RnSIR$ (restrained-susceptible-infected-recovered) news dissemination model is proposed to close the gap between the real world and the theoretical assumptions of the SIR model. Thus, the model will help to better understand the process of information spreading and the role of users.

Advantages of model are:

- $RnSIR$ take into consideration the limited nature of the phases in which users may be in the process of disseminating information;
- The presented model better reflects the role of users in the online social interaction network;
- Improving the assessment of the distribution of information.

The proposed $RnSIR$ model, as shown, is an extension of the SIR model. In addition to the three classes that are present in the SIR model, the proposed model has a new Rn class to represent users who restrain themselves from active activities within the affected network (social media, social networks, etc).

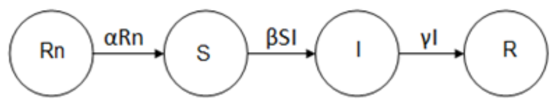

Fig. 1. Scheme of RnSIR information spreading model

The parameters in this model are defined as follows:

Rn is the number of people who refrain from active actions on the network;

S denotes the number of persons who are susceptible to infection;

I represents the number of infected persons at a certain time;

R represents the number of persons who leave the zone of influence at the given time;

α is coefficient of interaction of the person;

β is frequency of infection (receiving news by the user);

γ - the frequency of treatment or the rate of immunization (the rate of forgetting a message or its transition to an irregular state).

The $RnSIR$ model is described by a system of equations:

$$\frac{dRn(t)}{dt} = -\alpha Rn(t),$$
$$\frac{dS(t)}{dt} = \alpha Rn(t) - \beta SI,$$
$$\frac{dI(t)}{dt} = \beta SI - \gamma I(t),$$
$$\frac{dR(t)}{dt} = \gamma I(t),$$
$$N = Rn(t) + S(t) + I(t) + R(t),$$
$$\frac{dS(t)}{dt} + \frac{d(t)I}{dt} + \frac{dR(r)}{dt} = 0$$

In the context of social networks, individuals who are new to the network, have few network connections and low level of inter-user interaction. Therefore, at first, the entire audience is restrained, that is $Rn = N$. As time passes, a person creates new contacts and resources of communication and interaction, and increases the amount of his activity on the network. This indicator is the interaction factor, denoted as α. The parameter α is calculated depending on user activity. So, at this stage, people become sensitive to the influence of the media and the dissemination of information, resulting in $S \subset Rn$. Depending on the β value, a person receives new information. Therefore, in this state $I \subset S$. Again, over time, a person recovers from the influence of information (that is, he forgets the news) with a certain speed γ. The $RnSIR$ model is proposed to fill the gaps in the SIR model. To this end, the model adds a group of reserved users. Thus, the $RnSIR$ model is more appropriate in the real world representation. The results of the study show that the proposed model assumes that the information applies to a maximum of 50% of users [8].

Stochastic approach

The epidemiological approach does not take into account the desire of the network agent to distribute or not to disseminate the information received, which in studies may affect the accuracy of the results and does not show a complete picture of the interaction of the participants in the transmission of the information message. To improve the models obtained earlier and to clarify the real picture of the dissemination of news and rumors, models were built in which the desire to distribute or not distribute information is expressed through some probability of transferring the news from one group of agents to another.

SICR model of rumor spreading

Consider a community consisting of N individuals, which are divided into groups in relation to information: ignorants I are not carriers of some news, distributors S actively disseminate news, dublers C convey the news to those who already have received it , and stifliers R avoid distributing news. Assume that the news is spreaded directly by the distributors. The rules for communicating news from distributors to others are schematically presented in the figure:

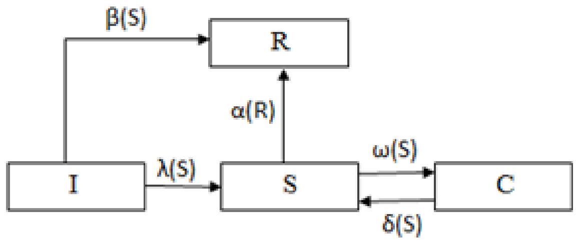

Fig. 2. Scheme of SICR information spreading model

As shown in figure, when the distributor contacts an ignorants, he may perceive the news and become a distributor of the order λ, or not to take it as important, and become a stifler of order β, with rule $\lambda + \beta \leq 1$. When the distributor contacts a different distributor, the initiating distributor becomes a dubler of order ω. When the distributor contacts a stifler, it also becomes a stifler of order α. When the dubler contacts the distributor, the dubler becomes a distributor of order δ. By the definite mechanism of the SICR model, it is described by following equations:

$$\frac{\partial\rho^i(k,t)}{\partial t} = -k(\lambda+\beta)\rho^i(k,t)\sum_{k`} P(k`|k)\rho^s(k`,t)$$
$$\frac{\partial\rho^s(k,t)}{\partial t} = k\lambda\rho^i(k,t)\sum_{k`} P(k`|k)\rho^s(k`,t) - k\omega\rho^s(k,t)\sum_{k`} P(k`|k)\rho^s(k`,t) -$$
$$k\alpha\rho^s(k,t)\sum_{k`} P(k`|k)\rho^r(k`,t) + k\delta\rho^c(k,t)\sum_{k`} P(k`|k)\rho^s(k`,t),$$
$$\frac{\partial\rho^c(k,t)}{\partial t} = k\omega\rho^s(k,t)\sum_{k`} P(k`|k)\rho^s(k`,t) - k\delta\rho^c(k,t)\sum_{k`} P(k`|k)\rho^s(k`,t)$$
$$\frac{\partial\rho^r(k,t)}{\partial t} = k\beta\rho^i(k,t)\sum_{k`} P(k`|k)\rho^s(k`,t) + k\alpha\rho^s(k,t)\sum_{k`} P(k`|k)\rho^r(k`,t)$$

where $\rho^s(k,t), \rho^i(k,t), \rho^c(k,t), \rho^r(k,t)$ are the number of nodes (users) belonging to class k for one of the specified states: S, I, C or R. These values satisfy the normalization conditions:

$$\rho^s(k,t) + \rho^i(k,t) + \rho^c(k,t) + \rho^r(k,t) = 1$$

In addition, sometimes small communities can accurately describe the dynamics of the information dissemination process on large networks, so a small homogeneous network of users can be used to simulate the distribution of news. Therefore, the above equations become the following:

$$\frac{d\rho^i(t)}{dt} = -(\lambda+\beta)\bar{k}\rho^i(t)\rho^s(t)$$
$$\frac{d\rho^s(t)}{dt} = \lambda\bar{k}\rho^i(t)\rho^s(t) - \omega\bar{k}\rho^s(t)\rho^s(t) - \alpha\bar{k}\rho^s(t)\rho^r(t) + \delta\bar{k}\rho^c(t)\rho^s(t)$$
$$\frac{d\rho^c(t)}{dt} = \omega\bar{k}\rho^s(t)\rho^s(t) - \delta\bar{k}\rho^c(t)\rho^s(t)$$
$$\frac{d\rho^r(t)}{dt} = \beta\bar{k}\rho^i(t)\rho^s(t) + \alpha\bar{k}\rho^s(t)\rho^r(t)$$

where k is an average degree of system (average number of interclass transitions) and the condition of normalization is fulfilled, so the initial conditions are given as follows:

$$\rho^s(0) = \frac{1}{N},$$
$$\rho^i(0) = \frac{N-1}{N},$$
$$\rho^c(0) = 0,$$
$$\rho^r(0) = 0$$

N is number of network agents [13].

Daley-Kendall stochastic model

The well-known Daley-Kendal method is a mathematical model for simulating the process of disseminating information (rumors, messages), also called the DK-model. This model divides all participants of the distribution of some informational message into three groups:

- 1st group that begins to spread the message (U);
- 2nd group that after spreading the message continues to distribute it (V);
- 3rd group that after receiving the message decides not to distribute it (W).

The model scheme is presented in Figure 3.

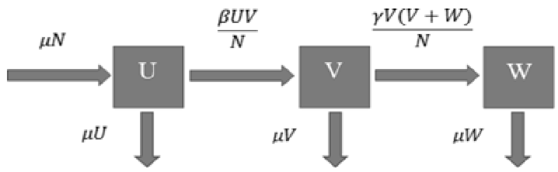

Fig. 3. Scheme of Daley-Kendall stochastic model information spreading model

In model N is the number of participants in the distribution process. The message is taken with the probability $\frac{\beta}{N}$. The degree of acceptance of a message is determined by the μ parameter. When the message distributor faces the W audience, then the distribution stops and the probability that it will happen is equal to $\frac{\gamma V(V-W)}{N}$. The news filed in the message loses its value over time. This probability is determined by the factor γ. This can be explained by the fact that the news ceases to be a novelty or there are no remaining parts that can be transmitted. The model can be represented as [14, 15]

$$\frac{dU(t)}{dt} = \mu N - \frac{\beta U(t)V(t)}{N} - \mu U(t),$$
$$\frac{dV(t)}{dt} = \frac{\beta U(t)V(t)}{N} - \frac{\gamma V(t)(V(t)-W(t))}{N} - \mu N,$$
$$\frac{dW(t)}{dt} = \frac{\gamma V(t)(V(t)-W(t))}{N} - \mu W(t).$$

Discrete time basis approach

To describe the processes of information spreading in a social network it can be viewed as a complex adaptive system, that is, a system consisting of a large number of agents, the interaction between which leads to large-scale, collective behavior that is difficult to predict and analyze. Cellular automata [14], Markov chains and Maki-Tompson-Daley-Kendall [15, 16] stochastic model are sometimes used to model and analyze such complex systems [16, 17].

Cellular automata approach

Cellular machines are considered an effective tool for studying complex systems. By studying the emergence of complex systems by observing cell interactions, it is possible to simulate the necessary characteristics of the system. Cellular automata consists of a set of objects (in this case, agents) that usually form a regular lattice.

The cellular automata is a discrete dynamic system, which includes homogeneous cells, connected with each other. All cells form an automaton. The condition of each cell is determined by the cells located near the cell. The set of "closest neighbors" is called vicinity of an automaton with the number j. The state of the cellular automaton j at time $t+1$ is determined as follows:

$$y_j(t+1) = F(y_i(t), O(j), t)$$

where where F is the rule that can be expressed (for example, in the language of the Boolean algebra), $O(j)$ is vicinity of cell, t is a step.

The cellular automaton is determined by the rules:

- The values change of each cell occurs simultaneously (the step is to change the time unit);
- The network of a cellular automaton is homogeneous, that is, the rules of change are the same for all cells;
- The number of cell states is finite.

The theory of cellular automata is used to analyze the diffusion of innovations, this process is very similar to the distribution of news on the Internet. The simplest function of the transformation of a model corresponds to the following rules: the individual corresponds to one cell, which can accept two states: 1 if news is accepted, 0 if news is not accepted. It is assumed that once receiving information, the state remains unchanged. The machine decides to accept news based on the opinion of its closest neighbors, if m neighbors support innovation and p is the probability of receiving news (generated during the model), then if $pm \geq R$, where R is a fixed threshold, the cell takes the news. An additional condition for the type of news may be imposed: the cell has fresh news (black), the cell has outdated information (gray), the cell has no information or has forgotten about it (white)[18].

Rules for distributing news:

1) in the beginning each cell is painted white, except for one black cell (which received a message with the news);
2) the white cell can change the color to black or stay white (that is, it has taken the news or remained in the dark);
3) the white cell changes its color if condition (1) is performed in the diffusion model (m is the number of black cells, if $m < 3$, then p is increased by 1.5 times);
4) if the cell is black and all the cells around are only black or gray, it changes its color to gray (the news is outdated);
5) if the cell is gray and the cells around are only black or gray, then it changes its color to white (information is forgotten).

Conclusions

Despite the complexity of modeling information dissemination and diffusion of innovations, more and more mathematical models applicable to modeling the dynamics of the dissemination of information and news in social media are being introduced and studied, also the peculiarities of interaction between users are exploring.

An epidemiological approach, an approach using the decision making theory and modeling with discrete time models was considered. An overview of mathematical models that could be used to model the distribution of news and rumors as separate cases of concealing information in the information environment was conducted. According to the chosen approaches, mathematical models were considered: extended SIR model of information spreading, RnSIR and SICR models of rumor spreading, Daley-Kendall stochastic model and cellular automata.

Analyzing the listed models following can be summarized:

- The extended SIR model makes it possible to estimate audience coverage, but does not correspond to reality in terms of displaying the dynamics of news spread over time, and also the presence of a recovery parameter makes the use of the model more complicated, since the assessment of the rate of transition of news in an unmet state is a complex indicator and needs expert evaluation [19];
- $RnSIR$ model predicts spread of information better than the SIR model and helps to understand the network of users and furthermore, and in a certain degree to interpret the user behavior identifying the users who are most likely to promote the information(or product);
- In comparison with the classical SIR rumor spreading model $SICR$ model is improved by adding a new recall group of people who stop spreading the rumor temporarily and could spread it again when incentives occur. In result this made the model more suitable for rumor spreading process, taking into consideration user's decision about news spreading expediency;
- The *Daley-Kendall model* has the same disadvantage as extended SIR model, although the number of engaging audiences for using the model is estimated fairly accurately, some time after the publication of the news, the model shows a significant leap in the number of news readings that are not true [19];
- The *model of a cellular automaton* with complication enables to evaluate not only the number of consumers of news but also its distribution channels,

but its significant disadvantage is the discrepancy in displaying the dynamics of news distribution in time, as well as considerable time and complexity of computational steps [19].

All of these models have a high potential in the field of modeling the distribution of messages and allow taking into account the features of interaction of users in social media, the presence or absence of communication channels, the likelihood of perceptions or rejection of news, the desire or reluctance to distribute it, taking into account the role of a centralized news distributor, etc. But within the domain, which considers the distribution of fakes and rumors as a tool for conducting information operations, using these models requires improving them or creating new models using combined approaches, since the above models do not fully reflect the specifics of the information dissemination process in cyberspace, which is characterized by the use of emotional and psychological influence, as well as the focus of influence on the informationally-control systems and civilian infrastructure [20]. In addition, the quality of such models on real data is offset by many parameters, not all of which are taken into account when constructing a mathematical model, so in the case of using they need to be finalized to obtain accurate results.

Accordingly, aforecited models do not reflect the user's behavior and the life cycle of the information message to the full within the considered subject area.

References

[1] C. Shao G.L. Ciampaglia A. Flammini, Menczer F. Hoaxy: A Platform for Tracking Online Misinformation. — 2016. — P. 745–750.

[2] Rapoza K. Can 'fake news' impact the stock market? // Forbes. — 2016. — no. 26 February.

[3] M. Mendoza B. Poblete, Castillo C. // Proceedings of the First Workshop on Social Media Analytics (Association for Computing Machinery, ACM). — 2010. — P. 71–79.

[4] K. Starbird J. Maddock M. Orand P. Achterman, Mason R.M. // iConference 2014 Proceedings, iSchools. — 2014.

[5] L. Zhao J. Yin, Song Y. An exploration of rumor combating behavior on social media in the context of social crises // Comput. Human Behav. — 2016. — no. 58. — P. 25–36.

[6] K. Wu S. Yang, Zhu K.Q. False rumors detection on Sina Weibo by propagation structures, // Proceedings - International Conference on Data Engineering. — 2015. — no. 2015-May.

[7] W.O. Kermack, A.G. McKendrick. A contribution to the mathematical theory of epidemics // Proc. of the Royal Society A: Mathematical, Physical and Engineering Sciences. — 1927. — no. 115.

[8] N. Sumith B. Annappa Swapan Bhattacharya. Rn-SIR:A New model of Information Spread in Online Social Networks // IEEE Region 10 Conference (TENCON). — 2016.

[9] N. Pissanidis H. Rõigas M. Veenendaal. The Social Side of «Cyber Power» // Social Media and Cyber Operations, on 8th International Conference on Cyber Conflict Cyber Power, NATO CCD COE Publications, Tallinn. — 2016.

[10] Rajdev Meet, Lee Kyumin. Fake and Spam Messages: Detecting Misinformation during Natural Disasters on Social Media // IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. — 2015.

[11] Ailian Wang Weili Wu Junjie Chen. Social Network Rumors Spread Model Based on Cellular Automata // 10th International Conference on Mobile Ad-hoc and Sensor Networks. — 2014.

[12] Silverman C. This analysis shows how viral fake election news stories outperformed real news on Facebook // BuzzFeed News. — 2016. — no. 16 November.

[13] Jiajia Wang Laijun Zhao Rongbing Huang, Chen Yucheng. Rumor spreading model on social networks with consideration of remembering mechanism // IET International Conference on Smart and Sustainable City 2013 (ICSSC). — 2013.

[14] Gorkovenko D. K. Obzor modeley rasprostraneniya informacii v sotsialnykh setyakh // Young scientist. — 2017. — no. 8(142). — P. 23–28.

[15] D.J. Daley, D.G. Kendall. Stochastic rumors // J. Inst. Math. Appl. — 1965. — Vol. 142. — P. 42–55.

[16] Rosnow R. L. Inside rumor: a personal journey // Am Psychol. — 1991. — Vol. 46, no. 5. — P. 484–496.

[17] V. V. Karelin V. M. Bure L. N. Polyakova, Svirkin M. V. A Stochastic Model of Rumour Spreading // Constructive Nonsmooth Analysis and Related Topics (dedicated to the memory of V.F. Demyanov) (CNSA). — 2017.

[18] Maki D. P., Thompson M. Mathematical models and applications. — Englewood Cliffs : Prentice-Hall, 1973.

[19] Y. V. Nakonechna A. B. Kachynsky. Mathematical models of the information dissemination dynamics in social media // Materials of XVI All-ukrainian scientific-practical conference of students, aspirants and young scientists 'Theoretical and applied problems of physics, mathematics and informatics'. — 2018.

[20] Bartolomew D. J. Stochastic models for social processes. — London school of economics and political science : John Wiley and Sons, 1985.